ISSN : 2165-4069(Online)
ISSN : 2165-4050(Print)

IJARAI

International Journal of
Advanced Research in Artificial Intelligence

Volume 1 Issue 1

www.ijarai.thesai.org

IJARAI

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

Editorial Preface

From the Desk of Managing Editor…

"The question of whether computers can think is like the question of whether submarines can swim." ― Edsger W. Dijkstra, the quote explains the power of Artificial Intelligence in computers with the changing landscape. The renaissance stimulated by the field of Artificial Intelligence is generating multiple formats and channels of creativity and innovation.

This journal is a special track on Artificial Intelligence by The Science and Information Organization and aims to be a leading forum for engineers, researchers and practitioners throughout the world.

As said by James F. Allen, Professor of Computer Science, University of Rochester, "Artificial Intelligence (AI) is not the science of building artificial people. It's not the science of understanding human intelligence. It's not even the science of trying to build artefacts that can imitate human behaviour well enough to fool someone that the machine is human, as proposed in the famous Turing test. AI is the science of making machines do tasks that humans can do or try to do. The field (of AI) focuses on the more complex things that people do."

The journal reports results achieved; proposals for new ways of looking at AI problems and include demonstrations of effectiveness. Papers describing existing technologies or algorithms integrating multiple systems are welcomed. IJARAI also invites papers on real life applications, which should describe the current scenarios, proposed solution, emphasize its novelty, and present an in-depth evaluation of the AI techniques being exploited.

IJARAI focusses on quality and relevance in its publications. In addition, IJARAI recognizes the importance of international influences on Artificial Intelligence and seeks international input in all aspects of the journal, including content, authorship of papers, readership, paper reviewers, and Editorial Board membership

The success of authors and the journal is interdependent. While the Journal is in its initial phase, it is not only the Editor whose work is crucial to producing the journal. The editorial board members , the peer reviewers, scholars around the world who assess submissions, students, and institutions who generously give their expertise in factors small and large— their constant encouragement has helped a lot in the progress of the journal and shall help in future to earn credibility amongst all the reader members.

I add a personal thanks to the whole team that has catalysed so much, and I wish everyone who has been connected with the Journal the very best for the future.

Thank you for Sharing Wisdom!

Managing Editor
IJARAI
Volume 1 Issue 1 April 2012
ISSN: 2165-4069(Online)
ISSN: 2165-4050(Print)

Editorial Board

Reviewer Board Members

CONTENTS

A Method for Chinese Short Text Classification Considering Effective Feature Expansion

Mingxuan liu 1
School of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing 40065, China

Xinghua Fan 2
School of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing 40065, China

***Abstract*—This paper presents a Chinese short text classification method which considering extended semantic constraints and statistical constraints. This method uses "HowNet" tools to build the attribute set of concept. when coming to the part of feature expansion, we judge the collocation between the attribute words of original text and the characteristics before and after expansion as the semantic constraints, and calculate the ratio between the mutual information of the original contents and the features before expansion versus the mutual information of the original contents and the features after expansion as statistical constraints, so as to judge whether feature expansion is effective with this two constraints , then rationally use various semantic relation word-pairs in short text classification. Experiments show that this method can use semantic relations in Chinese short text classification effectively, and improve the classification performance.**

Keywords-component; short text; classification; semantic relations; semantic constraints; statistical constraints; HowNet.

I. INTRODUCTION

The short-text classification is an automatic classification for short texts (The text length is usually less than 160 characters). It is required for filtering information such as mobile phone short message, web comments, network chat, etc. And it has very important application prospect. However, short-text has many inherent characteristics such as short length, weak signal of concept description, etc. Therefore, traditional text classification methods cannot be applied to short-text classification better. An effective way of short-text classification is to use some extra information to assist classification [1-2]. Some methods has already achieved certain effect, one of which is based on hyponymy relation and is proposed by Sheng Wang[3] and another one is Agent and Patient Relation Acquisition for Short-text Classification which is proposed by Dingbang Wei[4]. Theoretically, when combining the word-pairs sets extracted by the above different two semantic relationships into a bigger word-pairs extension set simply, the extension result of short text will be better. However, experimental results in this paper show: when doing feature extension for test text by using the word-pairs extension set directly[5-6], the classification performance of short text isn't improved, but is slightly reduced. Analyzing the results, the reason is that when introducing word-pairs of a variety of different semantic relations at the same time, it will be at the greatly increased risk of the introduction of noise. The classification performance is reduced, for the features expansion isn't effective. If we want to expand features for short-text effectively, the following two issues must be resolved:1、How to determine whether noise is introduced when expanded，2、How to apply different semantic relation word-pairs to classification of short text to improve the classification performance. For the above problems, this paper presents a classification method for short Chinese text considering effective feature expansion SCTCEFE(Short Chinese Text Classification Considering Effective Feature Expansion).

II. A CLASSIFICATION METHOD OF SHORT CHINESE TEXT CONSIDERING EFFECTIVE FEATURE EXTENSION (SCTCEFE)

The core idea of SCTCEFE is: through the constraint conditions, judge whether the knowledge expressed by the text before and after feature expansion changes, to ensure the availability of feature expansion and reduce the risk of noise introduction.

For the problem determining whether noise is introduced when expanded, by judging whether the knowledge expressed by extended text information and the original text are unified, solve the problem. Semantic relationship reveals the hidden semantic knowledge in the knowledge expressed by texts. Thus the article uses the following two methods to deal with different semantic knowledge: 1. relationship of similar concepts (Hyponymy relationship), by replacement of the similar concepts strengthen the description of the short text.

When using the concept information, we should determine whether the knowledge focus expressed by the original text is the similar concept between the relation word-pairs. for example, in the short text "宝马的外观非常重要", "外观" is the attribute of "汽车", we can determine the expansion is effective; But in the short text "我比较喜欢你那件宝马衣服的衣领", as "衣领" isn't the attribute of "汽车" , we can determine the expansion is invalid; 2. relationship of different concepts(agent-patient relationship), by introduction of new concept information mine the additional information of the short text[7].When using the information, we should judge whether there's strong correlation strength between new concept information and the other information of the original text.

For the problem of applying different semantic relation word-pairs to classification of short text to improve the classification performance, According to the different types of different relation word-pairs, the method establishes the semantic constraint for the similar concept relationship and the statistical constraint for the different concept relationship to judge whether feature expansion is effective.

The specific thread of SCTCEFE is: Build the attribute sets as the semantic constraint for the word-pairs of similar concept relationship, and calculate the ratio between the mutual information of the original content and the original feature versus the mutual information of the original content and the extended feature as the statistical constraint for the different concept relationship word-pairs. Combine with semantic constraints and statistical constraints; expand features effectively for short text with the use of relation words.

A. The standard of semantic constraint and statistical constraint

In this paper, we propose a concept of information integrity to describe the contribution value of a single word to the information integrity of a document. For example, the information integrity of a word A in the document D is described as P (A|D) ：

$$P(A\mid D)=\frac{\sum_{0<i<m} I(A,W_i)}{I(m)} \qquad (1)$$

I(m) represents the amount of information contained in the document D after the removal of a word A, m is the number of words in the document D, and I(A, Wi) represents the relevant information between a word A and the information Wi of a certain word in the text D. In this paper, we suppose that every sentence is an independent knowledge-expressed document. By feature expansion, we use $consist\left[\left(A,\overline{A}\right)\mid D\right]$ to express the consistency of text-expressed knowledge after expansion. When $consist\left[\left(A,\overline{A}\right)\mid D\right]>0$, there is a certain consistency between the knowledge of the extended text and the original text.

Semantic constraint: used for judging whether the similar concept information of the similar relationship word-pair is the keystone in the text. When there appears an attribute word which can match the relation word-pair simultaneously in one text, we can suppose that the similar or identical concept information of the word-pairs is used in the text, which can be expressed as:

$$consist\left[\left(A,\overline{A}\right)\mid D\right]=\frac{P\left(\overline{A}\mid D\right)}{P(A\mid D)}=\frac{\sum_{0<i<\sum_{0<i<m} I(A,W_i)} I(\overline{A},W_i)}{\sum_{0<i<m} I(A,W_i)} \qquad (2)$$

$\overline{A}$ represents the relation word of concept information similar to the word A. When there is an attribute-host relationship between an attributive (or attributive value) word and the word $\overline{A}$, then $I(\overline{A},W_i)=1$, otherwise it is 0. In this paper, when $\sum_{0<i<m} I(A,W_i)$ =0, $consist\left[\left(A,\overline{A}\right)\mid D\right]$=0.

statistical constraint: used for judging whether the word-pair which has different concept information is the hidden information of the text, when the ratio between the mutual information of the introduced new concept information and the original text versus the mutual information of the original concept information and the original text is within a certain range, it's considered that the concept information hides in the original text, which can be expressed as:

$$consist\left[\left(A,\overline{\overline{A}}\right)\mid D\right]=\frac{P\left(\overline{\overline{A}}\mid D\right)}{P(A\mid D)}=\Phi-\left|\frac{\sum_{0<i<m} MI\left(\overline{\overline{A}},W_i\right)}{\sum_{0<i<m} MI\left(A,W_i\right)}-1\right| \qquad (3)$$

$\overline{\overline{A}}$ is a word set which includes the new concept information and is associated with the word A. $MI\left(\overline{\overline{A}},W_i\right)$ is the mutual information between the word $\overline{\overline{A}}$ and the word W_i, and Φ is an adjustable threshold.

1) Establishing the attribute sets of concept information

Understanding of attribute: Any object must carry a set of attributes. Similarities and differences between the objects are determined by the attributes they each carry. There will be no object without attributes. Human beings have natural attributes such as race, color, gender as well as social attributes such as nationality, class origin, job, wealth etc. Under specific conditions, it is true to say that the attached attributes are even more important than the host itself. This paper presents a method by extracting the co-occur word-pairs which can constitute the relationship of concept-attributes to establish the attribute sets of all concept information in the training set.

A conventional method of Getting concept - attribute relations: HowNet determines all the hosts corresponding to attributes, and is signed with pointer &, for example: 外观: attribute|属性,appearance|外观,&physical|物质;This concept indicates that appearance belongs to the information of "属性", it can be understood as "外观", relying on its host "物质". On the other hand, the hyponymy relation of words indicates the universality and individuality between words. On the basis of universality inherits from higher word, the lower word has part of its own characteristics. Therefore, to any word-pair (C1,C2), if the first semanteme of C1 is an attribute, and its host is C2 or the higher concept of C2 , then C1 and C2 constitute the concept-attribute relationship directly, and C1 exists as the attribute of C2.

An improved method: because Chinese short texts often omit attribute words and adopt 'concept + attribute value' expression in order to give the essentials in simple language. Through common methods we cannot affirm the concept-attribute relationship between the two. However, in practice, attribute value embodies attribute, which should also be summed up in the attribute set of concept. This paper introduces an improved method to obtain concept-attribute relationship word-pairs:

a) For any word-pair（C1,C2）, if it contains one attribute value word, execute step (2)，if it contains one attribute word, execute step (3)，if it contains neither, the word-pair has no concept-attribute relationship；

b) Find attribute words corresponding to the attribute value words, then execute step(3);

Find the host of the attribute word, and judge whether the other word in the word-pair is the lower concept of the host, if yes, the word-pair has concept-attribute relationship, or else the word-pair has no concept-attribute relationship.

2) Building characteristics word-pairs and the attribute sets of the word-pairs

For feature word-pairs which have a similar concept relationship, in order to calculate its semantic constraint, we need to construct the attribute sets of the relationship, Take the hyponymy relations adopted by this paper for example. The feature of the attribute set applied to these feature word-pairs is constructing the concept-attribute relations with both the words in the same pair.

c) according to the method of literature [3], obtain feature word-pairs set I of hyponymy relation, I1={(W1,W2) | (W1,W2) appearing in the word-pairs obtained from the training set, and constitute the hyponymy relationship };

d) To every word-pairs (W1, W2), according to the method mentioned in 1.1 in this chapter, obtain the attribute set of W1, I11= {(C1) | (W1,C1) appearing in the word-pairs obtained from the training set, and (W1,C1) constitute the concept-attribute relation}. Also the attribute set of W2, I12 = {(C2) | (W2,C2) appears in the word-pairs obtained from the training set, and (W2,C2）constitute the concept-attribute relations};

e) Construct attribute sets of every feature word-pairs (W1, W2), I2 = {(C) | C∈I11 且 C∈I12}.

B. The Algorithm Description of Feature Expansion based on semantic relations

Input: short-text for test, feature word-pairs sets, attributes sets of feature word-pairs, mutual information between the words.

Output: Feature vector of the extended short-text.

1) For any word of the short-text, inquire the word-pairs sets. if there exists a record of a similar concept relationship, go to step 5, else if there exists a record of different concept relationships, go to step 2, else go to step 9;

2) If there is only one word-pair ti-tj, go to step 4, else if there are several word-pairs, go to step 3;

3) Extract the right words of all the word-pairs related to ti and form into Tx, if ∋tj∈TX, and tj can be found in the vector space of this short-text, go to step 7, else extract tj , the right word of the word-pair with the highest strength, and go to step 4;

4) Extract tj ,the right word of this word-pair, if tj cannot be found in the vector space of this short-text, go to step 6, else go to step 7;

5) Extract the word set TY in the text, if there exist tk∈TY and tk∈TZ（attribute set of the word-pair（ti，tj））, go to step 8, else go to step 10;

6) Calculate the mutual information between tj and other words in the text, and go to step 8 when meeting the requirements , else go to step 10;

7) Calculate the mutual information between tj and other words in the text, and go to step 9 when meeting the requirements , else go to step 10;

8) Insert tj into the vector space of this short-text;

9) Raise the frequency of tj in the vector space of this short-text at λ. (0<λ<1);

10) Don't extend this word, and input and seek the next word.

III. EXPERIMENT

When using the STCEFE with the use of statistical constraints and semanteme constraints, the experiment to test the effect of the short-text classification uses two different semantic relations.

Experiments dataset: The dataset used here is 4702052 Chinese short-texts which compose of 12 categories, including finance, real estate, international news, national news, military, science and technology, women, cars, book reviews, sports, games, entertainment. All of short-texts are titles of the news or Netizen comments about topics from Sina website and Netease website. We divide texts of each category into four parts randomly and averagely, one part as testing data, the rest as training data.

The Evaluation of Classification Performance: In this experiment, we use the following indexes in the evaluation of classification: Precision (P), Recall (R), F1-measure, and Macro-F1= $\frac{1}{n}\sum_{i=1}^{n} F1_i$

In experiments, we only adopt statistical constraint when selecting 4 kinds of features respectively. According to different thresholds, the revealed change trends of F1 are in nearly identical form, and when threshold Φ is set as 0.1, the classification result will work out for the best when Hyponymy Relationship word-pairs and Agent and Patient Relation word-pairs are both used for text classification. However, the result is still not ideal, we can conclude that only by calculating mutual information the response to whether the introduced new information matches the other information in the original text is not so accurate, thus the method should be in conjunction with other methods.

Experiment introduction: By taking sentences in the text as windows, extract the binary word-pairs in the training set and calculate all mutual information between every word-pair. Build the attribute set of concept according to the method in 1.1 of this article, using the hyponymy word-pairs and agent-patient word-pairs to build an attribute set of feature word-pairs by the method in 1.2 of this paper.

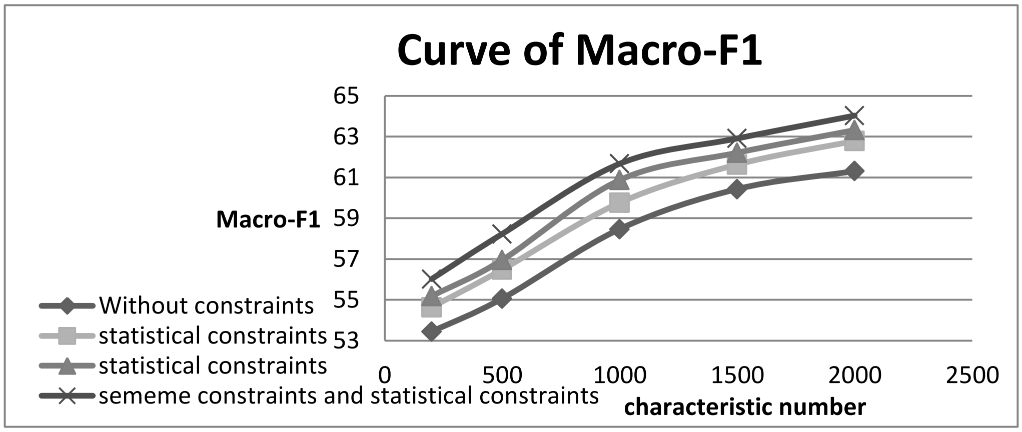


Figure 1. Curve of Macro-F1 by experiment

This experiment uses CHI feature selection, and takes the Naïve Bayes as the classifier. Then 5 groups of experiments for short-text classification are taken, and the sizes of the feature space are 200, 500, 1000, 1500, 2000.

Method 1: According to the expansion method described in [5], expand test text set by hyponymy word-pairs and agent-patient word-pairs.

Method 2: According to the expansion method only by the statistical constraint in this article and threshold is set as 0.1, expand test text set by hyponymy word-pairs and agent-patient word-pairs.

Method 3: According to the expansion method only by the semantic constraint in this article, expand test text set by hyponymy word-pairs and agent-patient word-pairs.

Method 4: According to the expansion method both of the semantic constraint and the statistical constraint in this article and threshold is set as 0.1; expand test text set by hyponymy word-pairs and agent-patient word-pairs.

The experimental results:

The classification results are shown in the figure, and we can arrive at the following review:

A. *Method 2 and 3 are both superior to method 1, indicating that when many kinds of semantic relationship words are used at one time to assist in categorizing the text, the traditional classification performance is far from ideal. By the expansion strategy judging the consistency between the expanded contents and the original text to decide to do feature expansion or not, method 2 and 3 could bring a certain improvement on the categorizing performance.*

B. *The result of method 3 is slightly superior to method 2, indicating the help of semantic constraint to the accuracy of text expansion is slightly superior to the help of statistical constraint.*

The result of method 4 is the best, indicating when using many semantic relationship words to assist in categorizing the text, a combination of semantic constraint and statistical constraint could be a bigger help to text categorizing. But the categorizing performance is still not high, indicating there is still a big part of hidden information to be mined in the short text.

IV. CONCLUSION

The paper proposes a short text classification method in consideration of characteristic effective extension, and constructs a concept of information completeness through semantic constraint and statistical constraint. Based on the concept when doing feature extension for the original text with the use of semantic relation word-pairs, judge whether the expansion is valid. Moreover, review the effectiveness of the classification method through experiments, and thus conclude that: 1) the short text classification method combining many semantic relationships and doing effective expansion with considered characteristics could enhance classification performance; 2) We can consider the combination of the other methods, to acquire more accurate constraint information judging the consistency of text contents.

ACKNOWLEDGMENT

The research is supported in part by the National Natural Science Foundation of China under grant number 6070301, cultivate project for significant research plan of the National Natural Science Foundation of China under the grant number 90924029 named “Research on the mechanism and related technology of public opinion on internet for abnormal emergency” and the Nature Science Foundation of Chongqing province in China under grant number CSTC, 2009BB2079.

REFERENCES

[1] ZELIKOCITZ S.Transductive LSI for short text classification problems[C]//Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference. Florida: AAAIPress, 2004:556-561.

[2] ZELIKOCITZ S, MARQUEZ F. Transductive learning for short text classificaton problems using latent semantic indexing[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2005, 19(2):143-151.

[3] Sheng Wang, Xinghua Fan , XianLin Chen. Chinese short text classification based on hyponymy relation[J]. Journal of Computer Applications, 2010, 30(3):603-606.

[4] Xinghua Fan, Dingbang Wei. A Method of Agent and Patient Relation Acquisition[C].2011 International Conference on Computer Science and Informatin Engineering.

[5] Xiwei Wang, Xinghua Fan, Jun Zhao. Method for Chinese short text classification based on feature extension[J]. Journal of Computer Applications, 2009, 29(3):0843-0845.

[6] Xinghua Fan, Maosong Sun. A High Performance Two-Class Chinese Text Categorization Method [J].Chinese Journal of Computer.2006, 29(1):124-131.

[7] Ramage D,Rafferty AN,Manning CD.Random walks for text semantic similarity//Proceedings of the 2009 Workshopon Graph— based Methods for Natural Language Processing.Suntec,Singapore,2009:23-31.

AUTHORS PROFILE

Mingxuan Liu born in 1988, is Graduate of Chongqing University of Posts and Telecommunications. His research interests is Chinese information processing.

Xinghua Fan born in 1972, is Professor and Ph.D. of Chongqing University of Posts and Telecommunications. His research interests include natural language processing and information retrieval.

A Sparse Representation Method with Maximum Probability of Partial Ranking for Face Recognition

Yi-Haur Shiau
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan

Chaur-Chin Chen
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan

***Abstract*—Face recognition is a popular topic in computer vision applications. Compressive sensing is a novel sampling technique for finding sparse solutions to underdetermined linear systems. Recently, a sparse representation-based classification (SRC) method based on compressive sensing is presented. It has been successfully applied in face recognition. In this paper, we proposed a maximum probability of partial ranking method based on the framework of SRC, called SRC-MP, for face recognition. Eigenfiaces, fisherfaces, 2DPCA and 2DLDA are used for feature extraction. Experiments are implemented on two public face databases, Entended Yale B and ORL. In order to show our proposed method is robust for face recognition in the real world, experiment is also implemented on a web female album (WFA) face database. We utilize AdaBoost method to automatically detect human face from web album images with complex background, illumination variation and image misalignment to construct WFA database. Furthermore, we compare our proposed method with the classical projection-based methods such as principal component analysis (PCA), linear discriminant analysis (LDA), 2DPCA and 2DLDA. The experimental results demonstrate our proposed method not only is robust for varied viewing angles, expressions, and illumination, but also has higher recognition rates than other methods.**

Keywords-Compressive sensing; Face recognition; Sparse representation classification; AdaBoost.

I. INTRODUCTION

Face recognition is a hot research area in recent years. Although many papers reported face recognition methods, researchers have focused primarily on projection-based methods rather than other methods [1]. As to the advantages of the projection-based methods, face images are reconstructed promptly and image features are extracted instantly, such as Principal Component Analysis (PCA) [2] and Linear Discriminant Analysis (LDA) [3]. Besides, the projection-based methods have achieved high recognition rates for several public face image databases. However, the disadvantage of the linear dimensionality reduction algorithms is that the projections are linear combination of all the original features. Meanwhile, all weighting coefficients in the linear combination are non-zero. Fortunately, compressive sensing theorem [4, 5, 6], a novel sampling technique, is proved to overcome the drawback. According to sparsity principle of compressive sensing, it is possible to recover certain signals and images exactly from far fewer samples of measurements beyond Nyquist rates [7]. A sparse representation-based classification (SRC) method based on compressive sensing is presented [8, 9]. It has been successfully applied in face recognition. In this paper, we propose a maximum probability of partial ranking method based on the framework of SRC, called SRC-MP, for face recognition. PCA (eigenfaces), LDA (fisherfaces), 2DPCA [10] and 2DLDA [11] are used for feature extraction. Experiments are implemented on three face databases: Extended Yale B, ORL, and web female album (WFA). The images on WFA database are obtained by using AdaBoost [12] to implement human face detection automatically from web album images in the real world. By applying our proposed method, it is robust for face recognition with varied faces in the real world, and we enable to gain the higher recognition rate than classical projection-based methods.

The rest of this paper is organized as follows: Section 2 briefly reviews SRC method [8]. Section 3 proposes our method based on the framework of SRC. Section 4 depicts experiment results and the conclusion is drawn in Section 5.

II. SPARSE REPRESENTATION BASED CLASSIFICATION

A. Compressive Sensing

Compressive sensing is a sampling technique for finding sparse solutions to underdetermined linear systems [7, 8]. A K-sparse signal is a signal that owns at most K nonzero coefficients where K << N, N is the size of signal. The compressive sensing theorem adopts the sparsity property, and is performed under the following optimization method based on l1-norm:

$$\min_{\hat{A} \in \mathbf{R}^N} \left\|\hat{A}\right\|_1 \text{ subject to } Y = \Phi^{\mathrm{T}} \hat{A} \quad (1)$$

Where

Y: an observed M-dimensional signal (M-dimensional column vector, M < N);

$\hat{A}$: an N-dimensional sparse signal (a column vector of N components);

$\left\|\hat{A}\right\|_1$: the l_1-norm of $\hat{A}$;

Φ: an $N \times M$ sensing matrix.

B. Sparse Representatin-based Calssification

An SRC method based on compressive sensing theorem is provided for face recognition [8]. The basic idea of SRC is to represent a testing image as a sparse linear combination of all training images. In order to obtain a sparse solution, the feature dimensions must be much smaller than the number of all training images.

Suppose there are K individuals in the face database, and let $B = [B_1, B_2, \dots, B_K]$ be the concatenation of the N training images from all of the K individuals, where $N = n_1+n_2+\dots+n_K$. $B_i = [s_1^{(i)}, s_2^{(i)}, \dots, s_{n_i}^{(i)}] \in \boldsymbol{R}^{m\times n_i}$, is the set of training images of the i^{th} individual, where $s_j^{(i)}$, j = 1, 2, …, n_i, is an m-dimensional vector stretched by the j^{th} image of the i^{th} individual. A new testing image $\boldsymbol{y} \in \boldsymbol{R}^m$ of the i^{th} individual could be represented as a linear combination of the training images in B_i, i.e. $\boldsymbol{y} = \sum_{j=1}^{n_i} \alpha_j^{(i)} s_j^{(i)} = B_i\,\boldsymbol{\alpha}^{(i)}$, where $\boldsymbol{\alpha}^{(i)} = [\alpha_1^{(i)}, \alpha_2^{(i)}, \dots, \alpha_{n_i}^{(i)}]^T \in \boldsymbol{R}^{n_i}$ are weighting coefficients. Let $\boldsymbol{y} = B\boldsymbol{\alpha}$ represent the testing image $\boldsymbol{y}$ by using B, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}^{(1)}; \boldsymbol{\alpha}^{(2)};\dots;\boldsymbol{\alpha}^{(K)}]$. Due to $\boldsymbol{y}$ belongs to the i^{th} individual and $\boldsymbol{y} = B_i\boldsymbol{\alpha}^{(i)}$, only the coefficients in $\boldsymbol{\alpha}^{(i)}$ have significant values in a noiseless case to $\boldsymbol{\alpha}$, and all the coefficients in $\boldsymbol{\alpha}^{(j)}$, j=1,2,…,K and $j \neq i$, are nearly zero. The SRC algorithm is listed as follows [8].

1) Normalize the columns of B to have unit l_2-norm.
2) Solve the following l_1-norm minimization problem:
$$\hat{\boldsymbol{\alpha}}_1 = \arg\min_{\alpha} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|B\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \le \varepsilon. \quad (2)$$
3) Compute the residuals
$$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - B\delta_i(\hat{\boldsymbol{\alpha}}_1)\|_2 \text{ for } i = 1,\dots,K. \quad (3)$$
4) Get output result by identity($\boldsymbol{y}$) = $arg\{min\, r_i(\boldsymbol{y})\}$.

III. MAXIMUM PROBABILITY OF PARTIAL RANKING METHOD

In the noiseless case, all the non-zero coefficients of $\hat{\boldsymbol{\alpha}}_1$ will completely be associated with the columns in B from a single individual. The testing image $\boldsymbol{y}$ can be easily assigned to the correct individual. As to the noise case, however, these non-zero weighting coefficient are not concentrated on any one individual and instead spread widely across the entire training set. $\boldsymbol{y}$ is difficult to represented as which one individual. Some classifiers are used to solve this problem. An SRC method classifies $\boldsymbol{y}$ to one individual by minimizing the residuals.

A simple, rapid method classifies $\boldsymbol{y}$ by using only the largest weighting coefficient value of $\hat{\boldsymbol{\alpha}}_1$, called SRC-LV. However, such heuristics do not harness the subspace structure associated with images in face recognition. In this paper, we propose a maximum probability of partial ranking method as a classifier, called SRC-MP. It is found by experiments that the largest weighting coefficient may not belong to the correct individual, however, the first largest weighting coefficients concentrate mostly on the correct individual.

Thus, we convert and normalize the weighting coefficient $v_j^{(i)}$ into the probability value $p_j^{(i)} = \frac{v_j^{(i)}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} v_j^{(i)}}$, where $v_j^{(i)}$ is the j^{th} non-zero coefficient greater than zero of the i^{th} individual of $\hat{\boldsymbol{\alpha}}_1$. Then, we assign a partial ranking value γ (first largest coefficients), and sum up these largest γ coefficients to obtain a new probability value for each of the individuals, respectively. Moreover, we employed the new maximum probability as the classifier. Figure 1 shows the weighting coefficients of the testing image $\boldsymbol{y}$. The green box represents the correct individual that $\boldsymbol{y}$ should be assigned, and the blue box represents the wrong individual.

By using SRC-LV classifier, the testing image $\boldsymbol{y}$ is classified to the wrong individual due to the blue box has the largest weighting coefficient. SRC-MP uses the maximum probability of partial ranking as the new weighting value. The maximum probability value of the green box is larger than that of the blue box, and $\boldsymbol{y}$ is classified to the correct individual.

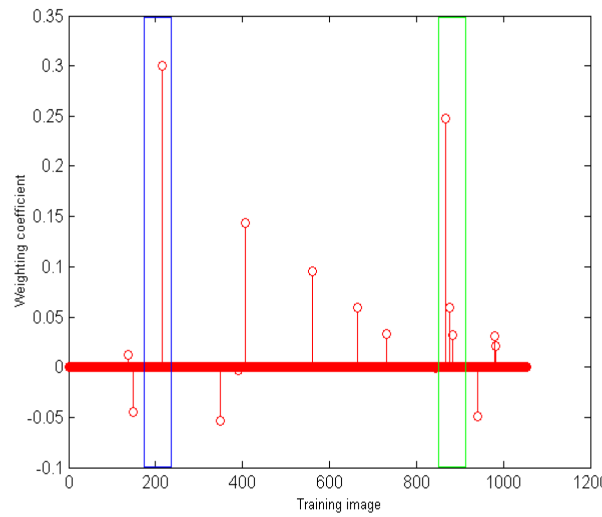


Figure 1. Weighting coefficients of the testing image $\boldsymbol{y}$. The green box indicates the correct individual that $\boldsymbol{y}$ is assigned, and blue box indicates the wrong individual.

The complete method we proposed is summarized as below,

1) Set $B_i = [s_1^{(i)}, s_2^{(i)}, \dots, s_{n_i}^{(i)}] \in \boldsymbol{R}^{m\times n_i}$ as a matrix of the training images for K individuals, and a testing image $\boldsymbol{y} \in \boldsymbol{R}^m$, as input data.

2) Solve the l_1-norm minimization problem.
$$\hat{\boldsymbol{\alpha}}_1 = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|B\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \le \varepsilon.$$

3) Compute the probability value $p_j^{(i)} = \frac{v_j^{(i)}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} v_j^{(i)}}$ for all non-zero values greater than zero.

4) Assign a partial ranking value γ, and compute new probability value for each individual $w_i(\boldsymbol{y})$, respectively.
for $k <= \gamma$, $w_i(\boldsymbol{y}) = w_i(\boldsymbol{y}) + p_k^{(i)}$ for i = 1, …, K, where $p_k^{(i)}$ is the k^{th} largest probability value that belongs to the i^{th} individual.

5) Label $\boldsymbol{y}$ by identity($\boldsymbol{y}$) = $arg\{max_i\, w_i(\boldsymbol{y})\}$.

IV. EXPERIMENTAL RESULTS

We evaluate the performance of our proposed method (SRC-MP) on Extended Yale B [13] and ORL [14] face databases, and PCA (eigenfaces), LDA (fisherfaces), 2DPCA and 2DLDA are used for feature extraction, respectively. We compare SRC-MP with classical projection-based methods such as PCA, LDA, 2DPCA and 2DLDA that adopt the nearest decision rule as the classifier. We also compare the recognition rates with different parameters γ introduced in Section 3. In order to show SRC-MP method is robust for face recognition in the real world, experiment is also implemented on WFA face database.

A. Extended Yale B Face Image Database

The Extended Yale B database has about 2,500 images of 39 different individuals. We use 34 individuals because there are some images missing. Our database consists of 2,108 face cropped and normalized images of 192 rows and 168 columns in PGM file format. There are 34 persons individually contributed 62 frontal-images by capturing under various laboratory-controlled lighting conditions. The first 10 images of individual 1 are shown in Figure 2. As for each subject, 31 images for training and the rest 31 images for testing are randomly selected.

Figure 2. The 10 face images of the 1th individual on the Extended Yale B face database.

We compute the recognition rates with the feature space dimensions d = 20, 30, 60, 120, 150, respectively. For SRC-MP method, we assign the partial ranking value γ = 10. Table I shows the recognition rates of all methods: (1) PCA, (2) Eigen + SRC-LV, (3) Eigen + SRC-MP, (4) LDA, (5) Fisher + SRC-LV and (6) Fisher + SRC-MP.

The recognition rates of SRC-LV are higher than classical projection-based methods, and SRC-MP obtains higher recognition rates than SRC-LV. In particular, the bold values indicate the best recognition rate accomplished by our proposed method. The curves of recognition rate versus the dimension of features are illustrated in Figure3.

TABLE I. THE RECOGNITION RATES (%) OF ALL METHODS ON THE EXTENDED YALE B DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 20	d = 30	d = 60	d = 120	d = 150
(1)	51.04	59.58	70.59	76.85	77.61
(2)	79.13	89.47	93.26	94.97	95.73
(3)	80.74	90.99	94.40	95.92	96.20
(4)	92.60	94.59	92.88	88.99	89.66
(5)	93.55	94.59	96.77	96.58	96.68
(6)	**94.59**	**95.16**	**97.25**	**97.25**	**97.34**

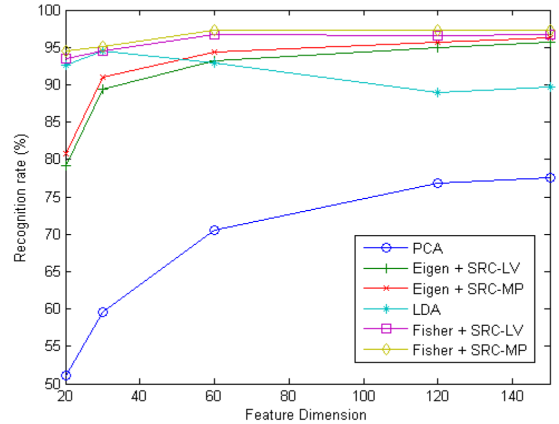


Figure 3. Recognition rates of all methods versus feature dimension on Extended Yale B database.

We also compare SRC-MP with 2DPCA and 2DLDA. Due to the feature dimension must be smaller than the number of training samples, we convert the images on Extended Yale B database into the size of 84x96. We compute the recognition rates with the feature space dimensions 96×d where d = 2, 3, 4, respectively. Table II shows the recognition rates of all methods: (1) 2DPCA, (2) 2DPCA + SRC-MP (γ = 10), (3) 2DLDA, (4) 2DLDA + SRC-MP (γ = 10). The curves of recognition rate versus the feature dimensions are illustrated in Figure 4.

TABLE II. THE RECOGNITION RATES (%) OF ALL METHODS VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 2	d = 3	d = 4
(1)	59.01	64.52	66.89
(2)	95.45	95.54	95.16
(3)	83.02	81.97	81.02
(4)	**95.83**	**95.92**	**96.02**

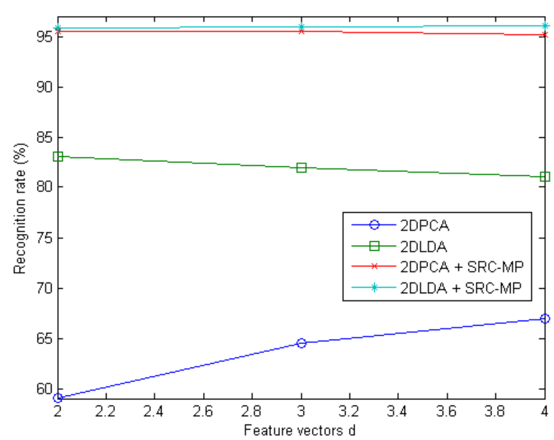


Figure 4. Recognition rates of all methods versus feature dimension on Extended Yale B database.

For SRC-MP method, a partial ranking value γ needs to be assigned to compute the new maximum probability value. We compare the recognition rates with different parameter γ values. Table III shows the recognition rates of different γ values. The curves of recognition rate versus the different γ values are illustrated in Figure 5. It shows that the larger the parameter γ, the higher the recognition rate when γ is in a certain range.

TABLE III. THE RECOGNITION RATES (%) OF DIFFERENT γ VALUES ON THE EXTENDED YALE B DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 20	d = 30	d = 60	d = 120	d = 150
$\gamma = 0$	79.13	89.46	93.26	94.97	95.73
$\gamma = 5$	80.27	90.89	94.21	95.64	96.20
$\gamma = 10$	80.74	90.99	94.40	95.92	96.20
$\gamma = 20$	80.65	90.99	94.50	96.02	96.77

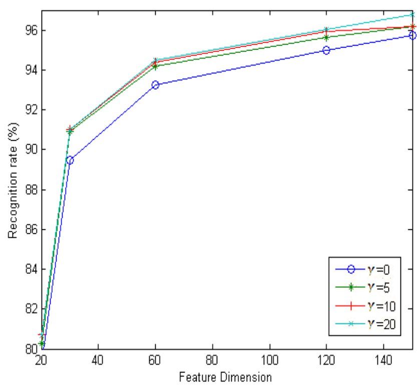


Figure 5. Recognition rates with different parameters γ on Extended Yale B database.

B. ORL Face Image Database

The ORL face database contains 400 8-bit gray level images of 112 rows and 92 columns in PGM file format. There are 40 persons individually contributed 10 images at different times, lightings, facial expressions, and some details on face. The 10 images of individual 17 are shown in Figure 6. As for each individual, the first 5 images for training and the next 5 images for testing were selected.

Figure 6. The face images of the 17^{th} individual on the ORL face database.

We compute the recognition rates with the feature space dimensions d = 16, 30, 60, respectively. Table IV shows the recognition rates of all methods: (1) PCA, (2) Eigen + SRC-MP ($\gamma = 10$), (3) LDA and (4) Fisher + SRC-MP ($\gamma = 10$). The bold values indicate the best recognition rate accomplished by our proposed method. The curves of recognition rate versus the dimension of features are illustrated in Figure 7.

TABLE IV. THE RECOGNITION RATES (%) OF ALL METHODS ON ORL DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 16	d = 30	d = 60
(1)	83.0	87.5	89.0
(2)	87.0	89.0	**90.0**
(3)	88.0	86.0	89.5
(4)	**90.0**	**91.5**	**90.0**

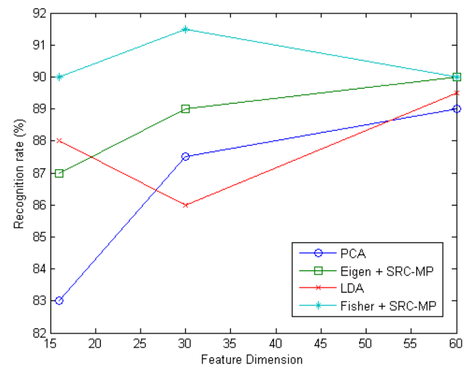


Figure 7. Recognition rates of all methods versus feature dimension on ORL database.

C. Web Female Album Face Image Database

Face detection in the real world is a difficult problem because faces are non-rigid objects and the complex background. In this paper, we collect some web female albums images with complex background, illumination variation, and image misalignment. AdaBoost method [12] is adopted to automatically detect human face on these images to construct WFA face image database. Figure 8 shows the workflow of the WFA database construction.

Figure 8. Workflow of the WFA database construction.

The WFA database contains 1320 images of 192 rows and 168 columns in JPEG file format. There are 33 persons individually contributed 40 images with varied viewing angles, expressions, and illumination. As for each individual, 20 images for training and the rest 20 images for testing are randomly selected. We compute the recognition rates with the feature space dimensions d = 12, 16, 20, 30, 40, respectively. Table V shows the recognition rates of all methods: (1) Eigen + SRC-LV, (2) Eigen + SRC-MP (γ = 10), (3) Fisher + SRC-LV and (4) Fisher + SRC-MP (γ = 10). The bold values indicate the best recognition rate accomplished by our proposed method. The curves of recognition rate versus the dimension of features are illustrated in Figure 9.

TABLE V. THE RECOGNITION RATES (%) OF ALL METHODS ON WAF DATABASE VERSUS THE CORRESPONDING FEATURE DIMENSIONS

	d = 12	d = 16	d = 20	d = 30	d = 40
(1)	59.70	69.39	76.06	78.33	81.36
(2)	**60.91**	**73.18**	**78.48**	80.15	**84.24**
(3)	56.52	64.24	71.21	74.39	76.82
(4)	59.24	68.64	75.00	**80.45**	81.67

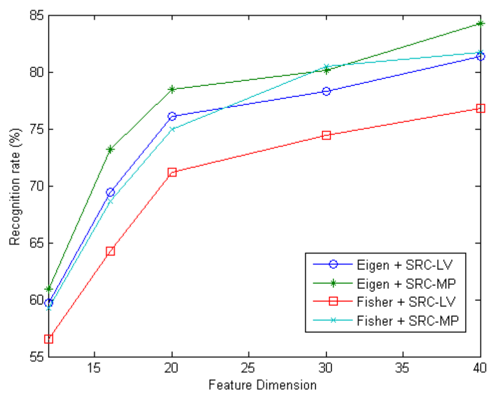


Figure 9. Recognition rates by different methods versus feature dimension on WFA database.

V. CONCLUSIONS

In this paper, we presented a maximum probability of partial ranking method (SRC-MP) based on the framework of sparse representation-based classification (SRC) for face recognition. PCA (eigenfaces), LDA (fisherfaces), 2DPCA and 2DLDA are utilized for feature extraction. We compared our proposed method with classical projection-based methodes such as PCA, LDA, 2DPCA and 2DLDA. Our experiments on Extended Yale B and ORL face databases demonstrated that our proposed method achieves higher recognition rate than classical projection-based methods under the same dimensionality. The experimental result showed that SRC-MP can obtain higher recognition rate than SRC-LV in all cases. It also demonstrated that the larger the parameter γ, the higher the recognition rate when γ is in a certain range. We also adopted AdaBoost method to automatically detect human face from web female album images in the real world to construct a web female album (WFA) face database. The experimental was implemented on WFA database for face recognition in the real world. The experimental result showed it was robust for varied viewing angles, expressions, and illumination, and enabled to achieve high recognition accuracy.

REFERENCES

[1] C.C. Chen, Y.S. Shieh, and H.T. Chu, “Face image retrieval by projection-based features,” The 3rd International Workshop on Image Media Quality and its Applications, pp. 138-144, 2008.

[2] M. Turk and A. Pentland, “Eigenfaces for Recognition,” Journal of Cognitive Neuroscience, vol.3, no.1, pp. 71-86, 1991.

[3] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers, “Fisher Discriminant Analysis with Kernels,” IEEE International Workshop on Neural Networks for Signal Processing, vol.9, pp. 41-48, 1999.

[4] E.J. Candes and M.B. Wakin, “An Introduction to Compressive Sampling,” IEEE Signal Processing Magazine, vol.5, no.2, pp. 21-30, 2008.

[5] M. Stojnic, “L1 Optimization and its Various Thresholds in Compressed Sensing,” International Conference on Acoustics, Signal and Speech Processing, pp. 3910-3913, 2010.

[6] Q. Zhang and B. Li, “Joint Sparsity Model with Matrix Completion for an Ensemble of Face Images,” IEEE International Conference on Image Processing, pp. 1665-1668. 2010.

[7] C. Moler, “”Magic” reconstruction: compressive sensing,” Cleves Corner, Mathworks News&Notes, pp. 1-4, 2010, http://www.mathworks.com

[8] J. Wright, A.Y. Yang, and A. Ganesh, “Robust Face Recognition via Sparse Representation,” IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.31, no.2, pp. 210-227, 2009.

[9] L. Zhang, M. Yang, Z. Feng, and D. Zhang, “On the Dimensionality Reduction for Sparse Representation based Face Recognition,” International Conference on Pattern Recognition, pp. 1237-1240, 2010.

[10] J. Yang, D. Zhang, A.F. Frangi, and J.Y. Yang, “Two-dimensional PCA: a new approach to appearance-based face representation and recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.26, no.1, pp. 131-137, 2004.

[11] J. Yang, D. Zhang, and J.Y. Yang, “Two-dimensional discriminant transform for face recognition,” Pattern recognition, vol.38, pp. 1125-1129, 2005.

[12] L. Ming, and Y. Baozong, “2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix,” Pattern Recognition Letters, vol.26, pp.527-532, 2005.

[13] P. Viola, and M. J. Jones, “Robust Real-Time Face Detection. Internal,” Jour. Computer Vision, vol.57, no.2, pp. 137-154, 2004.

[14] Entended Yale B face database: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html, last access on Mar. 6, 2012.

[15] ORL face database: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html, last access on Mar. 6, 2012.

Adaptive Neuro-Fuzzy Inference System for Dynamic Load Balancing in 3GPP LTE

Aderemi A. Atayero and Matthew K. Luka
Department of Electrical & Information Engineering
Covenant University
Ota, Nigeria

***Abstract*—ANFIS is applicable in modeling of key parameters when investigating the performance and functionality of wireless networks. The need to save both capital and operational expenditure in the management of wireless networks cannot be over-emphasized. Automation of network operations is a veritable means of achieving the necessary reduction in CAPEX and OPEX. To this end, next generations networks such WiMAX and 3GPP LTE and LTE-Advanced provide support for self-optimization, self-configuration and self-healing to minimize human-to-system interaction and hence reap the attendant benefits of automation. One of the most important optimization tasks is load balancing as it affects network operation right from planning through the lifespan of the network. Several methods for load balancing have been proposed. While some of them have a very buoyant theoretical basis, they are not practically implementable at the current state of technology. Furthermore, most of the techniques proposed employ iterative algorithm, which in itself is not computationally efficient. This paper proposes the use of soft computing, precisely adaptive neuro-fuzzy inference system for dynamic QoS-aware load balancing in 3GPP LTE. Three key performance indicators (i.e. number of satisfied user, virtual load and fairness distribution index) are used to adjust hysteresis task of load balancing.**

Keywords-ANFIS; 3GPP; LTE; Neural Network; Fuzzy Logic; Load balancing; Virtual load.

I. Introduction

The third generation project (3GPP) Long Term Evolution (LTE) has the core objective of meeting the increasing performance needs of mobile broadband. Some of the key features of LTE include: high spectral efficiency, very low latency, support of variable bandwidth, simple protocol architecture, and support for Self-Organizing Networks (SON) operation. SON operation was introduced to improve overall system performance through efficient operations and maintenance. Load balancing belongs to SON's self-optimizing functions, which are engineered towards reducing overall operational expenditures (OPEX) by minimizing workload for site survey, analysis of network performance and other operational and maintenance tasks that require human intervention. Generally, self-optimization involves the use of User Equipment (UE) and evolved Node B (eNB) measurements and performance measurements for network auto-tuning. The objective of load balancing is to ensure an equitable distribution of cell load among cells or to transfer part of the traffic from congested cells with the aim of improving the overall system capacity and network performance indices [1], [2].

The process of load balancing begins with detecting network load imbalance by periodically exchanging information between neighbouring eNBs (over the X2 interface) to compare the cells load. To realize an efficient intra LTE based load balancing, the load information must consist of both radio resource usage, which corresponds to the uplink and downlink physical resource block (PRB) usage as well as generic measurements representing non-radio-related resource usage. The non-radio-related load parameters include: Transport Network Load (TNL) indicator, Hardware (HW) load indicator, and available capacity for load balancing as a percentage of total cell capacity. For inter-RAT (radio access technology) load information must include another parameter known as Cell Capacity Class Value (CCCV), which is a relative capacity indicator. An algorithm to distribute the loads towards neighboring cell(s) with minimum number of cell reselection or handover is then implemented to achieve load balancing.

Several algorithms have been envisaged. In [3], a load balancing algorithm aimed at finding the optimum handover (HO) offset value between the overloaded cell and a possible target cell was implemented. Another paradigm to load balancing for LTE networks was investigated in [4]. The approach is based on a network formulation of heterogeneous services with different quality of service requirements. In [5], Wang et al. used a network utility-based load-balancing framework to develop an algorithm called Heaviest-First Load Balancing (HFLB). Another approach postulated in [6] involves the integration of another self-optimization function – handover parameter optimization to offset handover problems associated with load balancing. All of the aforementioned methods and algorithms are however based on iterative processes, which are computationally expensive. This is a serious limitation to a generalized load-balancing scheme.

In addition, since load balancing using handover is a computationally demanding task, it is desirable to divide and allocate resources between users who have data to transmit. If the desired load balancing is not achieved, then a handover is enforced. Moreover, to realize a generic load balancing, both radio resource usage and non-radio resource parameters must be incorporated. These challenges point to the need for the development of a robust, computationally less expensive and as a consequence cost effect approach. In this research work, an Adaptive Neuro-fuzzy Inference System (ANFIS) is proposed for the implementation of dynamic load balancing in LTE.

II. System Modeling and Load Metric Determination

A. Introduction

The proposed system consists of a five-layer ANFIS that takes three inputs viz.: a) SINR – the Signal to Interference Noise Ratio of the users; b) the virtual load of a cell and c) the load distribution index of the entire network. The output of the ANFIS system is a quality of service (QoS) indicator that is used to decide either scheduling or handover, in order to achieve load balancing. The Network model is based on a 3GPP downlink multi-cell network serving users with homogenous QoS requirement. Specifically, constant bit error rate (CBR) users are taken into account. Other QoS requirements can be easily added. The SINR is used as a metric measuring the link quality of the link model [7]. Performance analysis is hinged on two factors, namely: fairness distribution of the virtual load and the link Block Error Ratio (LBER).

B. Link Model

The post-equalization symbol SINR was determined from three parts of the link measurement model. These constituent models include: (i) shadow fading, (ii) macroscopic pathloss and (iii) small scale fading (for Multiple-Input-Multiple Output). The propagation pathloss due to distance and antenna gain can be modeled by the macroscopic pathloss between an eNodeB sector and a UE. The pathloss can be noted as $\mathrm{L}_{\mathrm{mp,T_i,U_j}}$ where $\mathrm{T_i}$ is the $\mathrm{i^{th}}$ transmitter (denoted as 0 for the attached eNodeB and $1, \dots, \mathrm{N}$ for the interfering eNodeBs. $\mathrm{U_j}$ is the jth UE which is located at an (x, y) position. The pathloss was generated using a distance dependent pathloss of $128.1 + 37.6\log_{10}(\mathrm{R[Km]})$ [8] and a $\theta_{3\mathrm{dB}} = 65^{\circ}/15$ dBi antenna [9].

Shadow fading occurs due to obstacles in the propagation path between the eNodeB and UE. Shadow fading can be seen as the changes in the geographical properties of the terrain associated with the mean pathloss derived from the macroscopic pathloss model. It is often approximated by a log-normal distribution of standard deviation 10 dB and mean 0 dB. A UE moving in the Region of Interest (ROI) will experience a slowly changing pathloss due to the shadow fading of the attached eNodeB being correlated with the shadow fading of the interfering eNodeBs. Shadow fading can be denoted by $\mathrm{L}_{\mathrm{sf,T_i,U_j}}$. The large scales fading (shadow fading and pathloss) are position dependent and time-invariant.

Small-scale fading results primarily due to the presence of reflectors and signal scatter agents that cause multiple versions of the transmitted signal to arrive at receiver. The small scale fading is modeled as a time dependent process for different transmission modes. One of the MIMO transmission modes is the Open Loop Spatial Multiplexing. The MIMO OSLM channel can be modeled to obtain the per-layer SINR. This transmission mode consists of a precoding for Spatial Multiplexing (SM) with large-delay Cyclic Delay Diversity (CDD) [10]. The OLSM MIMO precoding is defined by:

$$\begin{bmatrix} y_{(0)}(i) \\ \vdots \\ y_{(N_t-1)}(i) \end{bmatrix} = W(i)D(i)U \tag{1}$$

Where, $\mathrm{N_t}$ – Number of transmit antennas; v – Number of layers (a layer is a mapping of symbols to the transmit antenna); $\mathrm{W(i)} = \mathrm{N_t} \times \mathrm{v}$ – precoding matrix; D, U – $\mathrm{v} \times \mathrm{v}$ diagonal matrixes introducing the CDD.

For the MIMO OLSM, the SINR for the UE can be expressed as:

$$SINR_{c,u} = \frac{\alpha_i\, L_{sf,0,U}\, L_{pl,0,U} P_1}{\beta_i P_1 + \gamma_i \sigma^2 + \sum_1^{N_{int}} \theta_{i,1}\, L_{sf,T_i,U_j}\, L_{pl,T_i,U_j} P_1} \tag{2}$$

Where α_i and β_i model the channel estimation errors; $\mathrm{P_1} = \mathrm{P_{tx}}/\mathrm{v}$ represents the homogenously distributed transmit power; γ_i models a simple Zero Forcing (ZF) receiver noise enhancement σ^2 is the uncorrelated receiver noise and θ models the interference. $\mathrm{L}_{\mathrm{sf,T_i,U_j}}$and $\mathrm{L}_{\mathrm{pl,T_i,U_j}}\mathrm{P_1}$ stand for the shadow fading and pathloss between the UE, u and its attached eNodeB c (for $\mathrm{T_i} = 0$) and its interferers (for $\mathrm{T_i} = 1, \dots, \mathrm{N_t}$) respectively.

A given MCS (Modulation Coding Scheme) requires a certain SINR (measured at the receiver of the UE) to operate with an acceptably low BER (Bit Error Rate) in the output data. An MCS with a higher throughput needs a higher SINR to operate [11]. We assume that the best modulation coding scheme (MCS) is used for a given SINR and the highest data rate R(SINR) is achievable, this can be represented by Shannon formula as shown below:

$$R(SINR_u) = \log_2(1 + SINR_u) \tag{3}$$

For better approximation to realistic MCS, the mapping function is scaled by an attenuation factor (of say 0.75) and is bounded by the minimum required SINR of -6.5dB and a maximum bitrate of 4.8 bps/Hz.

C. Load Metric

The amount of Physical Resource Blocks (PRBs) required by user u can be expressed as:

$$N_u = \frac{D_u}{R(SINR)_u \cdot BW} \tag{4}$$

Where $\mathrm{D_u}$ – is required data rate; BW – is the transmission bandwidth of one resource block (180 kHz for LTE).

The load of cell c is thus expressed as the ratio of the sum of the required resources of all users connected to cell c to the total number of resources $\mathrm{N_t}$:

$$\rho_c = min\left(\frac{\sum_{u:X(u)=c} N_u}{N_t}, 1\right) \tag{5}$$

If we chose the number of unsatisfied users as assessment and simulation metric, then we can focus on the CBR traffic rather than the network throughput. In this case, the UEs either get exactly the CBR or they totally unsatisfied. Equation (5) implies that the cell load parameter should not exceed 1 for all users to be satisfied. This can be extended to give a general indication of how overloaded (or otherwise) a cell is, by defining a virtual load given by:

$$\widehat{\rho_C} = \frac{\sum_{u:X(u)=c} N_u}{N_t} \quad (6)$$

Where $\widehat{\rho_C} \leq 1$ implies that all users in the cell are satisfied, $\widehat{\rho_C} = U$ means $1/U$ of the users are satisfied. The total number of unsatisfied users in the whole network (with a total number of M_c users in cell c) is given by:

$$z = \sum_c max\left(0, M_c \cdot \left(1 - \frac{1}{\widehat{\rho_C}}\right)\right) \quad (7)$$

For performance analysis, the use of a fairness distribution proposed in [12] is employed. Thus, the load distribution index measuring the degree of load balancing of the entire network is given as:

$$\mu(t) = \frac{[\sum_c \rho_c(t)]^2}{|N| \cdot [\sum_c \rho_c(t)]^2} \quad (8)$$

Where $|N|$ is the number of cells in the network (used for simulation) and t is the simulation time. The load balance index $\mu(t)$ takes the value in the interval $\left[\frac{1}{|N|}, 1\right]$. A larger μ indicates a more balanced load distribution among the cells. Thus, the load distribution index is 1 when the load is completely balanced. The aim of load balancing (for CBR users) is to maximize is to maximize $\mu(t)$ at each time t.

In order to improve the load balancing performance among adjacent cells, it is necessary to find the optimum target cell. This can be achieved by adopting a two-layer inquiry scheme proposed in [13]. The source eNB (the cell requiring load balancing) request load state and environment state from all neighbouring eNBs (first layer cells). The load state is the load of the first layer cell and the environment state is the average load of the first layer cell's adjacent cells excluding the one to be adjusted (denoted as the second layer cells). The overall state of the first layer cell i is obtained by a weighted combination of the load state (LS_i) and environment state (ES_i) in one figure as follows:

$$OS_i = \alpha LS_i + (1 - \alpha) ES_i \quad (9)$$

Where the environmental state is given by:

$$ES_i = (\rho_{1i} + \rho_{2i} + \cdots + \rho_{ni})/n = \frac{\sum_{j=1}^{n} \rho_j}{n} \quad (10)$$

$LS_i = \rho_i$, the load of first layer cell i, and α is a parameter that indicates the relative contribution of LS_i and ES_i to OS_i.

OS_i gives a comprehensive load information of the first layer cell, thereby indicating whether the eNodeB can be a target cell. Taking the value of $\alpha = 0.2$ equation (9) can be expressed as:

$$OS_i = (0.2 \times \rho_i) + 0.8 \times \left(\frac{\sum_{j=1}^{n} \rho_j}{n}\right) \quad (11)$$

III. Adaptive Neuro-Fuzzy Inference System

Adaptive Neuro-Fuzzy Inference System (ANFIS) otherwise referred to as Adaptive Network-based Fuzzy inference System was proposed in [14]. ANFIS is a blend of Fuzzy Logic (FL) and Artificial Neural Network (ANN) that captures the strengths and offsets the limitations of both techniques for building Inference Systems (IS) with improved results and enhanced intelligence. Fuzzy logic is associated with the theory of fuzzy set, which relates to classes of objects with rough boundaries in which membership is a matter of degree. It is an extensive multivalued logical system that departs in concept and substance from the traditional multivalued logical systems. Much of fuzzy logic may be viewed as a platform for computing with words rather than numbers. The use of words for computing is closer to human intuition and exploits the tolerance for imprecision, thereby lowering the cost of the solution [15]. However, there are no known appropriate or well-established methods of defining rules and membership functions based on human knowledge and experience. Artificial Neural Networks are made up of simple processing elements operating concurrently. These elements model the biological nervous system, with the network functions predominantly determined by the connections between the elements. Neural Networks have the ability to learn from data by adjusting the values of the connections (weights) between the elements. Merging these two artificial intelligence paradigms together offers the learning power of neural networks and the knowledge representation of fuzzy logic for making inferences from observations.

A. *Basic ANFIS Architecture*

The ANFIS architecture described here is based on type 3 fuzzy inference system (other popular types are the type 1 and type 2). In the type 3 inference system, the Takagi and Sugeno's (TKS) if-then rules are used [16]. The output of each rule is obtained by adding a constant term to the linear combination of the input variables. Final output is then computed by taking the weighted average of each rule's output. The type 3 ANFIS architecture with two inputs (x and y) and one output, z, is shown in figure1.

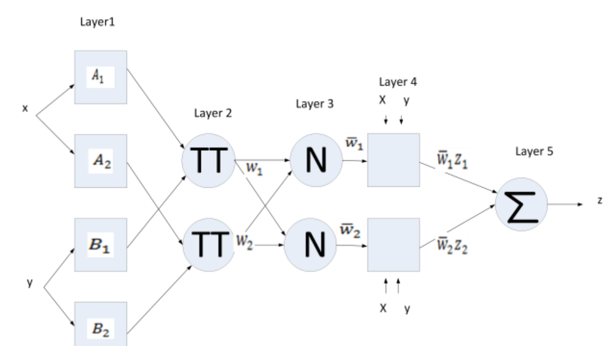


Figure 1. Type 3 ANFIS Architecture

Assuming the rule base contains two first order TKS if-then rules as follows:

$$Rule\ 1: if\ x\ is\ A_1\ and\ y\ is\ B_1, then\ z_1 = p_1x + q_1y + r_1$$

$$Rule\ 2: if\ x\ is\ A_2\ and\ y\ is\ B_2, then\ z_2 = p_2x + q_2y + r_2$$

The ANFIS structure is the functional equivalent of a supervised, feed-forward neural network with one input layer, three hidden layers and one output layer, whose functionality are as described below:

Layer 1: Every node in this layer is an adaptive layer that generates the membership grades of the input vectors. Usually, a bell-shaped (Gaussian) function with maximum equal to 1 and minimum equal to 0 is used for implementing the node function:

$$O_i^1 = \mu_{A_i}(x) = \frac{1}{1 + |(x - c_i)/a_i|^{2b_i}} \qquad (12)$$

Where O_i^1 is the output of the ith node in the first layer, $\mu_{A_i}(x)$ is the membership function of input x in the linguistic variable A_i. The parameter set $\{a_i, b_i, c_i\}$ are responsible for are responsible for defining the shapes of the membership functions. These parameters are called premise parameters.

Layer 2: Each mode in this layer determines the firing strength of a rule by multiplying the membership functions associated with the rules. The nodes in this layer are fixed in nature. The firing strength of a particular rule (the output of a node) is given by:

$$w = O_i^2 = \mu_{A_i}(x) \cdot \mu_{B_i}(y), i = 1, 2, \ldots \qquad (13)$$

Any other T-norm operator that performs fuzzy AND operation can be used in this layer.

Layer 3: This layer consists of fixed nodes that are used to compute the ratio of the ith rule's firing strength to the total of all firing strengths:

$$\overline{w} = O_i^3 = \frac{w_i}{w_1 + w_2}, i = 1, 2, \ldots \qquad (14)$$

The outputs of this layer are otherwise known as normalized firing strength for convenience.

Layer 4: This is an adaptive layer with node function given by:

$$\overline{w}_i z_i = O_i^4 = \overline{w}_i(p_i x + q_i y + r_i) \qquad (15)$$

This layer essentially computes the contribution of each rule to the overall output. It is defuzzification layer and provides output values resulting from the inference of rules. The parameters in this layer $\{p_i, q_i, r_i\}$ are known as consequent parameters.

Layer 5: There is only one fixed node in this layer. It computes the overall output as the summation of contribution from each rule:

$$\sum_i \overline{w}_i z_i = O_i^5 = \sum_i \frac{w_i z_i}{\sum_i z_i} \qquad (16)$$

B. Hybrid Learning Algorithm

The objective of learning is to tune all the adjustable parameters to make the ANFIS output match the desired data. In order to improve the training efficiency, a combination of learning algorithms is adopted to adjust the parameters of the input and output membership functions. The consequent parameters are optimized using the least square method with the antecedent parameters fixed. After updating the consequent parameters, the gradient descent method using back-propagation training algorithm is used to fine-tune the premise parameters. Assuming the premise parameters are held fixed, then the overall output of the ANFIS will be a linear combination of the consequent outputs given by:

$$\begin{aligned} z &= \overline{w}_1 z_1 + \overline{w}_2 z_2 \\ &= \overline{w}_1(p_1x + q_1y + r_1) + \overline{w}_2(p_2x + q_2y + r_2) \\ &= (\overline{w}_1 x)p_1 + (\overline{w}_1 y)q_1 + (\overline{w}_1)r_1 + (\overline{w}_2 x)p_2 + (\overline{w}_2 y)q_2 \\ &+ r_2 \end{aligned} \qquad (17)$$

IV. Design Of Load Balancing Inference Scheme

In the first stage, the crisp variables, the virtual load of the source cell, the load fairness distribution index and number of unsatisfied users are converted into fuzzy (linguistic) variables in the fuzzification process. The fuzzification maps the three input variables to fuzzy labels of the fuzzy sets. Each linguistic variable has a corresponding membership function. A sigmoidal membership function (precisely, the product of two sigmoidal function) was used in this work. As there are three inputs and 4 fuzzified variables, the inference system has a set of 64 rules (figure 2).

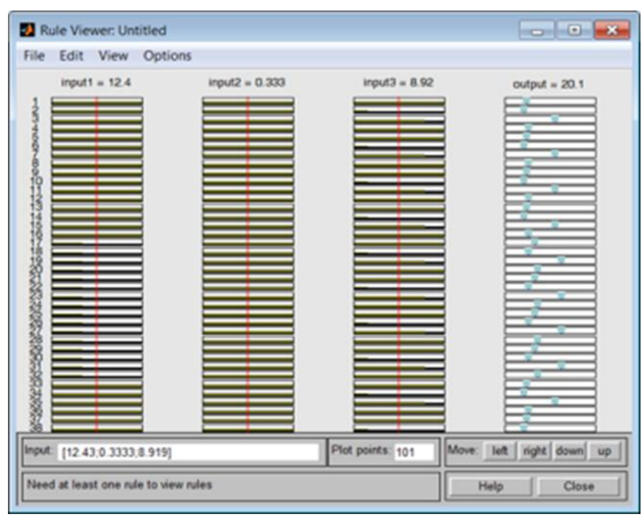


Figure 2. Rule Viewer for the inference system

The neural network training helps select the appropriate rule to be fired.

Next, the rules are de-fuzzified to produce quantifiable results. Defuzzification can be achieved using several techniques such as maximum methods, center of gravity

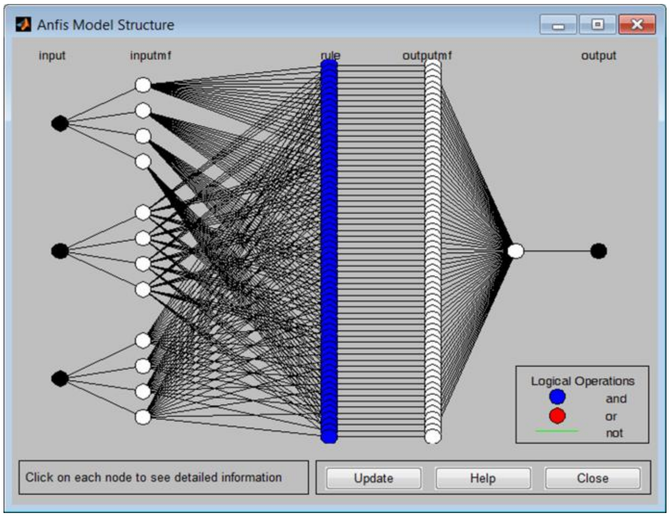


Figure 3. ANFIS structure for Proposed Dynamic Load Balancing

method, center of singleton method etc. The center of gravity method is adopted for this work. The defuzzified output is further is then used to schedule resources or handover users to achieve a dynamic load balancing. The structure of the Model used is depicted in figure 3.

The model consists of 158 nodes, 64 rules, 256 linear parameters and 48 nonlinear parameters. The total number of parameter is very important in deciding the number of training data pairs required. In order to realize a good generalization capability, it is recommended to have the number of training data points to be many times larger than the number of parameters being evaluated [15]. 1326 input/output pairs of training data was used for training. Thus, the ratio between the data points and parameters is about four times (1326/304).

For parameter optimization, hybrid training (which combines least mean squares and back-propagation) was used. To ascertain how well the training data models the load balancing system, model validation was incorporated. Model validation involves presenting input/output data sets on which the inference system not trained to the inference system to check the degree to which the inference system model predicts the corresponding data set outputs values. For this work, model validation was accomplished using a checking data set of 1326 input/output pairs. The checking data helps prevent the potential of model over-fitting of the data. This is accomplished by selecting model parameters that correspond to the minimum checking data model error. Fig.4 show training data tested on the checking data.

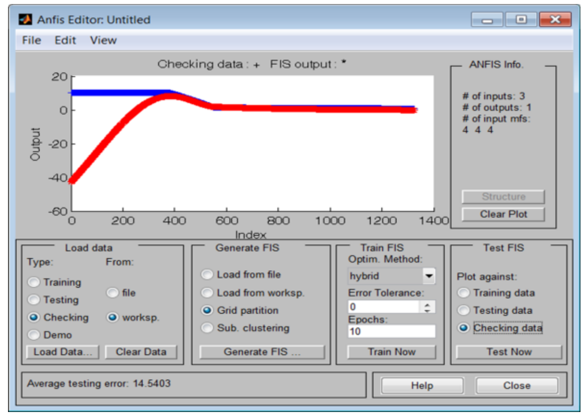


Figure 4. Checking data tested on training data

V. Results And Discussion

The ANFIS system uses the hysteresis value for a QoS aware dynamic load balancing. The inference system increases the hysteresis as the virtual load of the cell increases. The virtual load has an overriding effect over the fairness distribution index in determining the result of the inference. When these two factors are the predominant input metrics, the relationship is illustrated in figure 5.

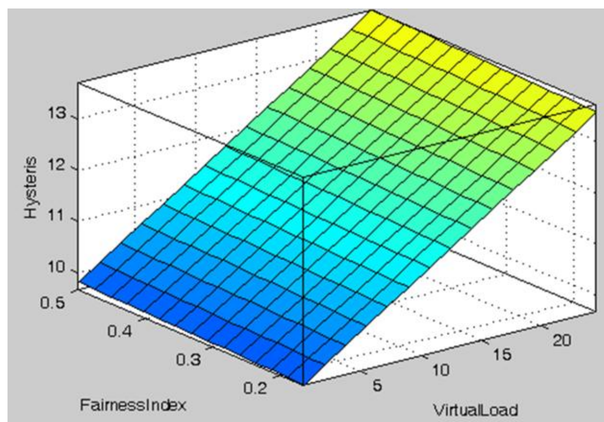


Figure 5. Hysterisis as a function of Fairness Index and Virtual Load

As the number of satisfied users increases, hysteresis value decreases (figure 6). Conversely, when the number of unsatisfied users in the network increases, the hysteresis value also increases to trigger. This results in sustaining or triggering load-balancing process.

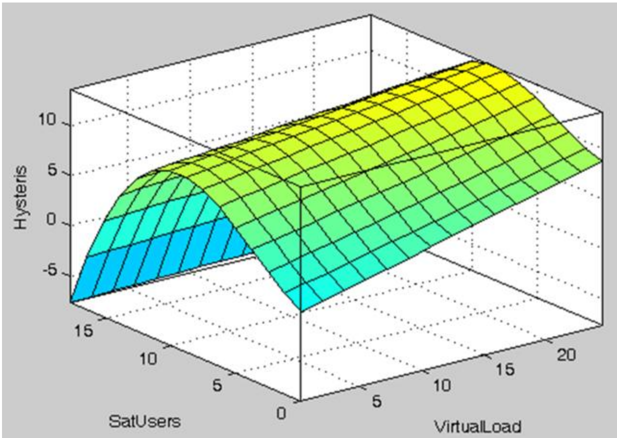


Figure 6. Effect of using virtual load and number of satisfied users for ANFIS load balancing.

Figure 7 shows the effect of fairness index and the number of satisfied users in determining the value. The result reveals that the numbers of satisfied users have a more domineering effect over fairness index in determining load balancing.

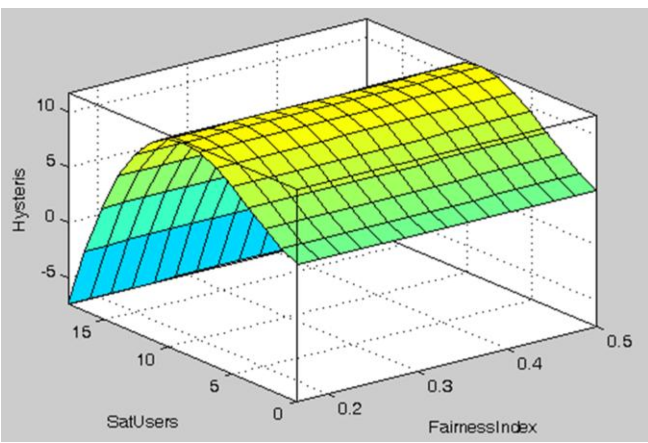


Figure 7. Effect of using fairness index and number of satisfied users for ANFIS load balancing

VI. CONCLUSION

In summary, we have presented in this paper three key performance indicators for consideration in LTE dynamic load balancing i.e. the number of satisfied (dissatisfied) users, the fairness index and the virtual load of the source are three key performance indicators that can be used for dynamic load balancing in LTE and by extension to all SONs. This becomes especially important in the consideration of different network architectures [17]. As seen from the results, the number of satisfied (dissatisfied) users plays a more dominant role as the key performance indicator (KPI) especially where QoS is a major consideration. The virtual load of the cell is the next most important key performance indicator for fine-tuning the load balancing decision. Although the Fairness index did not reflect well in comparison to the other KPIs, it is also important especially where the load balancing in the network can be achieved by a more even (fairer) distribution of resource to users. In other words, the fairness index can be used as the KPI for deciding scheduling, while the virtual load and number of satisfied/unsatisfied users can be employed for handovers when and where necessary, thereby achieving a dynamic QoS-aware load balancing.

REFERENCES

[1] 3GPP TS 36.300 version, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2", Technical Specification, version 10.4.0 Release 10, June 2011

[2] S. Sesia, I. Toufik and M. Baker, "LTE – The UMTS Long Term Evolution: From Theory to Practice", 1st edition, John Wiley & Sons, Ltd. , West Sussex, UK, 2009

[3] A. Lobinger et al., "Load Balancing in Downlink LTE Self-Optimizing Networks", IEEE 71st VTC 2010, Taipei, Taiwan, June 2010.

[4] H. Wang et al, "Dynamic Load Balancing in 3GPP LTE Multi-Cell Networks with Heterogenous services", ICST Conference, Beijing, 2010

[5] H. Wang, "Dynamic Load Balancing and Throughput Optimization in 3GPP LTE Networks", IWCMC 2010, Caen, France, July, 2010

[6] 3GPP TS 36.201 V9.1.0 (2010-03), "LTE Physical Layer: General Description."

[7] WINNER Technical Report, "Assessment of advanced beamforming and MIMO technologies," WINNER, Tech. Rep. IST-2003-507581, 2005

[8] Technical Specification Group RAN, "E-UTRA; LTE RF system scenarios," 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 36.942, 2008-2009.

[9] Technical Specification Group RAN, "Physical layer aspects for E-UTRA," 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 25.814, 2006.

[10] Technical Specification Group RAN, "E-UTRA; physical channels and modulation," 3rd Generation Partnership Project (3GPP), Tech. Rep. TS 36.211 Version 8.7.0, May 2009.

[11] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (EUTRA), "Radio Frequency (RF) system scenarios" (Release 8), Technical Report TR 36.942, available at http://www.3gpp.org.

[12] R. Jain, D.M Chiu and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Systems", Technical Report, Digital Equipment Corporation, DEC-TR-301, 1984.

[13] L. Zhang, Y. Liu, M. Zhang, S. Jia, and X. Duan, " A Two-layer Mobility Load Balancing in LTE Self-Organization Networks" IEEE Internal Conference on Communication Technology, Beijing, China, 2011, pg. 925 – 929.

[14] Jyh-Shing Roger Jang, "ANFIS: Adaptive Network-Based Fuzzy Inference System", IEEE trans., on Systems, Man and Cybernetics, Vol. 23, No. 3, May-June, 1993, pg. 665-685.

[15] Mathwork Inc., "Fuzzy Logic ToolboxTM User Guide ", Ver., 2.2.14, Sep., 2011. Available from www.mathworks.com

[16] T. Takagi and M. Sugeno, "Fuzzy Identification of Systems and its application to modelling and control," IEEE trans. on Systems, Man and Cybernetics, Vol. 15, pp.116-132, 1985

[17] A.A. Atayero, M.K. Luka, M.K. Orya, J.O. Iruemi "3GPP Long Term Evolution: Architecture, Protocols and Interfaces", International Journal of Information and Communication Technology Research (IJICTR), Vol.1, №7, pp. 306–310, Nov. 2011.

Estimation of soil moisture in paddy field using Artificial Neural Networks

Chusnul Arif
Dept. of Civil and Environmental Engineering
Bogor Agricultural University
Bogor, Indonesia

Masaru Mizoguchi
Dept. of Global Agricultural Sciences
The University of Tokyo
Tokyo, Japan

Budi Indra Setiawan
Dept. of Civil and Environmental Engineering
Bogor Agricultural University
Bogor, Indonesia

Ryoichi Doi
Dept. of Global Agricultural Sciences
The University of Tokyo
Tokyo, Japan

***Abstract*—In paddy field, monitoring soil moisture is required for irrigation scheduling and water resource allocation, management and planning. The current study proposes an Artificial Neural Networks (ANN) model to estimate soil moisture in paddy field with limited meteorological data. Dynamic of ANN model was adopted to estimate soil moisture with the inputs of reference evapotranspiration (ET_o) and precipitation. ET_o was firstly estimated using the maximum, average and minimum values of air temperature as the inputs of model. The models were performed under different weather conditions between the two paddy cultivation periods. Training process of model was carried out using the observation data in the first period, while validation process was conducted based on the observation data in the second period. Dynamic of ANN model estimated soil moisture with R^2 values of 0.80 and 0.73 for training and validation processes, respectively, indicated that tight linear correlations between observed and estimated values of soil moisture were observed. Thus, the ANN model reliably estimates soil moisture with limited meteorological data.**

Keywords – soil moisture; paddy field; estimation method; artificial neural networks

I. INTRODUCTION

In paddy field, soil moisture represents water availability for the plants and it is required for irrigation scheduling and water resource allocation, management and planning. Soil moisture variation affects the pattern of evapotranspiration, runoff and deep percolation in paddy field [1, 2, 3]. At the same time, soil moisture level is predominately influenced by water input through precipitation and irrigation.

However, hydrological data such as crop evapotranspiration, deep percolation, runoff and irrigation are often limited because acquisition of measurements in the field is costly, complicated, and time consuming. In addition, the detailed meteorological data required to determine crop evapotranspiration are not often available especially in developing countries. Therefore, estimation of soil moisture is needed using limited meteorological data.

Artificial Neural Networks (ANN) is suitable for use in dealing with complex system, such as agricultural system, than that mathematical method [4]. ANN has the capability to recognize and learn the underlying relations between input and output without explicit physical consideration [5]. The benefits of using ANN are its massively parallel distributed structure and the ability to learn and then generalize the problem [6]. In agricultural field, ANN has been applied to classify irrigation planning strategies [7] and to estimate subsurface wetting for drip irrigation [8].

The objective of this study was to estimate soil moisture from limited meteorological data using ANN model. Then, the proposed ANN model was validated by comparing observed and estimated values of soil moisture.

II. MATERIALS AND METHODS

A. Field Measurements

The field experiments were conducted in the experimental paddy field in the Nusantara Organics SRI Center (NOSC), Sukabumi, West Java, Indonesia during two paddy cultivation periods. The SRI center is located at 06º50’43” S and 106º48’20” E, at an altitude of 536 m above mean sea level. In both cultivation periods, meteorological data, consisted of air temperature and precipitation, were measured every 30 minutes. For validation model, soil moisture was measured using 5-TE sensor (Decagon Devices, Inc., USA) every 30 minutes.

TABLE I. PADDY CULTIVATIVATION PERIODS OF THE CURRENT STUDY

Period	Planting date	Harvesting date	Season
First cultivation	14 October 2010	8 February 2011	Wet
Second cultivation	20 August 2011	15 December 2011	Dry - Wet

All data were sent automatically to the server through quasi real-time monitoring [9, 10]. Daily maximum, average and minimum values of air temperature were used to estimate reference evapotranspiration (ET_o), then estimated ET_o and precipitation were used to estimate soil moisture.

B. Development of Artificial Neural Networks (ANN) models

Two ANN models were developed consisted of three layers, i.e. input, hidden and output layers, respectively (Fig.1).

The first model was developed to estimate ET_o according to maximum, average and minimum values of air temperature (Fig.1a). ET_o is a key variable in hydrological studies especially to quantitative knowledge of water supply, loss, and consumption in paddy field. Commonly, minimum meteorological data are required to determine ET_o consisted of solar radiation and air temperature data using Hargreaves model [11]. However, solar radiation data are not often available, thus we only used air temperature data for ET_o estimation. For validation, the output of model was validated by comparing to ET_o derived by Hargreaves model [12].

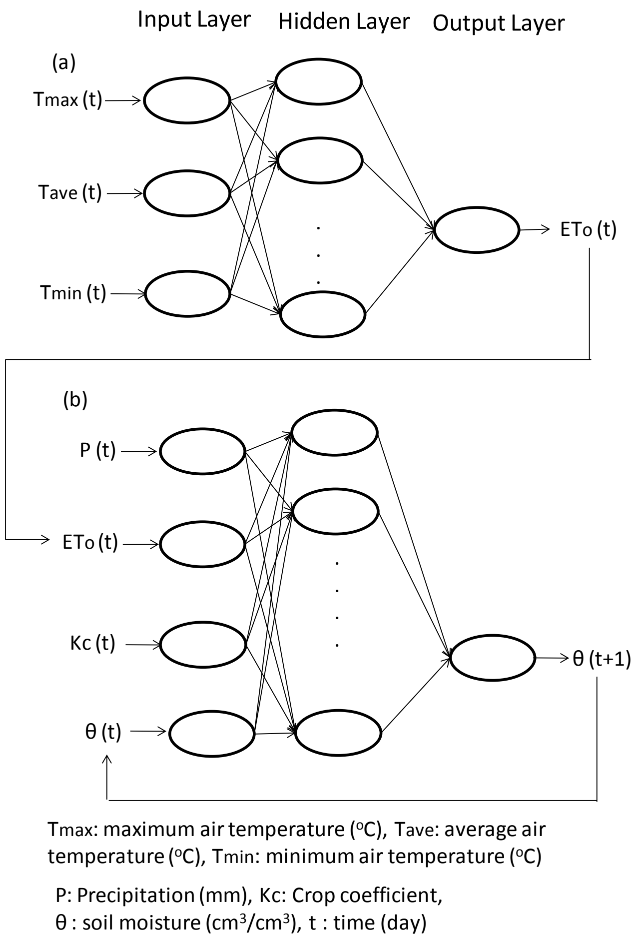


Figure 1. ANN models of the current study: a) ETo estimation, b) soil moisture estimation in paddy field.

Then, estimated ET_o and precipitation data were used to estimate soil moisture (Fig. 1b). Soil moisture is usually determined by performing water balance analysis [1, 13]. The inflow of water balance consisted of precipitation and irrigation, while outflow consisted of crop evapotranspiration, runoff and deep percolation.

However, the variables such as irrigation, runoff and deep percolation were not available because the typical of measurements are complicated and expensive tools, heavy labor and time consumption. Meanwhile, crop evapotranspiration is derived by multiplying ET_o to crop coefficient determined by the FAO procedure [14], thus crop coefficient was selected as the third inputs of ANN model (Fig.1b).

For this estimation, dynamic of ANN, similar to an auto-regressive moving average (ARMA) model procedure, was adopted by considering historical output data [15, 16]. All input parameters data were normalized between 0 and 1 by using fixed minimum and maximum values. Normalization of data is important to avoid larger numbers from overriding smaller ones and to prevent premature saturation of hidden nodes, which impedes the learning process [5]. For hidden layer, total nodes of 8 were selected as moderate number to prevent excessively time-consuming in training process and to avoid incapability in differentiating between complex patterns due to few number of node in hidden layer.

Back propagation was selected as the learning method, which is composed of two phases; first, propagation (forward and backward propagation), and second, weight update. A sigmoid function was selected as the activation by the following equations:

$$f(y) = \frac{1}{1 + e^{-gy}} \tag{1}$$

$$y = \sum_{i=0}^{n} x_i w_i \tag{2}$$

where x_i, w_i, n, g are the inputs, weights, number of inputs and gain parameter, respectively.

The gain parameters are usually set to 1.0 and not changed by the learning rule. However, this fixed value probably caused local minima problem. Therefore, the gain parameter should be adjusted according to the degree of approximation to the desired output of the output layer [17]. Here, gain parameter is adjusted based on the following condition:

$$g = \begin{cases} \frac{1}{Ap}\, if\ \mathrm{Ap} > 1.0 \\ 1.0\ if\, \mathrm{Ap} \leq 1.0 \end{cases} \tag{3}$$

$$Ap = 2e_p \tag{4}$$

$$e_p = \max\left|t_p - o_p\right| \tag{5}$$

where Ap is degree of the output layer, e_p is error for pattern *p*, t_p and o_p are observed and estimated output for pattern *p*, respectively.

The observed data were divided into two data sets. The observed data from first paddy cultivation period were used for training process, while data from the second period were used for validation process. This kind of validation is called as a cross validation method [18, 19]. Then, developed ANN model was evaluated by comparing observed and estimated values of soil moisture using the indicator of coefficient determination (R^2). The value of R^2 ranged from 0.0 to 1.0 with higher values indicating better agreement.

III. RESULTS AND DISCUSSION

A. Weather conditions during paddy cultivation periods

Meteorological conditions in the first and second paddy cultivation periods are shown in Fig. 2-3. In the first paddy cultivation period, the meteorological parameters were characterized by low air temperature and high precipitation compared to the second period. The monthly average air temperature changed during in the end of 2010 and 2011, where its value was highest on November 2010 for the first period, and then it occurred on December 2011 for the second period. The highest average air temperature values were 23.87°C and 24.49°C for first and second periods, respectively. The same situation occurred to precipitation in which highest total precipitation was 400.8 mm on December 2010 for the first period and its value was 369.6 mm on November 2011 for the second period.

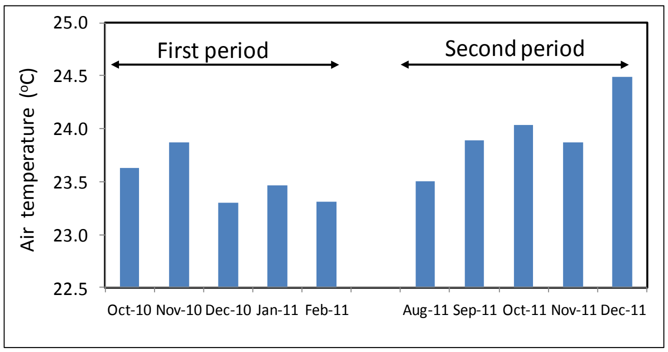


Figure 2. Monthly average air temperature during cultivation periods.

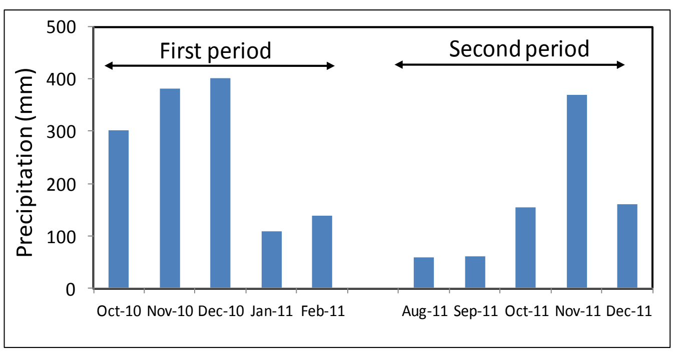


Figure 3. Total precipitation during cultivation periods.

B. Estimation of reference evapotranspiration

Training process should be carried out firstly by ANN model to learn the pattern of observation data between the input and the output. In this process, a thousand iterations were performed to minimize the error of estimation. As the result, ANN model estimated ET_o with R^2 of 0.96 as shown in Fig. 4a. Then, the weights, results of training process, were used to estimate ET_o using the second period data with the result as shown in Fig. 4b. Underestimation was occurred when ET_o value was higher than 4.5 mm.

However, with R^2 of 0.95, the estimated ET_o showed good agreement to the Hargreaves model. Therefore, ANN model can be used to estimate ET_o using the inputs of maximum, average and minimum values of air temperature.

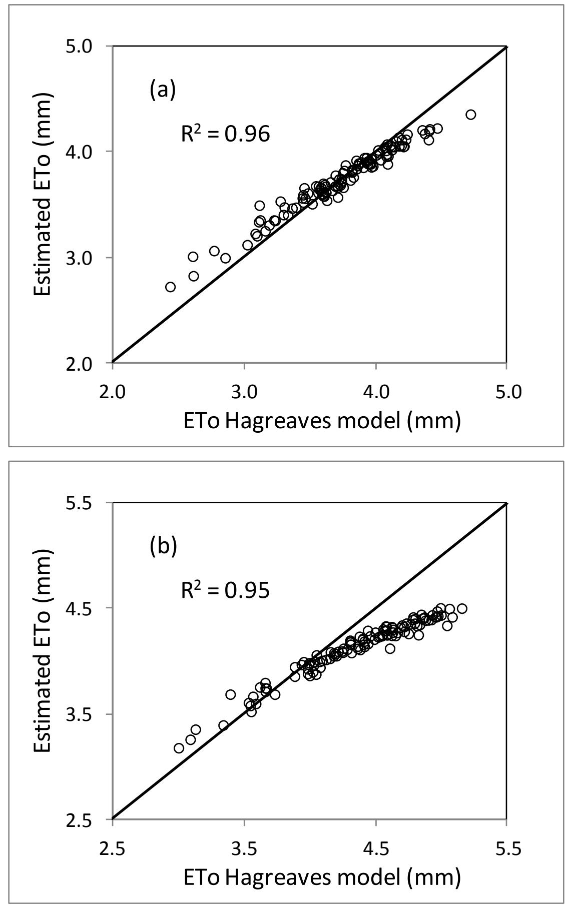


Figure 4. Model evaluation of ANN model for ETo estimation: a) training process, b) validation process.

C. Estimation of soil moisture

Dynamic of ANN model estimated soil moisture with R^2 of 0.80 and 0.73 for both training and validation processes, respectively (Fig. 5). In the training process, tight linear correlations between observed and estimated values of soil moisture were observed (Fig. 5a), thus more than 80% of the changes in observed soil moisture were well described by the model. Therefore, the weights, representation of relationship between the input and the output model, can be used to estimate soil moisture in paddy field with limited meteorological data.

Fig. 5b shows the correlation between observed and estimated values of soil moisture in the second cultivation period as the result of validation process. ANN model estimated soil moisture with R^2 values of greater than 0.72 indicate the model's performance.

Thus, tight linear correlations between observed and estimated values of soil moisture by the ANN model (Fig. 5b). Accordingly, the estimation results in this study showed good agreement to the observed data, thus the proposed method can be accepted as suggested by the previous study [1].

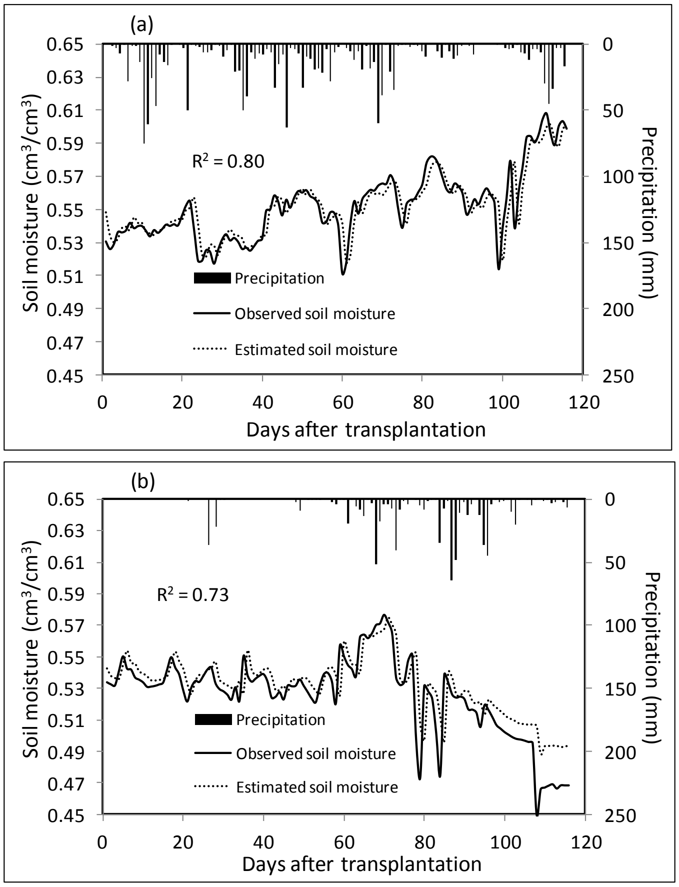


Figure 5. Observed and estimated soil moisture during cultivation periods: a) first period, b) second period.

IV. CONCLUSIONS

The current study proposed Artificial Neural Networks (ANN) models to estimate soil moisture in paddy field with limited meteorological data. Two ANN models were proposed, first model was developed to estimate reference evapotranspiration (ET_o) using the maximum, average and minimum values of air temperature as the inputs. Then, estimated ET_o and precipitation were used to estimate soil moisture by adopting dynamic of ANN in the second model.

The proposed models were performed under different weather conditions between the two paddy cultivation periods. In the first model, ET_o was estimated accurately with R^2 of 0.96 and 0.95 for training and validation processes, respectively. Then, the second model estimated soil moisture with R^2 values of greater than 0.72 for training and validation processes suggested that the ANN model was reliably to estimate soil moisture in paddy field with limited meteorological data and without complicated and expensive tools, heavy labor and time consumption.

ACKNOWLEDGMENT

The authors are grateful to the Directorate of Higher Education, Ministry of National Education, Republic of Indonesia for generous financial support through grant of International Research Collaboration and Scientific Publication. Also, the authors acknowledge the financial support by the Japan Society for the Promotion of Science.

REFERENCES

[1] H. K. Kim, T. I. Jang, S. J. Im, and S. W. Park, "Estimation of irrigation return flow paddy fields considering the soil moisture". Agr Water Manage vol. 96, pp. 875-882, 2009.

[2] T. V. Reshmidevi, R. Jana, and T. I. Eldho, "Geospatial estimation of soil moisture in rain-fed paddy fields using SCS-CN-based model". Agr Water Manage vol. 95, no. 4, pp. 447-457, 2008.

[3] Y. H. Li, and Y. L. Cui, "Real-time forecasting of irrigation water requirements of paddy fields". Agr Water Manage vol. 31, no. 3, pp. 185-193, 1996.

[4] Y. Hashimoto, "Application artificial neural network and genetic algorithms to agricultural systems". Computer and Electronics in Agriculture vol. 18, pp. 71-72, 1997.

[5] I. A. Basheer, and M. Harmeer, "Artificial neural networks: fundamentals, computing, design, and application". Journal of Microbiological Methods vol. 43, pp. 3–31, 2000.

[6] A. S. Nugroho, "Information analysis using soft computing - the applications to character recognition, meterological prediction and bioinformatics problem". PhD Disertation, Nagoya Institute of Technology, Nagoya, Japan, pp. 12, 2003.

[7] S. K. Raju, D. N. Kumar, and L. Duck, "Artificial neural networks and multicriterion analysis for sustainable irrigation planning". Computers & Operations Research vol. 33, pp.1138–1153, 2006.

[8] A. C. Hinnell, N. Lazarovitch, A. Furman, M. Poulton, and A. W. Warrick, "Neuro-Drip: estimation of subsurface wetting patterns for drip irrigation using neural networks". Irri. Sci. vol. 28, no. 6, pp. 535-544, 2009.

[9] M. Mizoguchi, T. Ito, C. Arif, S. Mitsuishi, and M. Akazawa, "Quasi Real-Time Field Network System for Monitoring Remote Agricultural Fields". Proc. SICE Annual Conference 2011, pp. 1586-1589, 2011.

[10] C. Arif, B. I. Setiawan, M. Mizoguchi, R. Doi, S. K. Saptomo, Ardiansyah, and T. Ito, "Field Network System to Monitor Paddy Fields in the System of Rice Intensification in Indonesia". Proc. CIGR 2011 conference on Sustainable Bioproduction, Tokyo, September 2011.

[11] J. I. L. Moreno, T. M. Hess, and S. M. White, "Estimation of reference evapotranspiration in a mountainous mediterranean site using the Penman-Monteith equation with limited meteorological data", Pirineos, vol. 164, pp. 7-31, 2009.

[12] G. H. Hargreaves, and R. G. Allen, "History and evaluation of Hargreaves evapotranspiration equation", Journal of Irrigation and Drainage Engineering, vol. 129, pp. 53-63, 2003.

[13] T. V. Reshmidevi, R. Jana, and T. I. Eldho, "Geospatial estimation of soil moistre in rain-fed paddy fields using SCS-CN-based model", Agr Water Manage, vol. 95, pp. 447-457, 2008.

[14] R. G. Allen, L. S. Pareira, D. Raes, and M. Smith, "Crop Evapotranspiration Guidelines for computing crop water requirements", FAO - Food and Agriculture Organization of the United Nations, Rome, 1998.

[15] S. Chen, S. A. Billings, P. M. Grant, "Non-linear system identification using neural network". Int. J. Control, vol. 51, no. 6, pp. 1191-1214, 1990.

[16] T. Morimoto, Y. Ouchi, M. Shimizu, M. S. Baloch, "Dynamic optimization of watering Satsuma mandarin using neural networks and genetic algorithms". Agr Water Manage, vol. 93: 1-10, 2007.

[17] X. G. Wang, Z. Tang, H. Tamura, M. Ishii, and W. D. Sun, "An improved backpropagation algorithm to avoid the local minima problem". Neurocomputing, vol. 56, pp. 455-460, 2004.

[18] T. Morimoto, and Y. Hashimoto, "AI approaches to identification and control of total plant production systems". Control Engineering Practice, vol. 8, pp. 555-567, 2000.

[19] H. Suhardiyanto, C. Arif, B. I. Setiawan, "Optimization of EC Values of Nutrient Solution for Tomato Fruits Quality in Hydroponics System Using Artificial Neural Network and Genetic Algorithms". ITB Journal Science, vol. 41A, no. 1, pp. 38-49, 2009.

AUTHORS PROFILE

Chusnul Arif, lecturer at Department of Civil and Environmental Engineering, Bogor Agricultural University, Indonesia. He is interested study on the application of Artificial Intelligence such as Artifical Neural Networks and Genetic Algorithms in agricultural field particularly in paddy field. Since April 2010, he is pursuing PhD degree at Department of Global Agricultural Sciences, the Univeristy of Tokyo with the main topic is optimization of water management in paddy field. This paper is part of his PhD dissertation.

Budi Indra Setiawan, Professor of soil physic and hydrology at Department of Civil and Environmental Engineering, Bogor Agricultural University, Indonesia. His common interest is to develop theory, model and technology which have direct benefits for the improvement of water resources management in agricultures as well as aquacultures.

Masaru Mizoguchi, Professor of International agro-informatics at Department of Global Agricultural Sciences, the University of Tokyo, Japan. He is developing quasi real-time monitoring of farmland for any purpose and study, such as climate change and plant growth model.

Ryoichi Doi, Assistant professor at Department of Global Agricultural Sciences, the University of Tokyo, Japan, has been working interdisciplinary in areas of appropriate technology, environment and agriculture related to socio-economic development.

Dynamic Decision Support System Based on Bayesian Networks

Application to fight against the Nosocomial Infections

Hela Ltifi - Ghada Trabelsi - Mounir Ben Ayed - Adel M. Alimi
REGIM: REsearch Group on Intelligent Machines,
University of Sfax, National School of Engineers (ENIS),
BP 1173, Sfax, 3038, Tunisia

***Abstract*—The improvement of medical care quality is a significant interest for the future years. The fight against nosocomial infections (NI) in the intensive care units (ICU) is a good example. We will focus on a set of observations which reflect the dynamic aspect of the decision, result of the application of a Medical Decision Support System (MDSS). This system has to make dynamic decision on temporal data. We use dynamic Bayesian network (DBN) to model this dynamic process. It is a temporal reasoning within a real-time environment; we are interested in the Dynamic Decision Support Systems in healthcare domain (MDDSS).**

Keywords- Dynamic Decision Support Systems; Nosocomial Infection; Bayesian Network.

I. Introduction

The questions that interest health scientists become increasingly complex. For many questions, we need much time of analysis to generate significant quantities of complex temporal data that describe the interrelated histories of people and groups of people [8]. In Intensive Care Units (ICUs), physicians focus on the continuously evolution of patients. The temporal dimension plays a critical role in understanding the patients' state.

The development of methods for the acquisition, modeling and reasoning is, therefore, useful to exploit the large amount of temporal data recorded daily in the ICU. In this context, a Medical Decision Support System (MDSS) can be developed to help physicians to better understand the patient's temporal evolution in the ICU and thus to take decisions.

In many cases, the MDSS deals with the decision problem according to its knowledge; some of this knowledge can be extracted using a decision support tool which is the Knowledge Discovery from Databases (KDD) [10] [14]. The goal of the KDD is to extract knowledge and to interpret, evaluate and put it as a valid element of decision support.

The MDSS is well applied particularly to the prediction and shows significantly positive results in practice [7]. The control of the Nosocomial1 infections (NI) is regarded as a promising research field in the ICU [16]. These infections are contracted during the hospitalization.

[1] The term "nosocomial" comes from the Greek word "nosokomeion" to indicate the hospital

From this point of view, a KDD-based MDSS aims at helping the physicians, users of the system, to especially understand and prevent the NI. The MDSS for the fight against NI require temporal data analysis. The dynamic aspect of the decisions is related to the measurements recorded periodically such as the infectious examinations, the antibiotic prescribed before admission, etc.

The objective is to daily predict the probability of acquiring a NI in order to daily follow-up the patient state using a KDD technique. With this intention, the data base must be pre-treated and transformed for a temporal data mining. The data mining technique must take into account the dynamic aspect of the decision. For this reason, we choose the Dynamic Bayesian Network (DBN) [9] [35] which are models representing uncertain knowledge on complex phenomena within the framework of a dynamic process. It is a question of obtaining knowledge models which evolve with time.

This article is organized into five sections. In the second, we will present the theoretical background of our decisional context. In section 3, we will concentrate on our problematic which is the fight against the nosocomial infections. We will also discuss the dynamic aspect of the decision. In section 4, we will describe, the use of the Dynamic Bayesian Networks as a KDD technique for supporting the dynamic medical decision-making. Concerning section 5, we will expose some results obtained by the application of the DBN for fight against NI. Finally, a conclusion and several perspectives will be proposed.

II. KDD-Based MDSS: Solution Exploited For The Medical Field

The decision is often regarded as a situation of choice where several solutions are possible; among them one is "the best" [34]. To decide is to choose in a reasonable way an appropriate alternative; it is a question of making a decision during a complete process [40]. Decision support systems play an increasingly significant role in medical practice. While helping the physicians or other professionals of the medical field to make clinical decisions, the MDSS exert a growing influence on the process of care for improved health care [30]. Their impact should be intensified because of our increasing capacity to treat more data effectively [21].

The MDSS can help the physicians to organize, store, and extract medical knowledge in order to make decisions. This can decrease the medical costs by providing a more specific and more rapid diagnosis, by a more effective treatment of the drugs prescriptions, and by reducing the need for specialists' consultations [32]. Within this framework, we are interested in the MDSS allowing controlling the NI which constitute a significant challenge of modern medicine and which are considered as one of the most precise indicators of the care quality of the patients [11].

In medical decision making, Knowledge Discovery from Databases (KDD) [10] [14] is critical. In fact, knowledge, which is hidden in patient records, is valuable to provide precise medical decisions such as the diagnosis and the treatments. Indeed, traditional tools of decision support (OLAP, Info-center, dashboard, ERP, etc.) leave the initiative to the user to choose the elements which he/she wants to observe or analyze. However, in the case of KDD, the system often takes the initiative to discover associations between data. It is then possible, in a certain manner, to predict the future, according to the past.

The KDD is an interactive and iterative process aiming at extracting new, useful, and valid knowledge from a mass of data. It proceeds in four phases [10] [20] (Fig. 1):

1) Selection of the data having a relationship with the analysis requested in the base;
2) Cleaning of the data in order to correct the inaccuracies or data errors and transformation of the data into a format which prepares them for mining;
3) The data mining, application of one or more techniques (neural networks, bayesian networks, decision tree, etc.) to extract the interesting patterns. A variety of KDD techniques were developed in the last few years and applied to the medical field; and
4) Evaluation of the result allowing estimating the quality of the discovered model. Once knowledge is extracted, it is a question of integrating it by setting up the model or its results in the decisional system.

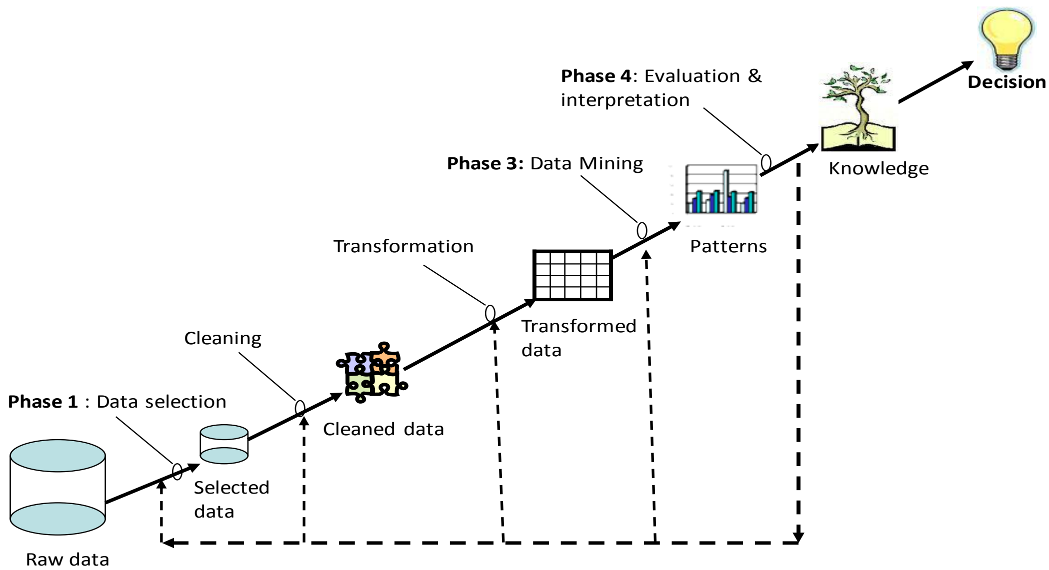


Figure 1. KDD process

Various research tasks previously applied the assistance to the medical decision-making based on the KDD for the fight against NI [3] [4], there is no study (as far as we know) which addressed the dynamic aspect of the medical decision in this context.

III. DBN-Based MDDSS: Solution Exploited For The NI Control

A. Dynamic context

This article lies within the scope of a project aiming at fighting against NI[2] in the Intensive Care Unit in the Teaching hospital Habib Bourguiba in Sfax, Tunisia [1] [23] [24] [27] [26] [41]. Some work proposed NI control systems based on the KDD techniques [3] [4]. This Work shows their effectiveness and their capacity to produce useful rules. But, their direct use by doctors appears difficult to us. A study on a prevalence of NI occurrence in the Teaching hospital Habib Bourguiba in Sfax, Tunisia, showed that 17,9 % of the hospitalized patients were victims of a NI during 24 hours [16].

The decision problematic on the patient state must envisage and prevent the NI occurrence. The risks of this infection can weaken the patient or delay his cure. The risk of infection is mainly conditioned by the fragility of the patient and the ICU techniques used for its survival. Our objective is to predict the NI occurrence each day during the hospitalization period.

The dynamic aspect is observed on various levels of decision-making [38]. It is indispensable to take into account a set of critical factors of decision which are identified by the assistance in particular interviews with some of the ICU physicians. The identification of the factors supporting the appearance of the infections is a very significant stage which

[2] An infection is typically regarded as nosocomial if it appears 48 hours or more after hospital admission

influences the results of the decision-making. These factors are classified into two categories:

1) *Static data:* patient admission data (age, gender, weight, entry and exit dates, antecedents), the SAPS II[3] score ([5] [31] proved that this score measured in the first 24 hours of the intensive care is an indicator of NI risk) and the Apache categorization[4] [18]. These data can help to determine the patients' fragility to the nosocomial infections.

2) *Temporal data:* control measurements to take each day (intubation, Central Venous Catheter (C.V.C.) [33], the urinary probe [13], Infectious examinations [16] [37] [42] and the catch of antibiotics.

At each day i ($1 \le i \le$ hospitalization duration), the decision on the patient state depends on the NI probability pi and thus on the values of the factors (static and temporal data) described above to the current day but also to the previous days, as well as to all the knowledge obtained by learning in time and recording former events. In fact, a basic decision is taken at the admission of the patient (t0). The future decision refers to a decision to be made after the consequences of a basic decision become (partially) known. A future decision is linked to the basic decision because the alternatives that will be available in the future depend on the choice made in the current basic decision. As time moves on, the future decision at current stage (t) becomes the basic decision at the next decision stage (t+1), when a new knowledge extracted by data mining (probability of acquiring a NI) and future decision should be addressed. This link repeats itself as long as the patient is hospitalized (cf. Fig. 2). The learn-then-decide-then-learn pattern describes how the decision-maker responds to new knowledge gained during the decision-making process. The elements described above, especially the existence of linked decisions, clearly show that decision-making in NI control is a dynamic process. In this scope, the decision-making process requires the consideration in time of linked or interdependent decisions, or decisions that influence each other. This dynamic decision-making pattern is a chain of decide, then learn; decide, then learn more; and so on. Such a system is so called Medical Dynamic Decision Support System (MDDSS).

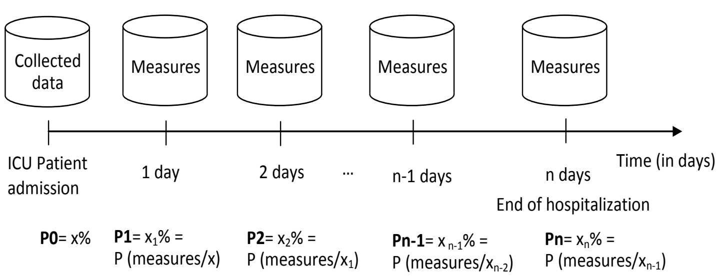


Figure 2. Temporal factors for the NI prevention

The MDDSS aims at the daily estimation of the NI occurrence probability, in percentage, during the ICU patient hospitalization. This probability is calculated using a KDD technique. It is the content of the following section.

B. Dynamic KDD technique

Because of their capacity to represent uncertain knowledge, Bayesian networks (BN) play an increasingly important role in many medical applications. They have been introduced in the 1980s as a formalism of representation and reasoning with models of problems involving uncertainty and adopting probability theory as a basic framework. Research to explore the use of this formalism in the context of medical decision making started in the 1990s [36] [29].

The medical literature contains many examples of the BN use. We can quote a BN model developed to assist clinicians in the diagnosis and selection of antibiotic treatment for patients with pneumonia in the ICU [28]. Burnside and al. [6] proposed the use of BN to predict Breast Cancer Risk.

A BN is a Graphical model (marriage between probability theory and graph theory). It is a graph with probabilities for representing random variables and their dependencies. It efficiently encodes the joint probability distribution (JPD) of a set of variables. Its nodes represent random variables and its arcs represent dependencies between random variables with conditional probabilities. It is a directed acyclic graph (DAG) so that all edges are directed and there is no cycle when edge directions are followed [15] [19].

The joint probability distribution of random variables $S = \{X_1, \ldots, X_N\}$ in a Bayesian network is calculated by the multiplication of the local conditional probabilities of all the nodes. Let a node X_i in S denote the random variable X_i, and let $Pa(X_i)$ denote the parent nodes of X_i. Then, the joint probability distribution of $S = \{X_1, \ldots, X_N\}$ is given by (1):

$$P(X_1, X_2, \ldots, X_N\} = \prod_{i=1}^{N} p(X_i \mid Pa(X_i)) \qquad (1)$$

Unfortunately, a problem with the BN is that there is no mechanism for representing temporal relations between and within the random variables. For this reason, to represent variables that change over time, it is possible to use Dynamic Bayesian Networks (DBNs) [9] [35].

DBN encodes the joint probability distribution of a time-evolving set of variables $X[t] = \{X_1[t], \ldots, X_N[t]\}$. If we consider T time slices of variables, the dynamic Bayesian network can be considered as a "static" Bayesian network with $T \times N$ variables. Using the factorization property of Bayesian networks [9] [35], the joint probability density of $X_T = \{X[1], \ldots, X[T]\}$ can be written as (2):

$$P(X[1], \ldots, X[N]\} = \prod_{t=1}^{T} \prod_{i=1}^{N} p(Xi[t] \mid Pa(X_i[t])) \text{ Where } Pa(X_i[t]) \text{ denotes the parents of } X_i[t] \qquad (2)$$

DBNs are a generalization of Kalman Filter Models (KFM) and Hidden Markov Models (HMM). In the case of (HMM), the hidden state space can be represented in a factored form

[3] Simplified Acute Physiology Score II: is used to evaluate and compare the gravity of the patients to the intensive care. It is about a predictive model of mortality of the patients.

[4] classification of the previous patient state that is statistically related to the appearance of NI

instead of a single discrete variable. Usually dynamic Bayesian networks are defined using the assumption that X[t] is a first order Markov process [35] [39].

In the context of the NI prevention, the DBN technique uses fixed and temporal variables presented in the following section.

IV. DBN Application For The Data Mining

A. DBN variables

Our study concerns the application of the DBN technique on a temporal medical data base containing 280 patients' data.

The data acquisition and selection is the first KDD-based MDDSS phase. It concerns the implementation of the temporal data base that consists on a large collection of time series. It is a succession of couples < $(v_1,t_1),(v_2,t_2)\ldots,(v_i,t_i),\ldots$ > where v_i is a value or a vector of values taken at a moment t_i. The values v_i of a sequence are often real numbers [12]. In our context, the time series are a set of daily sequentially recorded values. The data pretreatment allows applying scripts to prepare useful variables for the knowledge extraction: (1) fixed data having only one value during the hospitalization period of a patient; and (2) temporal data having a value for each time serie (day) during the hospitalization period.

The estimation of the NI occurrence probability of the patient is represented by the following variables (table 1):

TABLE I. Variables Of Bayesian Networks

Fixed variables	
Code	***Wording***
Sex	Patient gender
age1	Patient age
Periode_entr	Indicates the entry season in ICU
Orig	Origin
Detorig	Origin details
priseAnti	Antibiotic catch
Knaus	Apache categorization of the previous patient state.
Cissue	Issue : the patient is dead or survived
Diag	Diagnosis
Ant	Antecedent
Result	Static NI prediction probability
Temporal variables	
Code	***Wording***
dsj	Difference between ICU admission and exit dates
acti	Act carried out at the day i
cissuei	Issue
examinfi	Infectious examinations at the day i
sensi	Sensibility to the germ (causing the Infectious examinations at the day i) to the prescribed antibiotic
resulti	dynamic NI prediction probability at day i

The theory of the Bayesian Networks allows us to represent relationships between these observed variables in a probabilistic way which is well adapted to the uncertainty inherent to medical questions.

B. Construction of knowledge model on fixed data (Static BN)

Causal links between the fixed variables are represented on figure 3. However observations made on a one static BN for a patient are not sufficient to estimate the NI occurrence probability.

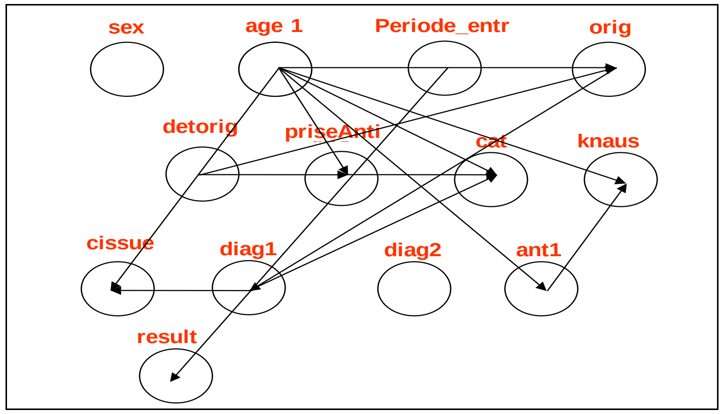


Figure 3. Causal links in static Bayesian network (static extracted knowledge model)

The extracted model could detect relations between logical variables like the relation between the age and the antecedent, between the age and cissue (the patient deceased or is survived). However obtained graph, contains "illogical" links between the nodes (for example, the age acts on the antibiotic catch). We also noted missing links which present interesting independence relations (For example, the relation between result and cissue).

The probabilities are calculated using $P(V_i|C)$ with:

- Vi : the node (sex, age1, periode_entr… diag1) having discrete values, and
- C: the class to be predicted (cissue and result) having Boolean values (yes/no): the patient catches a NI or not

We obtain a static Bayesian Network: a causal graph with the probabilities associated to each node. The use of the probabilities and the causal graph provide knowledge models which are not very rich. So, experiments made with this BN showed that the prediction was instable and could produce false alerts. In order to represent the influence of past events over the present state of the patient, it is necessary to extend this model into a dynamic BN.

C. Construction of knowledge model on temporal data (DBN)

The Figure 4 shows a dynamic extracted model based on temporal variables. The causal graph represents the interdependence between the temporal variables. We used for this dynamic structure the values of each time serie ($act_1\ldots act_{10}$, $exinf_1\ldots exinf_{30}$) [5] connected directly with the two predictive nodes which are the result and issue.

[5] In our context, we have 10 acts and 30 infectious examinations carried out daily to the patients.

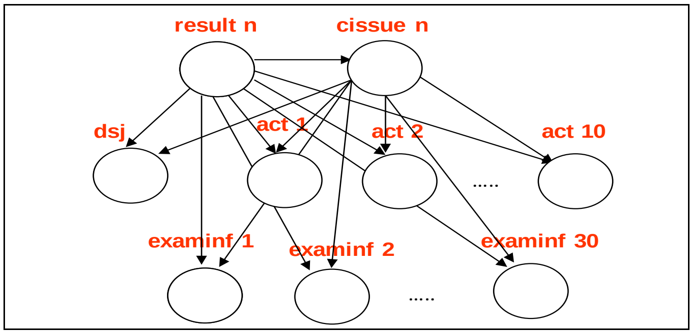


Figure 4. Causal links in Dynamic Bayesian Network (extracted model for t=n)

The principle of our Dynamic Bayesian Network can be defined by:

- At t=0, we use extracted static knowledge model (figure 3)
- For $1 \leq t \leq T$ (patient hospitalization duration): unrolling the extracted temporal knowledge models (figure 4).

We obtain a final Dynamic Bayesian Network that has the following causal graph (figure 5).

The distribution result of the joined probabilities is given by (3):

$$P(result_{1:T}) = \prod_{t=1}^{T}\prod_{i=1}^{N} P(result_t^i | Pa(result_t^i)) \quad (3)$$

With:

- T is the interval of hospitalization time,
- N is the total number of the variables for each extracted model.

The DBN application gives good prediction results presented in the next section.

V. PREDICTION RESULTS

This section presents the prediction results of an experimentation conducted over more than one year in the ICU of the teaching hospital Habib Bourguiba in Sfax, Tunisia.

After having generated many bases of examples, we applied our algorithm to real data coming from the ICU. We could extract knowledge models and transform them automatically to obtain probabilistic, quantitative and qualitative prediction results. These prediction results of our system are reliable to 74%, which is very encouraging.

Indeed, our study relates to the prediction of the patient state. This prediction is dynamic; it evolves throughout the patient hospitalization by new measurements.

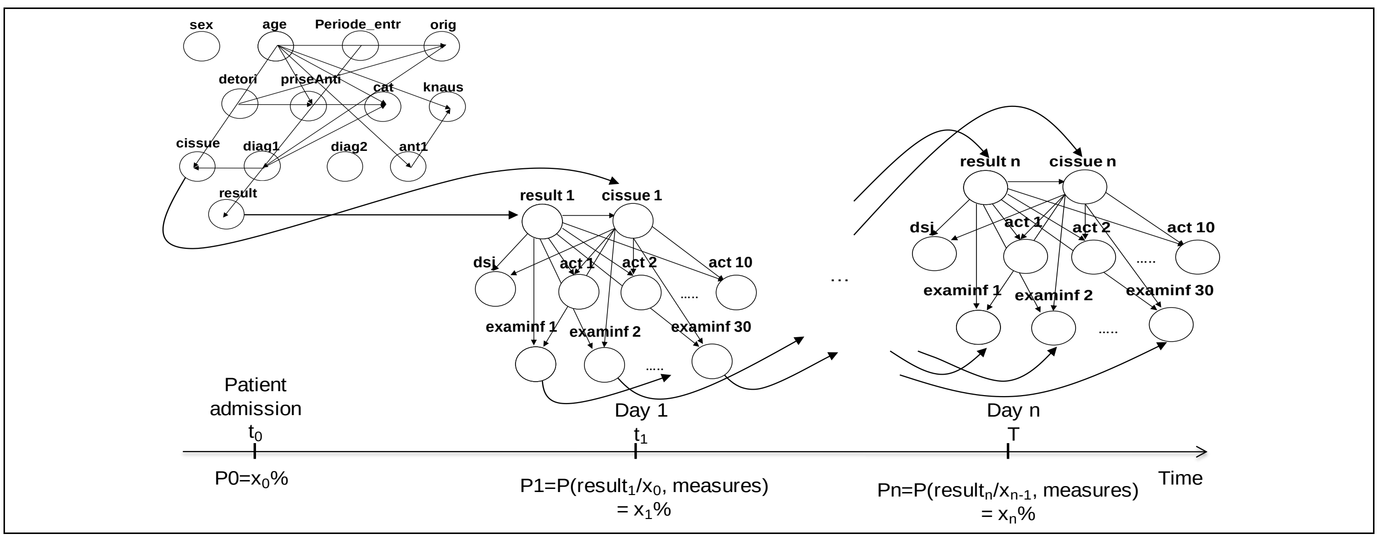


Figure 5. The causal graph of the Dynamic Bayesian Network

With each dayi of the patient hospitalization, we could envisage his state at the future by a probability, which will be used, in the prediction of the dayi+1, with these measured observations.

We used a base of test which contains 58 cases (patients), for the performance evaluation of the system. We obtained the results given by the matrix of confusion[6] represented by the table 2.

TABLE II. THE CONFUSION MATRIX OF THE RESULTS PROVIDED BY THE DYNAMIC BAYESIAN NETWORK

		Predicted	
		Negative	***Positive***
Actual	***Negative***	34	7
	Positive	8	9

We calculated the rates of evaluation starting from the prediction results obtained by our structure elaborated by the DBN. We found that the classification rate was correct to 0.74, the positive capacity of prediction = 0.56 and the negative

[6] Yes : to have a NI - No : not to have a NI - Total : the total of the predictions

capacity of prediction = 0.81. The generated observed vs. predicted results given by the table 2 are represented by the histogram (cf. Fig. 6).

An extension of the prediction phase could then be improved. With our current system, the prediction is made offline i.e. daily after the acquisition of all the data collected during the last 24 hours of hospitalization for a patient. This prediction can be improved so that it is carried out at each observation detected by our system and at every moment.

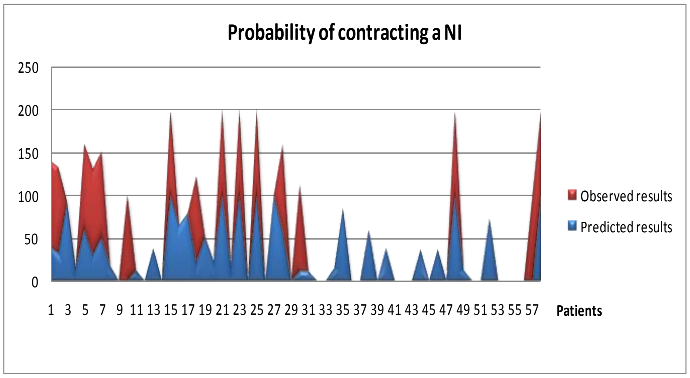


Figure 6. Prediction results

VI. CONCLUSION

In this paper, we described an application of decision support system to the hospitalized patients in the ICU. This system aims at helping the physicians to estimate the NI appearance. The decision given by this system is dynamic because it is based on the patient state described in terms of a set of temporal factors of which the unit of time is the day. The dynamic decision system evolves and proceeds in several stages corresponding to the increasing levels of the patient situation comprehension (scale of time). On each level, a set of knowledge can be generated.

In this study we used the KDD as a decisional tool. A data pre-treatment is used in order to transform medical data into standardized data usable by the system. The KDD technique used is the Dynamic Bayesian Networks (DBN). It is used for the modeling of complex systems when the situations are dubious and/or the data are of complex structure. In our case, the complexity of the data is due to the fact that they are temporal and not regular.

We have implemented the dynamic BNs based on fixed (at t=0 that gives a static BN) and temporal data (daily taken measurements during the hospitalization stay). The application of the developed models for the NI prediction gives good results.

VII. FUTURE WORK

Under the angle of the Human-Computer Interaction (HCI) and basing on this experiment in the medical field, our research perspectives are related to the design and the evaluation of a MDDSS based on a KDD process. We are confronted to the need to develop a specific methodology for the design and the evaluation of DSS based on the KDD while taking starting point the criteria, methods and techniques resulting jointly from the HCI field [2] [22] and the visualization field [25]. This last technology makes it possible to present the data and knowledge in a visual form making it possible to the user to interpret the data, to draw the conclusions as well as to interact directly with these data. It is considered that the visualization techniques can improve the current KDD techniques by increasing the implication of the user and his confidence in connection with the observations discovered [17]. Such a methodology of evaluation must allow the study of cognitive and emotional experience of the DSS users for the fight against the nosocomial infections [43].

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program. Thanks also to all the ICU staff of Habib Bourguiba Teaching Hospital for their interest to the project and all the time spent to help us design, use and evaluate our system.

REFERENCES

[1] M. Ben Ayed, H. Ltifi, C. Kolski, M.A. Alimi, "A User-centered Approach for the Design and Implementation of KDD-based DSS: A case Study in the Healthcare Domain", Decision Support Systems, vol. 50 no. 1, pp. 64-78, 2010.

[2] S. Bernonville, C. Kolski, N. Leroy, M. Beuscart-Zéphir, "Integrating the SE and HCI models in the human factors engineering cycle for re-engineering Computerized Physician Order Entry systems for medications: basic principles illustrated by a case study", International Journal of Medical Informatics, vol. 79, pp. 35-42, 2010.

[3] S.E. Brossette, A.P. Sprague, J.M. Hardin, Waites K.B., Jones W.T., Moser S.A., "Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance", Journal of the American Medical Informatics Association, vol. 5 no. 4, pp. 373-381, 1998.

[4] S.E. Brossette, A.P. Sprague, W.T. Jones, S.A. Moser, "A data mining system for infection control surveillance", Methods of Information in Medicine, vol. 39, pp. 303-310, 2000.

[5] A. Bueno-Cavanillas, Rodriguez-Contreras, A. Lopez-Luque, M. Delgado-Rodriguez, R. Galves-Vargas, "Usefulness of severity indices in intensive care medicine as a predictor of nosocomial infection risk", Intensive Care Medicine, vol. 17, pp. 336-339, 1991.

[6] E.S. Burnside, D.L. Rubin, J.P. Fine, R.D. Shachter, G.A. Sisney, W.K. Leung, "Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience", Radiology, vol. 240, no. 3, pp. 666-673, 2006.

[7] B. Chaudhry, J. Wang, S. Wu, et al., "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care", Annals of Internal Medicine, vol. 144, pp. 742-52, 2006.

[8] S.J. Clark, "A general temporal data model and the structured population event history register". Demographic research, vol. 15 no. 7, pp. 181-252, 2006.

[9] A. Darwich, "Constant-space reasoning in dynamic Bayesian networks", International journal of approximate reasoning, vol. 26, pp. 161-178, 2001.

[10] U.M. Fayyad, S.G. Djorgovski, N. Weir, "Automating the Analysis and Cataloging of Sky Surveys", Advances in Knowledge Discovery and Data Mining, MIT Press, pp. 471-494, 1996.

[11] R.P. Gaynes, T.C. Horan, "Surveillance of nosocomial infections", In: Mayhall CG, editor, Hospital Epidemiology and Infection Control, 2nd edition, Baltimore, Md: Williams & Wilkins, pp. 1285-1317, 1999.

[12] L. Ge, M.C.M. Mourits, R.B.M. Huirne, "Towards flexible decision support in the control of animal epidemics", Scientific and Technical Review of the OIE, vol. 26 no. 3, pp. 551-563, 2007.

[13] R.W. Haley, D.H. Culver, J.W. White, et al., "The nationwide nosocomial infection rate, A new need for vital statistics", American Journal of Epidemiology, vol. 121, pp. 159-167, 1985.

[14] D. Hand, H. Mannila, P. Smyth, Principles of Data Mining, MIT Press, Cambridge, 2001.

[15] F. Jensen, An Introduction to Bayesian Networks, Springer, Verlag, New York, 1996.

[16] H. Kallel, M. Bouaziz, H. Ksibi, H. Chelly, C.B. Hmida, A. Chaari, N. Rekik, M. Bouaziz, “Prevalence of hospital-acquired infection in a Tunisian Hospital”, Journal of Hospital Infection, vol. 59, pp. 343-347, 2005.

[17] D. Keim, “Information Visualization and Visual Data Mining”, IEEE Transactions on Visualization and Computer Graphics, vol. 8 no. 1, pp. 1-8, 2002.

[18] W.A. Knaus, J.E. Zimmerman, D.P. Wagner, E.A. Draper, D.E. Lawrence, “APACHE: acute physiology and chronic health evaluation: a physiologically based classification system”, Critical Care Medicine, vol. 9, pp. 591-597, 1981.

[19] L. Lauritzen, G. Cowell, P.A. Dawid, D.J. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer-Verlag New York, Inc., 1999.

[20] R. Lefébure, G. Venturini, Data Mining: Gestion de la relation client, Personnalisation des sites Web, Eds Eyrolles, Paris, 2001.

[21] G. Leroy, H. Chen, “Introduction to the special issue on decision support in medicine”, Decision Support Systems, vol. 43 no. 4, pp. 1203-1206, 2007.

[22] S. Lepreux, A. Abed, C. Kolski, “A human-centered methodology applied to decision support system design and evaluation in a railway network context”, Cognition Technology Work, vol. 5, pp. 248-271, 2003.

[23] H. Ltifi, C. Kolski, M. Ben Ayed, M.A. Alimi, “Human-centered design approach applied to Medical Dynamic DSS”, The 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, 31 August – 3 September, Valenciennes, France, 2010a.

[24] H. Ltifi, M. Ben Ayed, C. Kolski, M.A. Alimi, “Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d’ECD, application dans le domaine de la santé”, Journal d'Interaction Personne-Système (JIPS), vol. 1 no. 1, pp. 1-25, 2010b.

[25] H. Ltifi, M. Ben Ayed, S. Lepreux, M.A. Alimi, “Survey of Information Visualization Techniques for Exploitation in KDD”. IEEE AICCSA’09, Morocco, pp. 218-225, 2009a.

[26] H. Ltifi, M. Ben Ayed, C. Kolski, M.A. Alimi, “HCI-enriched approach for DSS development: the UP/U approach”, IEEE ISCC’09, Sousse, Tunisia, pp. 895-900, 2009b.

[27] H. Ltifi, M. Ben Ayed, C. Kolski, M.A. Alimi, “Prise en compte de l'utilisateur pour la conception d'un SIAD basé sur un processus ECD”, ERGO'IA 2008, Biarritz, pp. 85-92, 2008.

[28] P.J.F. Lucas, N.C. de Bruijn, K. Schurink, I.M. Hoepelman, “A Probabilistic and decision-theoretic approach to the management of infectious disease at the ICU”, Artificial Intelligence in Medicine, vol. 19 no. 3, pp. 251–279, 2000.

[29] P. Lucas, “Bayesian networks in medicine: A model-based approach to medical decision making”, Proceedings of the EUNITE Workshop on Intelligent Systems in Patient Care, Vienna, pp. 73-97, 2001.

[30] Y.A. Lussier, R. Williams, J. Li, S. Jalan, T. Borlawsky, E. Stern, I. Kohli, “Partitioning knowledge bases between advanced notification and clinical decision support systems”, Decision Support Systems, vol. 43 no. 4, pp. 1274-1286, 2007.

[31] O. Lortholary, J.Y. Fagon, A. Buu, et al., “Nosocomial acquisition of multiresistant Acinetobacter baumannii: risk factors and prognosis”, Clinical Infectious Diseases, vol. 20, pp. 790-796, 1995.

[32] G. Marckmann, “Recommendations for the Ethical Development and Use of Medical Decision-Support Systems”, Medscape General Medicine, 2001.

[33] J. Merrer, B. De Jonghe, F. Golliot, et al., “Complications of femoral and subclavian venous catheterization in critically ill patients: a randomized controlled trial”, The Journal of the American Medical Association, vol. 286, no. 6, pp. 700-707, 2001.

[34] C. Morel, Les décisions absurdes. Sociologie des erreurs radicales et persistantes, Gallimard, Paris, 2002.

[35] K.P. Murphy, Dynamic Bayesian Networks: Representation, Inference and Learning, PhD Thesis, UC Berkeley, Computer Science Division, 2002.

[36] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufman, San Mateo, California, 1988.

[37] D. Pittet, N. Li, R.P. Wenzel, “Association of secondary and polymicrobial nosocomial bloodstream infections with higher mortality”, European journal of clinical microbiology and infectious diseases, vol. 12, pp. 813-819, 1993.

[38] H. Qudrat-Ullah, M. Karakul, “Decision Making in Interactive Learning Environments towards an Integrated Model”, Journal of decision systems, vol. 16 no. 1, pp.79-99, 2007.

[39] L.R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, Proceedings of the IEEE, vol. 77 no. 2, pp. 257-286, 1989.

[40] H.A. Simon, The new science of management decision, Prentice Hall, New Jersey, 1977.

[41] G. Trabelsi, M. Ben Ayed, M.A. Alimi, “Système d’extraction de connaissance basé sur les Réseaux Bayésiens Dynamiques”. EGC’10, Tunisia, 2010.

[42] D.L. Veenstra, S. Saint, S.D. Sullivan, “Cost-effectiveness of antiseptic-impregnated central venous catheters for the prevention of catheter-related bloodstream infection”, The Journal of the American Medical Association, vol. 282, pp. 554-560, 1999.

[43] H. Yoshikawa, “Modeling Humans in Human-Computer Interaction”, In Jacko, J.A., Sears, A. (eds.), The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Mahwah, New Jersey, Lawrence Erlbaum Associates, pp. 118-149, 2003.

AUTHORS PROFILES

Hela Ltifi profile

Hela Ltifi has a Ph.D. in computer science. She is a member of the REGIM (ENIS, Tunisia). Her research activities concern DSS based on a KDD process and user-centred design; her application domain is hospital infections. She is a co-author of several papers in international journals and conferences. She is Assistant Professor in computer Sciences in the Faculty of Sciences of Gafsa, Tunisia. She is an IEEE member.

Ghada Trabelsi profile

Ghada Trabelsi is a Ph.D. Student in Computer Sciences. She is a member of two laboratories: REGIM (ENIS, Tunisia) and LINA (Laboratoire d’Informatique de Nantes Atlantique). Her research topics concern Bayesian Networks, espesialy and there appliccation in the temporal and dynamic data. She is contractual Assistant in IPEIS (Institut Préparatoir aux Etudes d’Ingénieurs de Sfax). She is an IEEE member.

Mounir Ben Ayed profile

Mounir Ben Ayed is gradted in Masrtery on Sciences and echnology. He obteined a PhD in Sciences - Biomedical Engineering. He is a member of the REGIM research unit, ENIS, University of Sfax, Tunisia. He is an assistant professor; he teaches DBMS, data warehouse and data mining. His research activities concern DSS based on a KDD process. Most of his research works are applied in health care domain. He is an author of many papers in international Journals and conferences. Mounir Ben Ayed has been a member of organization committees of several conferences, and particularly co-chair of the ACIDCA-ICMI'2005 organization comitee. He was the chair of the JRBA 2011(The 2nd Workshop on Bayesian Networks and their Applications) and he is the chair of the JFRB 2012 (The 6th French Speaker Workshop on Bayesian Networks). He is the former Director of Training at the Faculty of Sciences of Sfax. He is an IEEE member and chair of IEEE-EMBS Tunisia Chapter

Adel M. Alimi profile

Adel M. Alimi is graduated in Electrical Engineering 1990, obtained a PhD and then an HDR both in Electrical & Computer Engineering in 1995 and 2000 respectively. He is now professor in Electrical & Computer Engineering at the University of Sfax. His research interest includes applications of intelligent methods (neural networks, fuzzylogic, evolutionary algorithms) to pattern recognition, robotic systems, vision systems, andindustrial processes. He focuses his research on intelligent pattern recognition, learning,analysis and intelligent control of large scale complex systems.

He is associate editor and member of the editorial board of many internationalscientific journals (e.g. "IEEE Trans. Fuzzy Systems", "Pattern RecognitionLetters", "NeuroComputing", "Neural Processing Letters", "International Journal of Imageand Graphics", "Neural Computing and Applications", "International Journal of Robotics andAutomation", "International Journal of Systems Science", etc.).
He was guest editor of several special issues of international journals (e.g. Fuzzy Sets& Systems, Soft Computing, Journal of Decision Systems, Integrated Computer AidedEngineering, Systems Analysis Modelling and Simulations).He is the Founder and Chair of many IEEE Chapter in Tunisia section, he is IEEE SfaxSubsection Chair (2011), IEEE ENIS Student Branch Counselor (2011), IEEE Systems,Man, and Cybernetics Society Tunisia Chapter Chair (2011), IEEE Computer Society TunisiaChapter Chair (2011), he is also Expert evaluator for the European Agency for Research.

He was the general chairman of the International Conference on Machine IntelligenceACIDCA-ICMI'2005 & 2000.He is an IEEE senior member.

An Intelligent Location Management approaches in GSM Mobile Network

N. Mallikharjuna Rao[1]
Department Of MCA,
APGCCS,
Rajampet, India

Prof M. M Naidu[2]
Department of CSE,
S.V College of Engineering,
Tirupati, India

P.Seetharam[3]
Systems Engineer
AITS
Rajampet

***Abstract*—Location management refers to the problem of updating and searching the current location of mobile nodes in a wireless network. To make it efficient, the sum of update costs of location database must be minimized. Previous work relying on fixed location databases is unable to fully exploit the knowledge of user mobility patterns in the system so as to achieve this minimization. The study presents an intelligent location management approach which has interacts between intelligent information system and knowledge-base technologies, so we can dynamically change the user patterns and reduce the transition between the VLR and HLR. The study provides algorithms are ability to handle location registration and call delivery.**

Key words: Baste Station; MSC; HLR; VLR; IMEI; MT; Fuzzy Logic; Fuzzy databases.

I. INTRODUCTION

Recent advances in communication technology have created the opportunity for mobile terminals to receive many services that were, until not long ago, only available to tethered terminals. This system to support large scale mobility was the advanced mobile phone system. A new digital system, personal Communication System (PCS) provides voice as well as data services to wireless users. PCS works in the GSM 800/1900 MHz spectrum. There are competitive standards for analog, digital, and PCS system throughout the world [1].

One of the challenging tasks in a PCS environment is to efficiently maintain the location of the PCS subscribers in GSM who move around freely with their wireless unit. In India TRAI (Telephonic Regulatory Authority of India) is used for managing location information of the subscribers and enabling them to send and receive calls and other services such as messaging and data service.

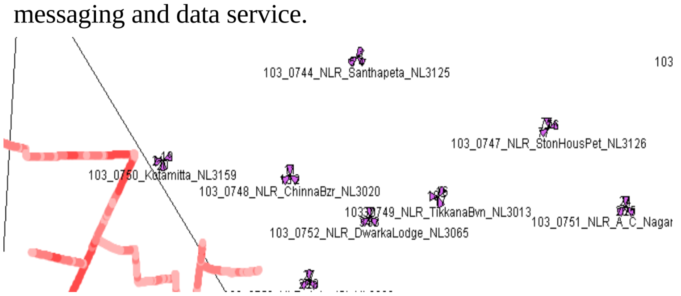


Figure 1. Sample Location Areas

The cells are established as we have shown in figure 1, using which mobile subscribers move between them and he make calls for transmitting voice/data.

The network reference model of a Personal Communication System in GSM network is shown in figure 2 we refined for supporting and understanding of my work .It consists of the following components [2].

Home Location Register (HLR): Maintains the profiles of the entire subscribers that are registered with the home network. When a mobile subscriber roams to another area, it has to register with the Visitor Location Register (VLR) of that area. The HLR maintains a pointer to the VLR which currently serve the mobile.

Visitor Location Register (VLR): Supports registration, authentication, and call routing to/from a mobile while it is away from its home area. Each MSC has a VLR to holds the data relevant for handling calls from and to the MSs that are currently located in its area.

The relevant data is downloaded from the home HLR when the mobile subscriber switches on the mobile handset in the area of the visited MSC thereby initiating the process of registration. VLR holds the exact location of the MS and keeps on updating the location as the mobile move across its area.

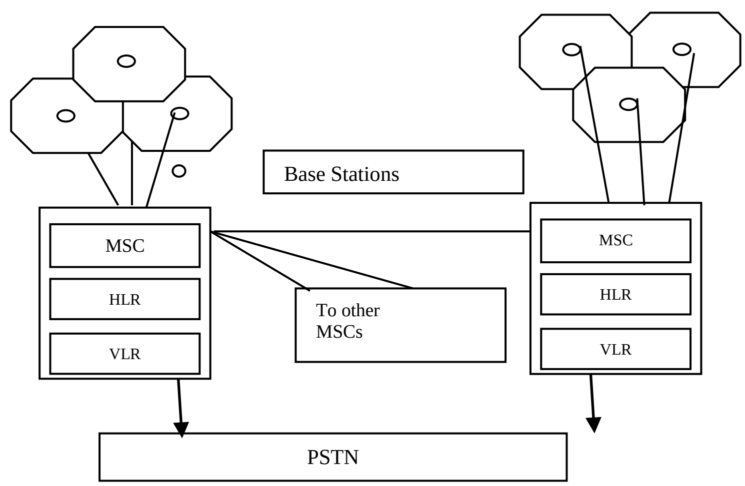


Figure 2. Network Reference Model

Mobile Switching Center (MSC): Responsible for switching the voice/data connection to the mobile host.

GMSC is Gate way Mobile Switching center which can route the calls from PSTN.

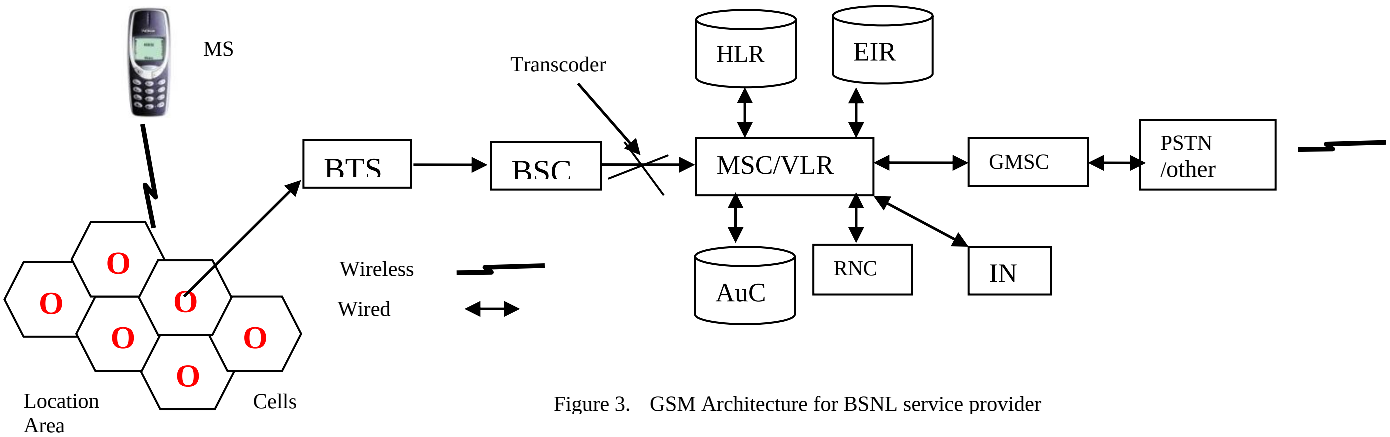


Figure 3. GSM Architecture for BSNL service provider

Base Station (BS): The base station is the gateway between the wireless network and wired network. It provides the wireless connection to the mobile subscribers within its coverage area (Cell). A set of base stations are connected to the MSC through a Base Station controller.

Authenticating Center (AC or AuC): The Authentication Center is a workstation system, which authenticates subscribers. AuC needs to access user information for authentication process so it is co-located with HLR.

Equipment Identify Register (EIR): It is a database which stores information for the identification of mobile units.

Public Switched Telephone Network (PSTN): This component refers to the regular wired line telecommunication network which is commonly accessed by landline calls.

Integrated Service Digital Network (ISDN): It is a wired line network which provides enhanced digital services to subscribers.

For supporting our work, we studied Nellore District BSNL office functionality and we find the existing architecture for GSM network as shown in figure 3.

Every subscriber is registered with a home network, the HLR of which maintains the subscriber's current physical location. This physical location is the LAI is the ID of the MSC currently serving the subscriber. If the subscriber has moved to another region then he/she has to register with the VLR that covers the new region. During registration, the VLR will contact the subscriber's HLR, and the HLR will update its database to reflect the new location of the subscriber. If the mobile has registered with some other VLR before, HLR will send a registration cancellation message to it.

Present system is having the disadvantage in call setup and call delivery using the traditional centralized database system.

One disadvantage is that since every location request as well as location registration are serviced through a HLR, in addition to the HLR being over loaded with database lookup operations [1] [3] [6]; the traffic on the links leading to the HLR is heavy. The other disadvantage is that any HLR to be unreachable even though mobiles may be roaming and away.

In view of above disadvantages, we are proposing intelligent location management approaches for to overcome all the disadvantages faced by the system. This effort will reduce the transition time between the VLR and HLR databases when the common MS roams to that MSC services area

In this paper, fuzzy logic concept is used for supporting proposed intelligent location management schemes, fuzzy logic deals with vague, doubtful and ambiguous data for giving better results for uncertain data in mobile networks. Vagueness or doubtfulness means that cannot be defined or determined data. In general, vagueness is as associated with the difficulty of making sharp or precise distinctions in the world. That is some domain of interest is vague if it cannot be delimited by sharp boundaries [14].

Ambiguity is associated with one-to-many relations that are situations in which the choice between two or three more alternatives is left unspecified. In this is paper we are proposing the techniques that can work easily to see that the concept of a fuzzy sets provides a base mathematical framework for dealing with vagueness.

Section 2 describes the existing databases system and approaches, section 3 describes the proposed intelligent system and concluding remarks is presented in section 4.

II. Background And Existing System

Location management includes two major tasks: Location registration and call delivery. Location registration procedures update the location database (HLR and VLRs) when an MT moves into different location area. Call delivery procedure locate a Mobile Station (MS) based on the information available at HLR and VLRs when a call for the MS is initiated.

In GSM network, there are two kinds of databases: HLR and VLR are used to store the location information of MSs. The whole network coverage area is divided into cells as shown in figure 1.

There is a base station installed in each cell and an MS within a cell communicates with the network through a base station. These cells are grouped together to from a larger area called a registration area (RA)/Location Area (LA).

All the base stations belonging to one LA are wired to a mobile switching center (MSC) through BTS and BSC which serves as the interface between the wireless and the wired networks. In this paper, we assume that one VLR is associated with each MSC as we shown in figure3.

A Location Area (LA) as shown in figure 3 is defined as a group of cells. Within the network, a subscriber location is known by the location area which they are in and is controlled by a Base Station Controller (BSC). The identification of a location area in which an MS is currently located is stored in the VLR.

When a MS crosses a boundary from a cell belonging to one LA into a cell belonging to another LA, it must report its new location to the network. When a MS crosses a cell boundary within a location area, it does need to report its new location to the network. But when a MS leaves the MSC service area, the scenario changes and updating information is increased. The VLR is always integrated with the MSC as shown in figure 3 and there is one VLR for each MSC service area.

The VLR can be regarded as a distributed HLR as it holds a copy of the HLR information stored about the subscriber [5]. The data stored includes:

1) *Mobile subscriber Roaming number (MSRN)*
2) *Service type (services that the subscriber is allowed to access)*
3) *Current location*
4) *HLR address Ciphering keys*
5) *Billing information*
6) *International Mobile Subscriber Identity (IMSI)*
7) *Subscribers phone number*
8) *Access point subscribed (GPRS)*
9) *Temporary Mobile Subscriber Identity (TMSI)*

When a subscriber roams into a new MSC service area, the following steps occur in figure 4 [5]

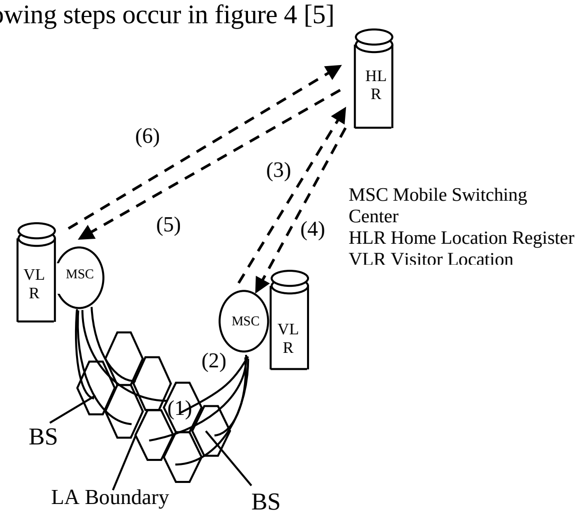


Figure 4. Call registration procedure

Step1: The VLR checks its database to determine whether or not it has a record for the MS. And this checking is based on the subscribers IMSI.

Step2: VLR sends a request to the subscribers HLR for a copy of the MS's subscription when it does not find any record for the corresponding MS

Step3: the HLR passes the information to the VLR and updates its location information for the subscriber.

Step4: the HLR instructs the old VLR to delete the information that was stored in the database.

Step5: the VLR stores its subscription information for MS, including the latest location and status (idle). This maintains of two databases at HLR and VLR gives a flexible mechanism to support call routing and dialing in a roaming situation.

Two major steps are involved in call delivery: determining the serving VLR of called MT, and locating the visiting cell of called MT. The following step occurs in call delivery shown in figure 5 [5]

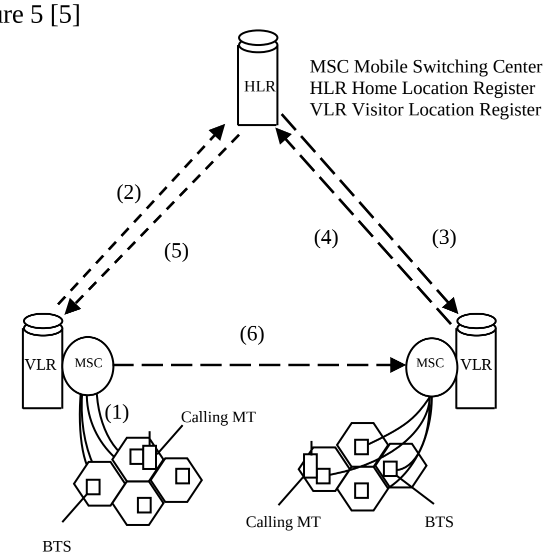


Figure 5. Call delivery procedure

Step1: Calling MT sends call initiated signal to its serving MSC through the base station.

Step2: MSC of calling MT, calling MSC, sends the location request message to HLR of called MT.

Step3: HLR of called MT determines the current serving VLR of called MT and sends the route request message to the associated MSC, called MSC.

Step4: Called MSC determines the cell location of called MT and assigns temporary location directory number (TLDN) to called MT. called MT MSC then sends this routing information (TLDN) to HLR

Step5: HLR forwards TLDN to calling MSC

Step6: calling MSC requests a call setup to called MSC through the SS7 network.

In the existing system, when subscriber is moved from one base station to other base station, subscriber's details are transferred to another base station VLR (data is transferred from HLR). If call is disconnected the same record stored at VLR is deleted automatically. For solving above problem we are proposing new approaches and these approaches are motivated from several works which were already done. In past studies [6] they proposed a novel mechanism for the VLR interaction methods where it has been given the potential to be intelligent to find the common users for the MSC service area. This effort will reduce traffic load and also improve spectrum efficiency. Moreover they concluded that, it reduces transition between VLR and HLR Therefore, it can play a significant role to support large number of traffic in the future cellular networks. In work [7] similar to [6] but they proposed fVLR, registration procedure of a mobile station location is described and a call setup procedure. They said to store and manage subscriber visits in fVLR database of which a mobile user frequently visits location. And then when set up the call path between mobile users, the VLR of the caller queries callee's fVLR for searching the location of callee instead of requesting to HLR of the callee. In this work fVLR can be well applied to the mobile users that live a well regulated life. In another study [10] proposed intelligent approach by taking the User Profile History (UPH); to reduce the location update cost. They obtained results the efficiency of UPH in significantly reducing the costs of both location updates and call delivery. In another study [12] proposed fuzzy logic approaches for user location tracking which is highly motivated us to work on these approaches for location management. In this study [13] demonstrate the sum of update and lookup costs of the location database must be minimized. Along with these

As per intelligent database system for location management, record will not be removed immediately from the VLR even subscribers is changed his/her MS. Proposed system will store for few days if any repetitions calls to the same base station occurs. Its means that system will act intelligently to take the decision, then removes the record from the VLR.

III. PROPOSED SYSTEM

Usually when the Mobile Station leaves one MSC service area to another, all the information related to the Mobile Station user is deleted (based on the subscribers IMSI or TMSI). Here in the proposed method deletion will be done depending on a decision that will be taken after doing some analysis during certain amount of time frame: 1) The VLR will have been observing the mobile station roaming habit for some days , for example 'n' days. 2) The VLR will also observe the time when the Mobile station roams into the MSC area. Depending on these data, a query will be done and the result of the query will help to decide that whether the information of mobile station will be stored inside the VLR or be dropped. The goal of the query is to find those mobile users who used to visit the MSC service area at least once regularly during the observation period. The found regularly visiting Mobile station will be termed as common mobile station. Some information like account balance, service validity and service that subscriber is allowed to access, is always needed to be updated.

A. Proposed Model Scenario

Let us consider a Mobile Station user travels 115 km daily. He lives at 'Nellore' and goes to his office at 'Rajampet' which is 115 km apart from his house. Since he is to drop his wife and daughter at 'Podalakur' and 'Rapur' respectively, he is to travel same path daily. Now the VLR1 will find that the person leaves this area at 8 am and returns at 8 pm. And the VLR2 of MSC service area will find that the person enters this service area at 8.30 am and leaves at 8.40 am daily. Similarly we assume, the possible roaming time in different MSC service area (here in this kadapa area is come MSC1/VLR1 and Rajampet comes in MSC2/VLR2 area). The same information is used for call delivery strategy.

In this paper, we are proposing intelligent location management scheme for finding the frequently visited mobile subscribers in particular location area, as we said above use fuzzy logic, the fuzzy sets within the field of decision making have for the most part consisted for extensions or "fuzzifications" of the classical theories of decision making.

Classical decision making generally deals with a set of alternatives comprising the decision space, a set of states of nature comprising the state space, a relation indicating the state or outcome to be expected from each alternative action.

B. Decision making of intelligent database VLR using Fuzzy Logic

Now we consider during the '*m*' days of observation period, the number of roaming users is 'N' on all (week) working days respectively in Mobile Switching Center service area. Then, the observation set for each day can be considered as follows. For example, a mobile subscriber roams from LA_1 to LA_2 frequently, intelligent system process the following:

Subscriber moves from LA_1 to LA_2, subscriber profile reads from HLR and sends to existing VLR for subscriber identity to make a call. The subscriber moves from existing LA_2 to LA_3 and then subscriber record will be deleted automatically and subscriber's identity will be send to LA_3. Suppose, if Subscriber again visits LA_2, subscriber profile reads from HLR and is send to VLR for identity, which we are presented in scenario model. But, here we have to remember one thing, that how many times the mobile subscriber visits or roams in one particular LA, this is called uncertainty. In this paper, we are presenting the fuzzy decision making for handling such uncertainties to retrieve subscribers records from VLR database. The fuzzy model of decision making is proposed by Belleman and Zadesh [1970] [11], which we are illustrating by a simple example. Suppose we choose different LA and their VLR sets which are mentioned below in mobile switching center:

Fuzzy Set= $\{(LA_1, VLR_1), (LA_2, VLR_2)\ldots (LA_n, VLR_n)\}$

The characteristics function of a crisp set assigns a value of either 1 or 0 to each individual in the universal set, thereby discriminating members and nonmembers of the crisp set under consideration. This function can be generalized such that the value assigned to the elements of the universal set fall within a specific range and indicate the membership grade of these elements in the set in question. Larger values denote higher degree of set membership. Such a function is called a membership function and the set defined by it's a fuzzy set [11]. The range of values of membership functions is the unit interval [0, 1]. Here each membership function maps elements of a given universal set X, which is always a crisp set, into real numbers in [0,1]

The membership function of fuzzy set A is defined by a,

A:$\rightarrow$ X [0, 1]

Once a fuzzy decision has been arrived at, it may be necessary to choose the "best" single crisp alternative from this fuzzy set. A fuzzy set may be represented by a meaningful fuzzy label. A reasonable expression of these concepts by trapezoidal membership S_1, S_2, S_3 these functions are defined on the interval [0, 20]. For example, "Low_visits", "Medium_Visits", "High_Visits" are linguistic variable for fuzzy set frequently visited locations maintained in mobile VLR databases. The linguistic representation is as follows:

$$\text{Low\_Visits} = \begin{cases} 1, & \text{No\_of\_visits} \leq 4 \\ (8-\text{No\_of\_visits})/5, & 4< \text{No\_of\_visits} < 8 \\ 0, & \text{No\_of\_visits} \geq 8 \end{cases}$$

$$\text{Medium\_Visits} = \begin{cases} 0, & \text{No\_of\_visits} \leq 8 \\ (\text{No\_of\_visits}-12)/5, & 9< \text{No\_of\_visits} < 12 \\ (14-\text{No\_of\_visits}/5, & 12\leq \text{No\_of\_visits} < 14 \\ 1, & 14 < \text{No\_of\_visits} \geq 15 \end{cases}$$

$$\text{High\_Visits} = \begin{cases} 0, & \text{No\_of\_visits} <16 \\ (18-\text{No\_of\_visits})/5, & 16\leq \text{No\_of\_visits} < 18 \\ 1, & \text{No\_of\_visits} \geq 18 \end{cases}$$

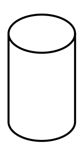
Figure 6. Triangular membership functions for Visits

Suppose, if we assume three mobile subscribers they are visiting and registered at VLR_1 database.

Mobile-subscribers set = {(S_1, 5), (S_2, 12), (S_3, 17)} as we shown in table 1.

TABLE I. VISITORS INFO IN VLR DATABASE

ID	Common Visits	Current visits
S_1	3,5,6	5
S_2	2,3,4	12
S_3	3,4,5	17



In this above table, if we observe S_1 has visited 5 times on the days of 3, 5, and 6 in a 7 day format. That means he/she has visited more than one time in day. The decision making scenario using fuzzy logic is, so we can calculate maximum number of times visited to VLR_1 in LA_1, the membership degree is:

Subscriber S_1 visited to VLR_1 5 times

Subscriber S_2 visited to VLR_1 12 times

Subscribers S_3 Visited to VLR_1 17 times

Membership degree for subscriber S_1 is 8-5/5 = 3/5 =0.6

Membership degree of subscriber S_2 is 14-12/5= 2/5=0.4

Membership degree of subscriber S_3 is 18-17/5=1/5=0.2

The intersection for fuzzy set is

$\mu_{A\cap B}(x) = \text{Min}\,[\mu_A(x), \mu_B(x)]$

Therefore, to find the frequently visited subscriber is

= Min [S_1/0.6, S_2/0.4, S_3/0.2] or Min [0.6, 0.4, 0.2]

= S_3 (0.2)

= Frequently visited Subscriber is S_3

This fuzziness allows the decision maker to frame the goals and constraints in vague, linguistic terms; which may more accurately reflect the actual state of knowledge. In this model, we have shown to stored the linguistic values in fuzzy VLR in figure 7.

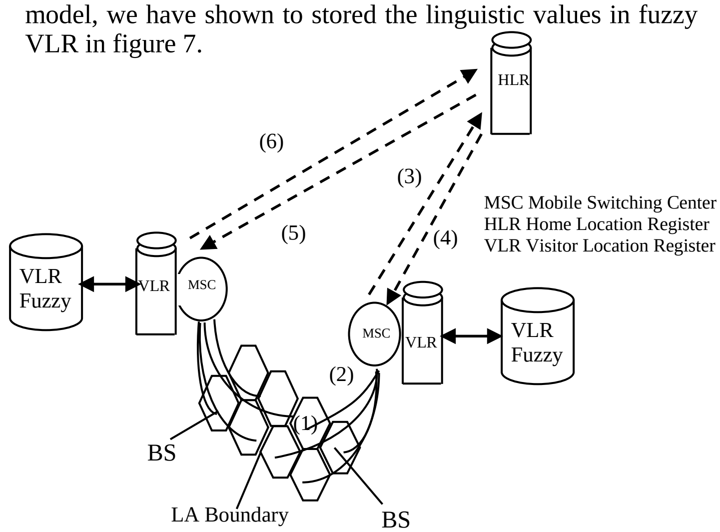


Figure 7. Proposed Intelligent Call

As we shown in figure 7 VLR database is divided into two tiers one is low-tier and second one is high-tier DBs. They are represented as V (Traditional VLR) and V_1 (Fuzzy VLR). When the subscribers come into LA, MS functions checks in V and V will check at V_1, if it found that there is no need to search the subscriber's profiles from the Home Location Register (HLR). i.e. there are two operations are performed in this system. They are $\text{VLR}_{\text{traditional (tier1)}} \longrightarrow \text{VLR}_{\text{fuzzy (tier2)}}$. Immediately $\text{VLR}_{\text{fuzzy (tier 2)}} \longrightarrow \text{VLR}_{\text{traditional (tier1)}}$ transfers the old record to VLR. In this scheme, VLR fuzzy is stores the information if subscribers are frequently visited subscribers in that location area. Record will be deleted after specified period automatically from both the VLRs. In this case, the fuzzy constraints may be defined on the set *V* and fuzzy goals on the set V_1 such that

μ_C: $V \rightarrow [0, 1]$ and μ_G: $V_1 \rightarrow [0, 1]$

function *f* can then the defined as a mapping from the ser of actions *V* to the set of outcomes $V_1 \; f\!\!\rightarrow V \quad V_1$ such that a fuzzy goal G defined on set V induces a corresponding fuzzy

$$\mu_D(x) = \text{Min}\,[\mu_G(x), \mu_C(x)]$$

goal G' on the set V, thus $\mu_G'(x) = \mu_C(x)$ a fuzzy decision D may them be defined as the choice that satisfies both the goals G and the Constraints C. if we interpret this as a logical, we can model it with the intersection of the fuzzy sets G and C D= G∩C the fuzzy decision D is the specified by the membership function

C. Algorithm for Intelligent VLR

The Location management algorithm makes use of the fact that an average mobile user has limited number of frequently visited locations. Usually a type of activity such as work, school, or shopping, occurring at a particular location. As an example, the shopping locations are located at specific sites, and for most users, the work, school, and home locations are fixed. We are considering ‘m’ days of observation period for a VLR to find its common mobile station and the searching must be based on IMSI of the visited MS during that time frame. A possible algorithm is given below:

```
Integer get_common_MS ()
{           Integer i=0; k=0; m=0;
        Interger array AA[m][];
        AA[m] AA[m] = days_IMSI_copy(1,2,3,4,5,6,7);
      b_day = get_busiest_day_user();
     Max=max1=maximum_user_in_busy_day ();
          // comparing to comparing to get common IMSI;
      for ( ; i<=max; i is incremented by 1)
     {
          for (; j<=m; j is incremented by 1)
          for (;k<max1;k is incremented by 1)
            {
                     If AA[j][k]=b_day[i])
                 {   d is incremented by 1
                  If (d==m) copy _IMSI (A[i]);
                 }
            } }
```

Figure 6: Proposed Intelligent VLR algorithms

D. Algorithm for Intelligent call delivery using fuzzy Logic

Frequently visited locations of a user are not pre-defined in location management, since this information is not always collected from mobile subscriber. whenever a subscriber want to make a call, profile identification is required to check whether the subscriber identity is verified and accepted by HLR or not, but in our proposed system, if subscriber is a known visitor of that particular location, call will be registered in VLR and will not be deleted immediately by the VLR register after subscriber moves out of VLR, as we shown in figure 7, the visited records are saved in fuzzy VLR, we will fix up time to each and every subscriber for keeping his profiles in VLR database, if they do not visit that location within the specified time, the subscriber record then be deleted from VLR as well as Fuzzy VLR. Therefore, it is the responsibility of the location management algorithm to intelligently determine such information.

The intelligent algorithm determines these frequently visited locations of a user at a given time of day based on the individual user profile.

E. Proposed call delivery algorithm

In this system, VLR maintains two entries for storing the data as we shown in figure 8. One is for the high-tier database is called VLR and another is for the low-tier database is called Fuzzy VLR for storing fuzzy related values. Mobile subscriber roams in the same area frequently; it not required in retrieving the subscriber record from HLR for identity that record will be there in fuzzy VLR database up to some period. If the subscriber will not visit same location in specified period of time record then the record will be deleted automatically. Call delivery setup will have two Databases one as HLR and another as VLR, in our proposed system the VLR database will be divided into two tier architecture database which will have High tier database as VLR and Low Tier database as VLR Fuzzy. Subscribers make a call data transfer in the following way:

VLR $\longrightarrow$ VLR_{Fuzzy} if subscribers profile is available call will be delivering without asking HLR because subscriber is visits the same location frequently. If subscriber is new for that location that means VLR_{fuzzy} send signal VLR and VLR ask subscriber profile from HLR. (i.e. $VLR_{fuzzy} \longrightarrow$ VLR). These two tier system as we shown in figure 8 for call delivery process.

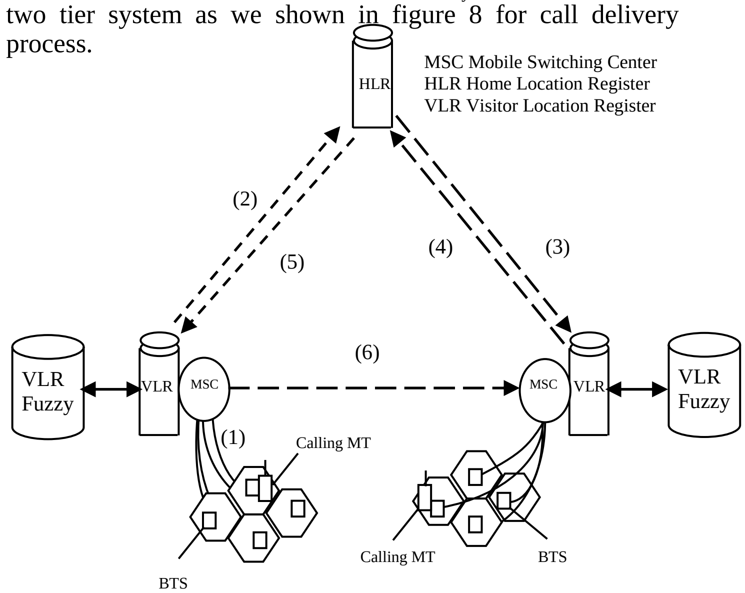


Figure 8. Proposed Intelligent Call

We present trapezoidal membership for mobile subscribers, who are visiting frequently on the interval [0, 7]. For example, “Low_visits”, “Medium_visits”, “High_Visits” are linguistic variable for fuzzy set frequently visited locations maintained in mobile VLR databases. The linguistic representation is as follows for identify the subscribers are visiting the same location area or not.

By using fuzzy relations we tested for 7 days and stored it in and as we show the proposed fuzzy VLR database values in table 2. Intelligent call delivery algorithm is derived basing on the values shown in the Table 2.

Call delivery setup and deletion of subscriber profile at VLR database will be depended upon the algorithm given through Fuzzy. For example S_1 Subscriber visits are limited to 1, S_2 Subscriber visits are limited 0, in case of S_2 Subscriber the VLR database will delete S_2 profile immediately after 7 Days (we have considered 7 days for observation). Likewise subscriber frequency of visits is defined as Low, Medium and High as fuzzy linguistic variables.

$$\text{Low\_Visits} = \begin{cases} 1, & \text{No\_of\_visits} < 1 \\ (\text{No\_of\_visits}-1)/3, & 1 < \text{No\_of\_visits} \le 2 \\ 0, & \text{No\_of\_visits} > 2 \end{cases}$$

$$\text{Medium\_Visits} = \begin{cases} 0, & \text{No\_of\_visits} > 2 \\ (\text{No\_of\_visits}-2)/3, & 2 < \text{No\_of\_visits} < 4 \\ (6-\text{No\_of\_visits}/3, & 4 \le \text{No\_of\_visits} < 5 \\ 1, & \text{No\_of\_visits} < 5 \end{cases}$$

$$\text{High\_Visits} = \begin{cases} 0, & \text{No\_of\_visits} > 5 \\ (7-\text{No\_of\_visits})/3, & 5 \le \text{No\_of\_visits} < 6 \\ 1, & \text{No\_of\_visits} < 6 \end{cases}$$

Figure 9. Triangular membership functions for visits in a week

TABLE II. SUBSCRIBERS VISITING STATISTICS

Mob ID	Visits in a week							No. of Visits	Freq. Visits	Time Expiry
	D_1	D_2	D_3	D_4	D_5	D_6	D_7			
S_1	0	0	0	0	0	0	1	1	Low	week
S_2	0	0	0	0	0	0	0	0	Low	week
S_3	1	0	1	0	1	0	1	4	Medium	week
S_4	1	1	0	1	1	1	1	16	High	week

A possible algorithm is given below:

```
Integer get common_Call ()
{
        Integer i=0; k=0; n=0;
        Integer array hlr[n][]
         ?? n =7
        ?? ans = ??;
        ?? Test =??;
          ?? i = ??;
          ?? k =??;
         ??[][] hlr = new??[??][??];
        Test = 1;
        While (test > n)
  {
        ans = get???("specify that a Subscriber is
arrived or not (0/1)");
hlr[i][k] = ans;
           k = k + 1;
Test = test + 1;
           }
 }
```

Figure 9: Proposed Intelligent Call Delivery algorithm

We developed above algorithms by using Raptor, Raptor is a simple-to-use problem solving tool that enables the user to generate executable flowcharts and algorithm.

IV. RESULTS

We felt that we have the following Advantages with the proposed location management approaches:

- As we understand that the proposed schemes are reduce the transition between VLR and HLR database because search/update/delete operations of the subscriber profiles every time from the HLR database whenever the subscribers are reached to same location area or if it is already roamed by subscriber earlier. However, this proposed mechanism also reduces the transition between MS and MSC.
- Call delivery algorithm, which can reduce the call delivery latency in the intelligent registration scheme because subscriber record is already available at VLRFuzzy database hierarchy. It is avoided multiple registrations for call delivery.
- As we said earlier, fuzzy databases are more flexible than traditional database to querying the data from the databases. Hence, the proposed approaches will reduce the retrieval cost also.
- Moreover, Fuzzy databases are support multi key file structures to retrieve the records very fast and accurately.

V. CONCLUSIONS

Mobile subscriber moves anywhere in mobile networks, and then location registration is need always for making calls. In this study, we noticed how these two databases Home Location Register (HLR) and Visitors Location Register (VLR) reflects when subscribers make call. When subscriber makes a call two basic operations are performed.

In this paper, we proposed intelligent location management schemes; they can reduce the cost of maintaining location of mobile users by using Fuzzy Logic and Fuzzy databases for call registration and call delivery procedure of a mobile station for describing location and a call setup. We demonstrate the how to store and manage fuzzy data crisp values of which a mobile user frequently visits the location. Our proposed fuzzy based VLR can be well applied to the mobile subscribers that live a well regulated life. In the future we will further research improved on mobility management using fuzzy logic and fuzzy databases.

REFERENCES

[1] William C.Y Lee, Mobile cellular Communications: Analog and Digital systems, McGraw-Hill, Inc., 2nd edition, 1995

[2] Jan Thornier, Intelligent Networks, Artech House, Inc., 1994

[3] C.N Lo, R.S. Wolff, and R.C Bernhard's, "An estimate of network database transaction volume to support personal communication service" in IEEE First international conference on universal Personal communications, 1992, pp. 236-241

[4] Theodore S Rappaport, "Wireless Communication Principle and Practice", 2nd, Person Education, LTD

[5] Kalveh Pahlavan and Prashant Krishnamurthy, "Principle of wireless Networks a unified Approach"

[6] Jae-Woo LEE, "Mobility Management Using Frequently Visited Location Database", IEEE, International Conference on Multimedia and Ubiquitous Engineering, 2007

[7] S.H Shah Newaz and Md. Asif Hossain " Intelligence adaptation in visitor Location Register to Enhance the Performance of Next generation Cellular Network" IEEE pp1517-1521

[8] Karunaharan Ratnam and Ibrahim Matta, "a fully distributed location management scheme for large PCS", vol 12, pp: 215-220

[9] Sangheon Pack and Wonjun Lee "An efficient call Delivery algorithm in Hierarchical Cellular Networks" ICUFN 2009.

[10] Singh, J.A.P Karnan M, "Intelligent Location Management Using Soft Computing Technique", IEEE International Conference on Communication software and networks , 2010

[11] Pijush Kanti Bhattacharjee, & Chandan Tilak Bhunia "Artificial Intelligence Based Authentication Technique with Three Entities in 3-G Mobile Communications", International Journal of Computer and Electrical Engineering, Vol. 2, No. 2, April, 2010, ISSN: 1793-8163

[12] Sajal Saha, "Intelligent Paging Based Mobile User Tacking Using Fuzzy Logic", International Conference on methods and models in science and Technology, ISBN: 978-0-7354-0879-1, 2010

[13] Kevin Lee, Sanjay Jha, Nirupama Bulusu, "Adaptive, Distributed Location Management in Mobile, Wireless Networks", 2009

[14] http://www.scribd.com/doc/50284358/Fuzzy-Sets-and-Fuzzy-Logic-Theory-and-Applications-George-j-Klir-Bo-Yuan

[15] TRAI, Telephonic Regulatory authority of India.

AUTHOR PROFILE

N. Mallikharjuna Rao is presently working as Associate Professor in the Department of Master of Computer Applications at Annamacharya PG College of Computer Studies, Rajampet and having more than 12 years of Experience in Teaching UG and PG courses. He received his BSc (Computer Science) from Andhra University in 1995, Master of Computer Applications (MCA) from Acharya Nagarjuna University in 1998, Master of Philosophy in Computer Science from Madurai Kamarj University, Tamilnadu, in India and Master of Technology in Computer Science and Engineering from Allahabad Agricultural University, India. He is a life Member in ISTE and Member in IEEE, IACSIT, and IAENG. His research interests are on Mobile networks, Mobile databases and Fuzzy Databases. He is a research scholar from Acharya Nagarjuna University under the esteemed guidance of Dr. M. M. Naidu, Dean & Professor, Dept of CSE, SV University, Tirupati, AP.

Modified Genetic Algorithms Based Solution to Subset Sum Problem

Harsh Bhasin
computergrad.com
Faridabad, India

Neha Singla
Student, YMCAUST
Faridabad, India

***Abstract*—Subset Sum Problem (SSP) is an NP Complete problem which finds its application in diverse fields. The work suggests the solution of above problem with the help of genetic Algorithms (GAs). The work also takes into consideration, the various attempts that have been made to solve this problem and other such problems. The intent is to develop a generic methodology to solve all NP Complete problems via GAs thus exploring their ability to find out the optimal solution from amongst huge set of solutions. The work has been implemented and analyzed with satisfactory results.**

Keywords-subset sum problem; genetic algorithms; NP complete; heuristic search.

I. INTRODUCTION

Theoretical computer scientists are in an agreement on the issue that a minimum requirement of any efficient algorithm is that, it runs in polynomial time that is O (n^c), for some constant c. But there are certain problems that cannot be solved in polynomial time. Cook, Karp, and others, defined such class of problems as NP-hard problems [1]. Some of the NP hard problems include Travelling Salesman Problem (TSP), Boolean Satisfiability problem, Subset Sum problem, Knapsack problem, Hamiltonian Path problem, Post Correspondence Problem (PCP) and Vertex Cover Problem (VCP). TSP and VCP have already been dealt with in the previous works [1], [2]. As regards PCP the solution that was proposed had some constraints but considering the fact that on analysis it gave acceptable results, it can be said that even PCP can be dealt with Genetic Algorithms (GAs) [3]. The problem that has been discussed in the following work is subset sum problem. A Genetic Algorithm based solution has been proposed and analyzed. If accepted, it will help in solving many other such problems via the concept of reducibility.

II. LITERATURE REVIEW

Many papers as regards NP Hard problems have been studied and analyzed. In the earlier works various NP Hard and NP Complete problems have been solved using GAs which has been explained below.

A. *Vertex Cover Problem*

A vertex cover of a graph G is a set of vertices such that each edge of G is incident to at least one vertex in the set. The resultant set is said to cover the edges of G. A minimum vertex cover is a vertex cover of smallest promising size. The vertex cover number is the size of a minimum vertex cover. This problem has been solved using GA [2], giving an effective and efficient solution. In the process, Initial population was generated and encoding was performed on it. An index has been assigned to each chromosome and its fitness value and threshold were calculated. A mutation and crossover operator were also applied on above generated population and above process was repeated. Then reproduction was carried out giving the solution to vertex cover problem [2].

B. *Post Correspondence Problem*

Given a collection of dominos each of the form [x/y], where x and y are strings the problem is to determine if there is a sequence of dominos that results in a match where the top string is the same as the bottom string [4]. An Artificial Intelligence based solution using GA has been used to solve the PCP problem [3]. Initial population was generated and divided into cells. Each cell was encoded and converted into 1D array. After that, each part of row of 1D array is matched with every other part of the other rows. If strings match, solution is achieved. GA operators like crossover and mutation were applied to get the desired results [3].

C. *Travelling Salesman Problem*

Given a list of cities and their pair wise distances, the mission is to find the shortest possible tour that visits each city exactly once still keeping the new cost minimum. TSP has been solved using randomness by applying Cellular Automata (CA) and heuristics by applying GAs. Elementary CA was generated and was reduced and analyzed, out of which some rules were considered and crossover and mutation operators were applied to reanalyze the rules giving the optimal solution to the TCP problem. From the selected rules, paths were generated and from them path with minimum cost was given as solution [1]. The solution to the above problem using GAs is also being developed in a related work.

D. *Subset sum Problem*

The previous attempts to solve subset sum problem have also been analyzed. In total 6 papers have been studied and analyzed. It was observed that all the implementations work well under certain constraints. The following work uses GAs based approach to find out the solution of Subset Sum Problem.

III. NP COMPLETE

The class P consists of those problems that are solvable in polynomial time. They are problems that can be solved in time

$O(n^k)$ for some constant k, where n is the size of the input to the problem [5].

The class NP consists of those problems that are "verifiable" in polynomial time. This means that if we were somehow given a "certificate" of a solution, then we could verify that the certificate is correct in time polynomial in the size of the input to the problem [6].

A problem in P is also in NP, since if a problem is in P then we can solve it in polynomial time without even being given a certificate [6]. P is subset of NP.

A problem is in the class NP-Complete if it is in NP and it is as "hard" as any problem in NP. No polynomial-time algorithm has yet been discovered for an NP-complete problem, nor has anyone yet been able to prove that no polynomial-time algorithm can exists for any one of them [6].

A problem is in class NP-Hard if the problem is “at least as hard as the hardest problems in NP”. A problem H is said to be NP-hard if and only if there is a NP-complete problem L that is polynomial time Turing reducible to H. NP-hard problems can be of any type: decision problems, search problems and optimization problems [7]. Subset Sum Problem is an NP-Complete problem.

IV. GENETIC ALGORITHMS

After Genetic Algorithms (GAs) are search algorithms based on the theory of natural selection with an innovative flair of human touch. The central idea of research on GAs has been robustness. This class not only takes into accounts the efficiency but also afficacy [8]. The implications of robustness are the elimination of costly resigns and higher level of adaptations.

The depiction of a natural population is done using, what is called chromosomes which are nothing but a set of numbers, generally binary. Each number represents a cell and can be perceived as an affirmative or negative answer. For example, a chromosome 10110 if applied to knapsack problem can be assumed as selecting the first, third and fourth item from amongst a set of five items, as we have 1 at the first, third and fourth position. The initial population can be generated using any Pseudo Random Number Generator. Each chromosome is then assigned a fitness value. Based on this fitness value replication is done as explained in the following Table 1.

TABLE I. ROLLET WHEEL SELECTION IMPLENTATION

Chromosomes	Frequency	Cumulative Frequency
Chromosome 1	33	33
Chromosome 2	24	57
Chromosome 3	17	74
Chromosome 4	10	84
Chromosome 5	16	100
Total	100	

Now generate a random number % 100. Let it be 63. Now Cumulative Frequency 63 lies in Chromosome 3. Therefore, Chromosome 3 is replicated.

The above population is enhanced by using basic operations like crossover and mutation.

A. Crossover

Crossover operator has the significance as that of crossover in natural genetic process. In this operation two chromosomes are taken and a new is generated by taking some attributes of first chromosome and the rest from second chromosome. In GAs a crossover can be of following types

1) *Single Point Crossover:* In this crossover, a random number is selected from 1 to n as the crossover point, where n being the number of chromosome. Any two chromosomes are taken and operator is applied.

2) *Two Point Crossover:* In this type of crossover, two crossover points are selected and the crossover operator is applied.

3) *Uniform Crossover:* In this type, bits are copied from both chromosomes uniformly.

B. Mutation

Mutation is a genetic operator used to maintain genetic diversity from one generation of population to the next. It is similar to biological mutation [9]. Mutation allows the algorithm to avoid local minima by preventing the population chromosomes from becoming too similar to each other [10]. GAs involves string-based modifications to the elements of a candidate solution. These include bit-reversal in bit-string GAs or shuffle and swap operators in permutation GAs [2], [3].

C. Selection

It is quantitative criterion based on fitness value to choose which chromosomes from population will go to reproduce. Intuitively the chromosome with more fitness value will be considered better and in order to implement proportionate random choice, Roulette wheel selection is used for selection [2], [3], [11].

GAs are different from the other search processes owing to the fact that they work on coding of the parameter set and not on the parameters [12]. It is also general belief that GAs use payoff and not auxiliary knowledge. Moreover, determinism is not needed in GAs.

The initial population for GAs is generated by applying the following procedure.

Initial population is stored in a 2D array, let it be called init_pop[][].

```
for i = 0 to n
begin
    for j = 0 to m
    begin
        Generate a random number x modulo 100
        if(x <= 50) then init_pop[i][j] = 0
        else init_pop[i][j] = 1
    end
```

end

V. PROPOSED WORK

The proposed work is based on the premises that GAs imitates the process of natural selection in robust and efficient manner. In the work, the list of numbers on which subset sum is to be applied is told in the array A []. s denotes the expected sum.

The array is sorted since sorting the array by quick sort has the complexity of O (nlogn). Even if this is taken into consideration then also the proposed solution gives the better result than the existing ones, since subset sum is an NP Complete problem. The various steps of the proposed algorithm have been explained below.

A. Sorting

The given array A[] is sorted.

B. Calculating Limit Point

The following procedure is applied to find out the limit point, where the limit point is defined as the point in the array after which the members need not to be considered.

For i = 0 to n

begin

if(A[i] > s) then take i as the position of the limit point and break;

end

If no limit point is found then all the array elements need to be considered.

C. Generating Initial Population

Generate Genetic Population as explained in section IV.

D. Mapping

Mapping of genetic population to the array till the limit point is performed as follows

Let A = {1, 3, 4, 5, 9, 10}

i = 0 1 2 3 4 5

Let s = 6

The limit point become i = 4

Now/ the numbers {1, 3, 4, 5} needs to be considered. Let's call this modified array.

Let us assume that the chromosome of genetic population is 101101

Now take that many cells that are equal to the length of modified array.

i.e. 1011 is considered.

Now, 1 denotes accepting the element and 0 denotes not accepting the element.

As per the above chromosome items {1, 4, 5} are selected.

E. Calculating the Sum

Calculate the sum of selected items. For above chromosome, sum comes to be 10.

F. Reducing the Population

Accept or Reject the chromosome on the following basis.

if(sum > s) then reject the chromosome

else accept the chromosome.

G. Crossover

Crossover operator is applied on the above reduced population and step *E.* and *F.* are repeated.

H. Mutation

Mutation operator is applied on the population obtained above.

I. Moderation

The above process is used for small values of s. For large values moderation is used which is explained as follows:

Initial Population generated in Step *C.* is considered and fitness value is calculated for each chromosome using the formula,

Fitness = (1 / (1 + N1)) * 100

Where, N1 = Number of One's in second half of chromosome.

High fitness indicates that the chromosome is more relevant. Roulette Wheel Selection is applied on initial population based on the fitness value.

The above process is explained in the figure 1.

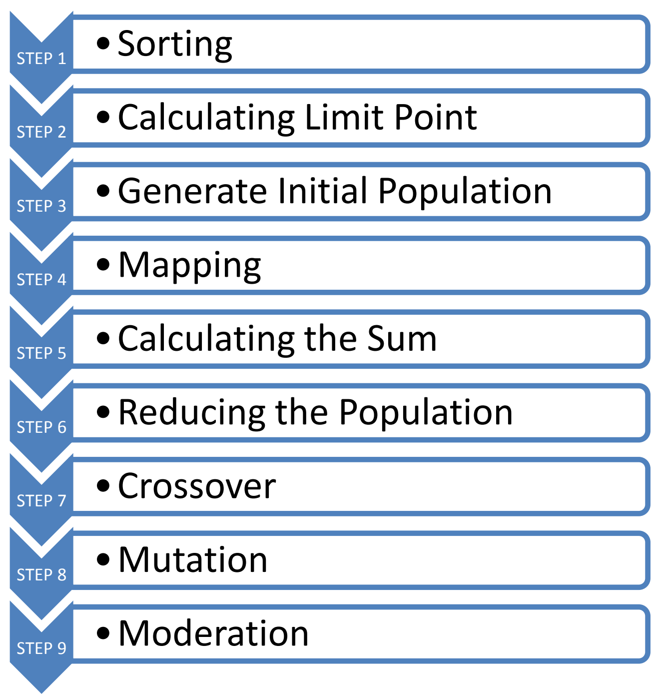


Figure 1. Modified Genetic Algorithms Based Subset Sum Solution

VI. Conclusion And Future Scope

In the analysis, 30 items were considered and were randomly generated such that the maximum number generated was 100. Analysis was done for various values of the factor s. For small values of s, i.e. less than 30, limit is calculated i.e. items which can contribute to this sum were taken and rest were not. Genetic process was applied and results were analyzed. The results have been shown in Table 2 and Fig. 2

In Fig. 2, a) Represents the percentage of sample runs which give the accurate result, b) represents the percentage of sample runs which give optimal solution and c) represents the cases where required sum is not possible, but gives the best possible result.

If the value of sum is taken as 60 or greater in this experiment, then the results were not satisfactory. On analysis, it was found that this is due to fact that the number of one's in the right half are more. So as to handle this situation, a process of moderation was applied. The results obtained have been listed in Table 3.

The rate of replication was taken as 4% and Roulette Wheel Selection was applied which resulted in the replication of favorable data, thus making the population fitter.

The overall results are encouraging. Some of them have also been shown in the following figures and tables.

It must also be remembered that it is not always the case the a finite solution to subset sum problem can be found. Moreover GAs give best solutions but are not guaranteed to do so always.

TABLE II. DATA FOR ANALYSIS 1

Number of items	30
Required Sum	20

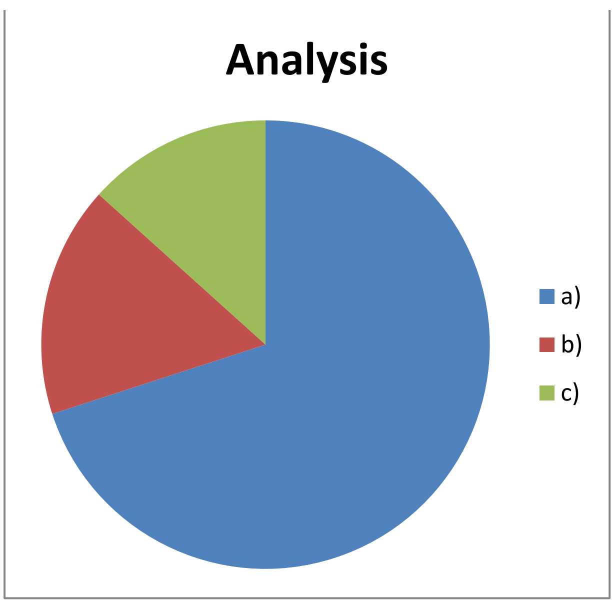


Figure 2. Analysis 1

TABLE III. RESULTS

Fitness Value	Number of chromosomes in initial population having this fitness value	Number of chromosomes in replicated population having this fitness value
25	2	9
20	11	36
16	16	44
14	29	70
12	42	96
11	36	74
10	36	60
9	19	36
8	5	8
7	4	7

References

[1] Harsh Bhasin and Neha Singla, "Harnessing Cellular Automana and Genetic Algorithms to solve Travelling Salesman Problem", communicated.

[2] Harsh Bhasin and Gitanjali , "Harnessing Genetic Algorithm for Vertex Cover Problem", International Journal on Computer Science and Engineering (IJCSE), Vol. 1, Issue 2, pp. 218 - 223.

[3] Harsh Bhasin and Nishant Gupta, "Randomized algorithm approach for solving PCP", IJCSE 2012; Vol 4, Issue 1, pp. 106-113.

[4] Nasseef Abukamail, " Specialized Version of the Post Correspondence Problem (PCP)", Presented to: Dr. David Juedes CS506, Computation Theory, 1988.

[5] P C. Bell and Igor Potapov, "The Identity Correspondence Problem and its Applications", 2010.

[6] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, "Introduction to Algorithms", the MIT Press, Cambridge, Massachusetts, USA, 2009.

[7] Michael R. Garey and David S. Johnson (1979), "Computers and Intractability: A Guide to the Theory of NP-Completeness". W.H. Freeman.

[8] Cormen, Thomas H.; Leiserson, Charles E., Rivest, Ronald L., Stein, Clifford (2001) [1990]. "35.5: The subset-sum problem". Introduction to Algorithms (2nd ed.). MIT Press and McGraw-Hill. ISBN 0-262-03293-7.

[9] H. Bhasin and S. Bhatia, "Application of Genetic Algorithms in Machine learning", IJCSIT, Vol. 2 (5), 2011.

[10] Pisinger D (1999). "Linear Time Algorithms for Knapsack Problems with Bounded Weights". Journal of Algorithms, Volume 33, Number 1, October 1999, pp. 1–14

[11] Harsh Bhasin, "Use of Genetic Algorithms for Finding Roots of. Algebraic Equations", IJCSIT, Vol. 2, Issue 4.

[12] S. Thrun, "Learning to Play the Game of Chess", In G. Tesauro, D. Touretzky, and T. Leen, editors, Advances in Neural Information Processing Systems(NIPS) 7, Cambridge, MA, 1995. MIT Press.

Authors Profile

Harsh Bhasin has completed his B. Tech in Computer Science and Engineering, M. Tech (C. E.).He is presently working in computergrad.com, proprietor Sahib Soft and a member of International Association of Computer Science and Information Technology. He has published many papers in the domain of Random Number generation and NP Hard Problem by Cellular Automata; Cryptography, Machine Learning, Equation Solving and NP Hard Problems by Genetic Algorithm; conversion of south Asian languages by Natural Language Processing. His various papers have been published in IJCA online, IJCSIT, IJCST, IJCSE, IJCNWC, IJIEE , IJCSI (To be published on 31st March) and IEEE and National and International conferences.

Neha Singla is a student of M. Tech. , YACA UST , Faridabad.

Solving the Resource Constrained Project Scheduling Problem to Minimize the Financial Failure Risk

Zhi-Jie Chen, Chiuh-Cheng Chyu
Department of Industrial Engineering and Management
Yuan-Ze University
Taiwan

***Abstract*—In practice, a project usually involves cash in- and out-flows associated with each activity. This paper aims to minimize the payment failure risk during the project execution for the resource-constrained project scheduling problem (RCPSP). In such models, the money-time value, which is the product of the net cash in-flow and the time length from the completion time of each activity to the project deadline, provides a financial evaluation of project cash availability. The cash availability of a project schedule is defined as the sum of these money-time values associated with all activities, which is mathematically equivalent to the minimization objective of total weighted completion time. This paper presents four memetic algorithms (MAs) which differ in the construction of initial population and restart strategy, and a double variable neighborhood search algorithm for solving the RCPSP problem. An experiment is conducted to evaluate the performance of these algorithms based on the same number of solutions calculated using ProGen generated benchmark instances. The results indicate that the MAs with regret biased sampling rule to generate initial and restart populations outperforms the other algorithms in terms of solution quality.**

Keywords-RCPSP; cash availability; memetic algorithms; variable neighborhood search.

I. Introduction

Cash flow is critical to the success of executing a project. The term cash flow is used to describe the net difference, at any point in time, between income (revenue) and project expenditures; negative cash flow is outgoing (cash out-flow), while positive cash flow is income (cash in-flow). In practice, cash in-flows often arise from payments due to the completion of specified parts of the project. On the other hand, cash out-flows are caused by the execution of activities, such as resource usage and necessity expenditure. Both cash in- and out- flows may occur at several points in time during execution of an activity. Usually, discount rates are taken into consideration. Some commonly used NPV maximization RCPSP models include progress payments, lump-sum payment at the pre-specified project deadline, payments at activity completion times, etc. [1-4] carried out comparative studies on the above payment models. For an overview of RCPSP with objectives based on NPV, we refer to [5, 6].

This research presents a different financial model, which aims to minimize the risk of payment failure or maximize the cash availability during project execution. The goal is to provide a cautious and less complex model to minimize the payment failure risk during the project execution. To achieve this goal, the money-time value, which is the product of the cash in-flow and the length from the time the cash received to the project makespan, can provide a financial evaluation of project cash availability. The cash availability of a project schedule is defined as the total money-time values associated with all activities. This financial metric does not consider discount rate, and it will provide a conservative estimate of cash in-flows during the project execution, since cash on hand will grow in value over time. In the proposed model, the cash in-flows are assumed to occur at the completion time of each activity, and the cash amounts can be used during the rest of project execution time. Hereafter, we shall refer to this model as the project cash availability maximization problem (PCAMP) for the resource constrained project scheduling problem (RCPSP).

The PCAMP is mathematically equivalent to the RCPSP with the objective of minimizing total weighted completion time (also known as total weighted flow time). This problem is strongly NP-hard since its sub-problem, single machine scheduling with total flow time minimization objective subject to precedence constraints, is strongly NP-hard [7]. Thus, the heuristic approach will be appropriate for solving the PCAMP for RCPSP of large size. As the objective function of the PCAMP for RCPSP is regular, the optimal solution must be an active schedule [8]. The serial SGS procedure will generate the active schedules, where each activity is scheduled as early as possible [9]. A regular objective function is a non-decreasing function of the activity start times or finish times.

To solve the PCAMP for RCPSP, we propose several memetic algorithms (MAs) that differ in initial population and restart strategy. The performance of these MAs will be compared to each other, as well as a double variable neighborhood search (DVNS) [10]. MAs are a population-based meta-heuristic that incorporates evolutionary algorithms with local search [11]. In MAs, the term “memes” refers to the strategies that are employed to improve individuals or solutions. The strategies involve local refinement, perturbation, constructive methods, restart policy, etc. Ong and Keane [12] discussed the importance of selecting local search methods in MAs. An appropriate local search method will significantly improve the efficiency of the solution search. Merz and Freisleben [13] presented an MA framework with a restart strategy, in which mutation operator is used to generate a new and diverse population when the current population is judged to be convergent.

Variable neighborhood search (VNS) was introduced by Mladenović and Hansen [14], and its principles and applications were further detailed in [15]. The basic idea of VNS is to perform a systematic change of neighborhood and to find a local optimal solution with respect to each neighborhood using a local search algorithm. This algorithm explores increasingly distant neighborhoods of the incumbent solution and jumps from there to a new one when an improvement has been made.

The remainder of this paper is organized as follows: Section 2 defines the problem; Section 3 illustrates the solution methods; Section 4 presents the numerical results; Section 5 concludes this research.

II. Problem Description

The RSPSP can be described as follows. A project consists of a set $N = \{0, 1,\ldots, J+1\}$ of activities (nodes), where activities 0 and J+1 are dummies, and respectively represent the start time and the completion time of the project. A set of precedence relationships between activity pairs must be specified for the project. An activity list (AL) is a sequence of activities that follows the precedence constraints. Activity preemption is not allowed. The duration of an activity j is denoted by d_j, and its requirement for renewable resource type k is r_{jk}, $k = 1,\ldots, K$. The availability of a resource type k for each time period is R_k units.

This model aims to minimize the risk of payment failure during project execution, which in turn to maximize the project cash availability. The following are notations and formulation of the PCAMP model.

Notation

- j : Index of activity, $j = 0, 1, 2,\ldots, J + 1$
- t : Index of time, $t = 1,\ldots, D$
- D : Project deadline
- k : Index of renewable resource type, $k = 1,\ldots, K$
- d_j : Duration of activity j
- ω_j : Cash in-flow of executing activity j
- r_{jk} : Per period usage of resource k by activity j
- R_k : Availability level of resource k per period
- DP_j : Direct predecessors of activity j
- f_j : completion time of activity j
- S(t) : Set of activities in progress at time t

Mathematical model

$$\text{Maximize} \quad \sum_{j=1}^{J} \omega_j \cdot (D - f_j) \tag{1}$$

Subject to

$$f_j \geq f_i + d_j \quad \text{for all } i \in DP_j \tag{2}$$

$$\sum_{j \in S(t)} r_{jk} \leq R_k \quad k = 1,\ldots, K;\ t=1,\ldots, D \tag{3}$$

$$f_j \geq 0 \text{ and integers} \quad j = 0, 1,\ldots, J+1 \tag{4}$$

Equation (1) is the model objective that computes the cash availability during the project execution. Constraint set (2) describes the precedence relationships among activities. Constraint set (3) specifies the usage limit per period for each renewable resource type at any time during project execution.

III. Solution Methods

All algorithms presented in this research will employ local search to refine a newly produced solution. Three local search methods are used: (1) left move at random, (2) 2-swap, and (3) forward/backward improvement (FBI; also known as justification procedure [16]). This section first introduces the three local search methods, and then the proposed MAs and DVNS. These algorithms use activity list (AL) as the encoding scheme and forward serial list scheduling (F-SLS) as the decoding scheme. An AL offers the order to schedule the activities in the project. Fig. 1 displays a simple example with a single resource availability of 5 units, where activities 0 and 9 are dummy and they represent the start and finish events of the project, respectively. Fig. 2 shows the encoding and decoding schemes through an AL = {1, 3, 4, 2, 6, 5, 8, 7}.

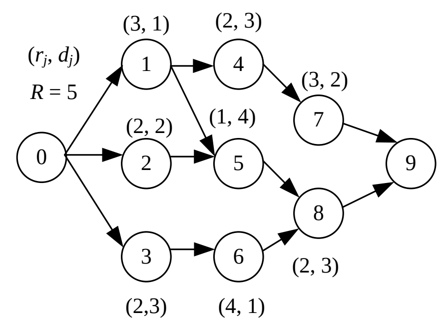


Figure 1. Example of project network

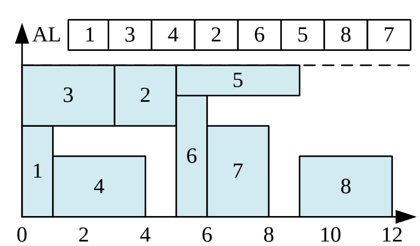


Figure 2. Example of decoding AL

A. Local Search methods

In the left move method, a position is randomly selected from the AL, and the activity on that position is then randomly placed leftward into a position allowed by precedence relations.

In Fig. 2 example, if position 5 is selected, the corresponding activity 6 can be placed on either position 3 or 4, but not position 2 (i.e. activity 3).

In the 2-swap method, an activity is randomly selected from the AL, and then the corresponding left- and right-move limits are determined. Then, within these two limits, a second activity is randomly selected for possible swapping. The procedure performs swapping when these two activities have different start times in the current schedule and the swap does not violate the precedence constraints. In Fig. 2 example, if activity 6 is selected, then it can swap with activities 4, 2, and 5.

The FBI method consists of two steps: in the forward step, the activities are scheduled as late as possible according to the order of their finish times in the current schedule; then in the backward step, the activities are scheduled as early as possible based on the order of their start times in the schedule generated in the forward step. The FBI method has been proved very effective in improving solutions for RCPSP with the objective of minimizing makespan [17]. Fig. 3 shows the forward step (i.e. right justification) and Fig. 4 shows the backward step (left justification) using the schedule in Fig. 2.

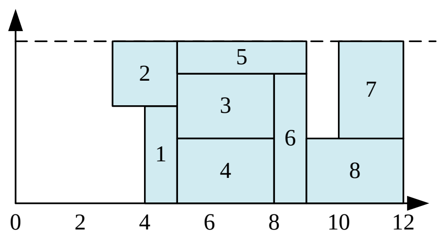


Figure 3. Forward step schedule

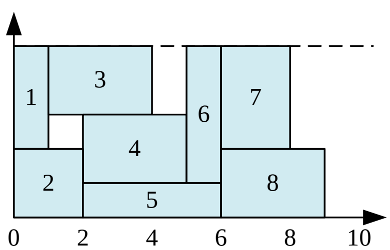


Figure 4. Backward step schedule

B. Memetic Algorithms

Memetic algorithms (MAs) are sometimes referred to as genetic local search algorithms. In this research, we present a multi-start MA framework which employs the FBI to improve each newly produced individual. In each generation, a modified order based recombination operator is applied to produce offspring. The MA will restart with a new population when the current population contains at least 80% identical individuals. This situation is regarded as the population having entered into a convergence state, and continuing evolution will have little chance of making further improvement under the recombination operation.

Our MA considers the following policies to produce individuals of the initial population and restart population: (1) randomly generated individuals followed by FBI for initial population, and a series of 2-swap followed by FBI for restart population; (2) regret biased sampling [18] followed by FBI for both initial and restart populations (also denoted as RBS + RBS); (3) regret biased sampling followed by FBI for initial population, and a series of 2-swap followed by FBI for restart population (also denoted as RBS + 2-swap); (4) randomly generated individuals followed by FBI for initial population and no restart. For any method, the best individual will be retained when generating a new population. We shall refer to the MA with policy k as MA^k for $k = 1,\ldots, 4$. Fig. 5 describes the pseudo code of the MAs.

When generating a population using the RBS rule, at each step an activity i is selected from current candidate list CL according to the following probability:

$$P(i) = \frac{[\mathrm{Max}_{m\in \mathrm{CL}}(\omega_m \cdot f_m) - \omega_i \cdot f_i + \varepsilon]}{\sum_{j\in \mathrm{CL}}(\mathrm{Max}_{m\in \mathrm{CL}}(\omega_m \cdot f_m) - \omega_j \cdot f_j + \varepsilon]} \tag{5}$$

```
Initialize Population P (Policy k);
While (termination conditions not met) do
{ for i = 1 to psize do
    {Select a, b ∈ P by tournament;
     Offspring c = recombine (a, b);
     c = FBI(c);
}end for;
 Elitist selection to produce next population P;
 If P converges (80% of individuals are identical)
    Construct P by Policy k;
} end while;
Output the best solution in P.
```

Figure 5. Framework of memetic algorithms

The recombination is performed by a modified order-based procedure described as follows: Randomly select k positions and determine the set A(k) that contains the elements that correspond to those positions in parent b. Place the elements of A(k) one by one to the k positions according to their order in parent a. Fill out the remaining positions by the activities in parent b at the corresponding positions. Repair the offspring if it violates precedence relations of activities. Fig. 6 illustrates this operation via Fig. 1 network. Suppose k = 3 and the positions are 2, 4, and 7. The corresponding activities in parent b, A(k) = {1, 5, 6}. Since $DP_3 = 6$, the child {2, 1, 4, 6, 7, 3, 5, 8} violates the precedence order and the violation will be fixed by moving activity 3 to the left position of activity 6.

C. Double variable neighborhood algorithm

Similarly, the proposed VNS uses AL and serial F-SLS for the coding schemes. Hansen and Mladenović [15] mentioned that three problem-specific questions must be considered when designing a VNS: (a) What neighborhood structures should be used and how many of them? (b) What should be their local search? (c) What strategy should be used in changing neighborhoods? The proposed VNS consists of the following features: (1) The initial solution is constructed by a greedy heuristic using the objective function values. (2) Two fundamental generators are used to construct the variable neighborhoods: 2-swap and left move. The distance of the generated neighborhood to the current solution is evaluated by the number of the operations performed. (3) An enhanced local search procedure named short-term or inner VNS is used. The inner VNS uses the efficient local search method FBI, and aims to seek the best solution in the nearby area.

Parent *a*	**1**	3	4	2	**6**	**5**	8	7
Parent *b*	2	**1**	4	**5**	7	3	**6**	8
child	2	**1**	4	**6**	7	3	**5**	8
child	2	**1**	4	3	**6**	7	**5**	8

Figure 6. Modified order-based recombination

Fig. 7 describes the framework of the DVNS. The relevant parameters are explained as follows: N_k is the k-th neighborhood, $k = 1,\ldots, Kmax$ and $s < K1 <$ Kmax, where $\{N_1,\ldots,N_s\}$ are used for the inner VNS and the remaining $\{N_{s+1},\ldots, N_{Kmax}\}$ are used for the outer VNS; each $N_k \in \{N_1,\ldots, N_{K1}\}$ is generated by performing k times of two-swap operations, and each $N_k \in \{N_{K1+1},\ldots, N_{Kmax}\}$ by performing $k - K1$ times of left-move operations. In general, a 2-swap operation takes a longer computational time than a left-move operation, but the former will make a bigger change on the current solution structure. Each N_k will generate m neighboring solutions, each of which will produce two additional solutions by FBI. The DVNS will continue to move ahead to next neighborhood regardless of whether or not the current neighborhood search has found a new best solution. In the DVNS, parameters are set to $m = 20$, $s = 5$, $K1 = 20$, $Kmax = 30$. In the performance comparison study, each algorithm will calculate MaxSol solutions. The DVNS will return to N_{s+1} if the number of solutions calculated after N_{Kmax} is smaller than *MaxSol*.

```
Initialization:
Select N_k, k = 1,…, N_Kmax; determine s, K1, MaxSol.
Generate an initial solution x by greedy heuristic;
Set k = s+1; current best solution x* = x;
While (# of solutions calculated < MaxSol) do
{ y = N_k(x); y = FBI(y); set x = best of {x, y};
  for i = 1 to s do
  { for j = 1 to m do
     { y = N_i(x); y = FBI(y); set x = best of {x, y};
       Set x* = x if objective (x) > objective (x*);
     } end for j;
   } end for i and inner VNS;
   k = k+1; if (k > Kmax) set k = s+1;
} end While;
Output the best solution x*.
```

Figure 7. Framework of algorithm DVNS

IV. Numerical results

This section presents the experimental results for the four MAs and DVNS. The test sets come from J.60 and J.120 of the PSPLIB [19]. Test set J.60 contains 480 instances, while the set J.120 has 600 instances. Each instance set is characterized by three factors: (1) Network complexity (NC) with three levels, 1.5, 1.8, 2.1; (2) Resource factor (RF) with four levels, 0.25, 0.50, 0.75, 1.0; (3) Resource strength (RS). The RS for J.60 has four levels: 0.2, 0.5, 0.7, 1.0, whereas the RS for J.120 has five levels: 0.1, 0.2, 0.3, 0.4, and 0.5. The factor NC is the average number of successors of each job, RF is the average number of resource types used by a job, and RS signifies the strength of resource availability. Each combination of factors has 10 instances. The J.120 test instances are much more difficult than J.60, mainly due to larger problem size and lower resource strength. The test environment is as follows: CPU: Intel Core i5 3G, RAM: 2G DDRIII, HD: 500G 7200 rpm, OS: Win7, Language: Visual C#.Net.

A. Generation of activity profit

The profit of each activity is assumed to be a function of the resources consumed. The following describes the method to generate the profit for activity j using r_{jk} units of resource type k, $k = 1,\ldots, K$. For each resource type k, we generate a value b_k at random from (1,000, 1,300), and set the unit cost for resource type k to $c_k = b_k/R_k$. The larger the resource limit, the cheaper the unit cost. The cost of activity j is defined as $C_j = \sum_{k=1}^{K} c_k \cdot r_{jk}$, and the profit $\omega_j = \lambda_j \cdot C_j$, where the multiplier λ_j is randomly generated from the interval of [0.3, 0.5]. The project deadline of each instance is set to 1.5 multiples of the minimum makespan, which can be found in PSPLIB (URL: http://129.187.106.231/psplib/). The critical path method (CPM) based upper bound is used to compare the results of these algorithms.

B. Computation results

An experiment was conducted to compare the performance of the proposed algorithms based on 1000 and 5000 schedules for J.60 and J.120 test sets. Each test set has 10 repetitions on each instance. The performance of an algorithm on a test set is evaluated based on the average deviation from the CPM upper bound. The deviation (in percentage) of an algorithm A for solving an instance is defined as follows:

{(CPM upper bound – best objective value found by A) / CPM upper bound} · 100%.

The 1000-schedule experiment evaluates short running time performance of an algorithm, whereas the 5000-schedule evaluates moderate running time performance of an algorithm. A small average deviation indicates that the algorithm is able to achieve a high quality performance.

TABLE I displays the performance of the algorithms for J.60 based on 1000 and 5000 schedules. The “min” column shows the best performance among 10 runs, the “avg” column presents the average performance, and the “max” column gives the worst performance. The results indicate that MA^2 (i.e. RBS + RBS) and MA^3 (RBS + 2-swap) produce the best results in short and moderate running times. The DVNS, which uses enhanced local search on each new neighboring solution, outperforms the MA with randomly generated initial population and 2-swap restart strategy (MA^1), as well as the MA without restart strategy (MA^4). It is also observed that DVNS improves its performance the most when increasing schedules from 1000 to 5000.. MA^2 and MA^3 are next in this respect, and MA^1 and MA^4 are least effective in calculating additional solutions. Similar results can be observed for J.120 in TABLE II. The deviations grow significantly for all algorithms since the problem size of J.120 is much larger.

TABLE I. ALGORITHM PERFORMANCE ON J.60

	min	avg	max	CPU
	1000 schedules			
DVNS	5.63%	5.98%	6.43%	0.11s
MA^1	5.85%	6.11%	6.39%	0.12s
MA^2	5.54%	5.76%	6.00%	0.18s
MA^3	5.52%	5.75%	6.01%	0.15s
MA^4	5.89%	6.13%	6.37%	0.21s
	5000 schedules			
DVNS	5.46%	5.69%	5.96%	0.55s
MA^1	5.75%	6.01%	6.29%	0.36s
MA^2	5.39%	5.59%	5.84%	0.86s
MA^3	5.37%	5.58%	5.85%	0.76s
MA^4	5.78%	6.03%	6.31%	0.47s

TABLE II. ALGORITHM PERFORMANCE ON J.120

	min	avg	max	CPU
	1000 schedules			
DVNS	14.68%	15.29%	15.94%	0.43s
MA^1	14.97%	15.41%	15.82%	0.50s
MA^2	14.38%	14.77%	15.16%	0.46s
MA^3	13.76%	14.76%	15.15%	0.45s
MA^4	15.08%	15.50%	15.91%	0.49s
	5000 schedules			
DVNS	14.19%	14.65%	15.17%	2.02s
MA^1	14.51%	14.95%	15.40%	2.43s
MA^2	13.76%	14.16%	14.59%	2.11s
MA^3	13.73%	14.14%	14.60%	2.01s
MA^4	14.48%	15.04%	15.49%	2.07s

TABLE III shows the performance of MA^3 for J.120 based on the factor RF. Each instance contains four resource types. Thus an RF of 0.25 implies that each activity in the project consumes one of the four resource types, whereas an RF of 1.0 indicates that each activity consumes all four resource types. The problem becomes more difficult when RF increases, and the computation time increases as well. TABLE IV shows the results of MA^3 based on RS for J.120. The smaller the RS value, the more scarce the resource. When RS decreases, the problem becomes harder and the project completion time will be longer. TABLE V presents the results of MA^3 on four hardest problems. The most difficult problem is (RF, RS) = (1.0, 0.1). In addition, the RS influences the problem difficulty more significantly than the RF. The average deviation for (1.0, 0.1) is 30.48% and for (0.75, 0.1) is 29.46%, but for (1.0, 0.2) is 24.18% and for (0.75, 0.2) is 22.58%.

TABLE III. RESULTS OF MA^3 BASED ON RF FOR J.120

RF	min	avg	max
0.25	6.40%	6.65%	6.95%
0.5	13.35%	13.82%	14.34%
0.75	16.74%	17.24%	17.77%
1	18.42%	18.85%	19.32%

TABLE IV. RESULTS OF MA^3 BASED ON RS FOR J.120

RS	min	avg	max
0.1	24.12%	24.76%	25.45%
0.2	18.39%	18.91%	19.46%
0.3	12.62%	13.02%	13.49%
0.4	8.60%	8.91%	9.24%
0.5	4.91%	5.11%	5.34%

TABLE V. RESULTS OF MA^3 FOR HARD INSTANCES OF J.120

(RF,RS)	min	avg	max
(0.75,0.1)	28.74%	29.46%	30.22%
(0.75,0.2)	21.99%	22.58%	23.25%
(1,0.1)	29.87%	30.48%	31.10%
(1,0.2)	23.69%	24.18%	24.72%

V. CONCLUSIONS

This research presents a new model to minimize the financial failure risk during the project execution. The model assumes that cash out-flows may occur at any times during activity execution, and all returns (in-flows) will be received upon activity completion times. In practice, the project owner would like to schedule activities to best prepare for cash out-flows until project completion time from a simple and conservative financial viewpoint.

The proposed project cash availability maximization model can be shown to be mathematically equivalent to the RCPSP with the objective of minimizing total weighted flow time, which is strongly NP-hard. Thus, the meta-heuristic approach is appropriate for solving this problem. Two solution approaches are presented: memetic algorithms (MAs) and a double variable neighborhood search termed DVNS. Our experimental results indicate that MA with good constructive heuristic for initial population and with restart strategy will produce high quality results.

ACKNOWLEDGMENT

This work was supported by the National Science Council in Taiwan under Grant NSC 99-2221-E-155-029.

REFERENCES

[1] M. Vanhoucke, E. Demeulemeester, and W. Herroelen, "Maximizing the net present value of a project with linear time-dependent cash flows," Int. J. Prod. Res., vol. 39, pp. 3159-3181, 2001.

[2] M. Vanhoucke, E. Demeulemeester, and W. Herroelen, "Progress payments in project scheduling problems," Eur. J. Oper. Res., vol. 148, pp. 604-620, 2003.

[3] G. Ulusoy, F. Sivrikaya-Serifoglu, and S. Sahin, "Four payment models for the mult-mode resource constrained project scheduling problem with discounted cash flows," Ann. Oper. Res., vol. 102 , pp. 237-261, 2001.

[4] M. Mika, G. Waligóra, and J. Węglarz, "Simulated annealing and tabu search for multi-mode resource-constrained project scheduling with positive discount cash flows and different payment models," Eur. J. Oper. Res., vol. 164, pp. 639-668, 2005.

[5] E. Demeulemmester, and W. S. Herroelen, Project Scheduling: A research handbook. Kluwer's International Series, 2002.

[6] S. Hartmann, and D. Briskorn, "A survey of variants and extensions of the resource-constrained project scheduling problem," Eur. J. Oper. Res., vol. 207, pp. 1-14, 2010.

[7] M. Pinedo, Scheduling: Theory, Algorithms, and Systems, 3rd edition, Springer, 2008, pp. 605.

[8] K. Neumann, C. Schwindt, and J Zimmermann, Project scheduling with time windows and scarce resources, Springer, 2002.

[9] R. Kolisch, "Serial and parallel resource-constrained project scheduling methods revisited – Theory and computation", Eur. J. Oper. Res., vol. 90, pp. 320-333, 1996.

[10] C. C. Chyu, and Z. J. Chen, "Scheduling jobs under constant period-by-period resource availability to maximize project profit at a due date", Int. J. Adv. Manuf. Technol., vol. 42, pp. 569-580, 2009.

[11] P. Moscato, On evolutions, search, optimization, genetic algorithms and martial arts: toward memetic algorithms, Technical Report, Caltech Concurrent Computer Program Report, California Institute Technology, Pasadena, CA, 1989.

[12] Y. S. Ong, and A. J. Keane, "Meta-Lamarckian learning in memetic algorithms," IEEE Trans. Evol. Comput., vol. 8, pp. 99-110, 2004.

[13] P. Merz, and B. Freisleben, "A comparison of memetic algorithms, tabu search, and ant colonies for the quadratic assignment problem," In 1999 Congress on Evolutionary Computation (CEC'99), IEEE Press, Piscataway, NJ, 1999, pp. 2063–2070.

[14] N. Mladenović, and P. Hansen, "Variable neighborhood search," Comput. Oper. Res., vol. 24, pp. 1097–1100, 1997.

[15] P. Hansen, and N. Mladenović, "Variable neighborhood search: principles and applications," Eur. J. Oper. Res., vol. 130, pp. 449–467, 2001.

[16] V. Valls, F. Ballestin, and S. Quintanilla, "Justification and RCPSP: A technique that pays," Eur. J. Oper. Res., vol. 165, pp. 375-386, 2005.

[17] R. Kolisch, S. Hartmann, "Experimental evaluation of state-of-the-art heuristics for the resource-constrained project scheduling: An update," Eur. J. Oper. Res., vol. 174, pp. 23-37, 2006.

[18] R. Kolisch, "Project scheduling under resource constraints – efficient heuristics for several problem classes," Physica, Heidelberg, 1995.

[19] R. Kolisch, and A. Sprecher, "PSPLIB – a project scheduling problem library," Eur. J. Oper. Res., vol. 96, pp. 205-216, 1996.

AUTHORS PROFILE

Zhi-Jie Chen is currently a PhD student of the Industrial Engineering and Management department at Yuan-Ze University, Chung-Li, Taiwan. His research interests include applied operations research, project scheduling, and meta-heuristics for combinatorial optimization problems.

Chiuh-Cheng Chyu is currently an associate professor of the department of Industrial Engineering and Management at Yuan-Ze University. His current research interests are in the areas of applied operations research, multiple criteria decision-making, scheduling, and meta-heuristics for combinatorial optimization problems.

Spatial Metrics based Landscape Structure and Dynamics Assessment for an emerging Indian Megalopolis

Ramachandra T.V[1,2,3*], Bharath H. Aithal[1,2], Sreekantha S.[1,3]
[1]Energy & Wetlands Research Group, Centre for Ecological Sciences [CES],
[2]Centre for Sustainable Technologies (astra),
[3]Centre for infrastructure, Sustainable Transportation and Urban Planning [CiSTUP],
Indian Institute of Science, Bangalore, 560 012, India
http://ces.iisc.ernet.in/energy; http://ces.iisc.ernet.in/grass
*corresponding author: +91-080-22933099

***Abstract*—Human-induced land use changes are considered the prime agents of the global environmental changes. Urbanisation and associated growth patterns (urban sprawl) are characteristic of spatial temporal changes that take place at regional levels. Unplanned urbanization and consequent impacts on natural resources including basic amenities has necessitated the investigations of spatial patterns of urbanization. A comprehensive assessment using quantitative methods and methodological understanding using rigorous methods is required to understand the patterns of change that occur as human processes transform the landscapes to help regional land use planners to easily identify, understand the necessary requirement. Tier II cities in India are undergoing rapid changes in recent times and need to be planned to minimize the impacts of unplanned urbanisation. Mysore is one of the rapidly urbanizing traditional regions of Karnataka, India. In this study, an integrated approach of remote sensing and spatial metrics with gradient analysis was used to identify the trends of urban land changes. The spatial and temporal dynamic pattern of the urbanization process of the megalopolis region considering the spatial data for the five decades with 3 km buffer from the city boundary has been studied, which help in the implementation of location specific mitigation measures.**

The time series of gradient analysis through landscape metrics helped in describing, quantifying and monitoring the spatial configuration of urbanization at landscape levels. Results indicated a significant increase of urban built-up area during the last four decades. Landscape metrics indicates the coalescence of urban areas occurred during the rapid urban growth from 2000 to 2009 indicating the clumped growth at the center with simple shapes and dispersed growth in the boundary region with convoluted shapes.

Keywords-Landscape Metrics; Urbanisation; Urban Sprawl; Remote sensing; Geoinformatics; Mysore City, India.

I. INTRODUCTION

Patterns and processes of globalization and consequent urbanization are the factors influencing contemporary land use trends and also posing challenges for sustainable land uses [9]. Analysis of landscape patterns and dynamics has become the primary objectives of landscape, geographical and ecological studies in recent times. Landscape changes involving large scale deforestation are the primary drivers of the climate change [52], [11] earth dynamics [51]. The spatial patterns of landscape transformation through time are undoubtedly related to changes in land uses [41].Landscape changes are diverse but very often influenced by regional policies [6]. The main driving factors for global environmental changes are been identified as agriculture intensification [17], [19], urbanisation [40] in the context of local policies[24,30,34]. The socio-economic impacts are often determinants of the type of land use within a given region, which in turn affect environmental issues [32], [35]. In order to address these urbanization challenges without compromising the environment values and their local sustainance, land use planning and necessary supporting data are crucial, especially to developing countries under severe environmental and demographic strains [12].

Urbanization is a irreversible process involving changes in vast expanse of land cover with the progressive concentration of human population. Urbanising landscapes will invariably have high population density that might lead to lack of infrastructure and provision of basic facilities. The urban population in India is growing at about 2.3% per annum with the global urban population increasing from 13% (220 million in 1900) to 49% (3.2 billion, in 2005) and is projected to escalate to 60% (4.9 billion) by 2030 [42]. Population of Mysore is 1 million as per census 2001 compared to 0.653 million (1991).

The increase in urban population is in response to the growth in urban areas due to migration from either rural area or other cities. There are 48 urban agglomerations (Mega cities, Tier I) having a population of more than one million in India (in 2011). Tier 1 cities have reached the saturation level evident from lack of basic amenities, traffic bottlenecks, higher concentrations of pollutants, higher crime rates due to burgeoning population. In this context, well planned Tier 2 cities offer humongous potential with the scope for meeting the basic amenities required. This entails the provision of basic infrastructure (like roads, air and rail connectivity), adequate social infrastructure (such as educational institutions, hospitals, etc.) along with other facilities. Modeling and visualization of urban growth based on the historical spatio-temporal data would help in identifying the probable regions of intense urbanization and sprawl.

Urban sprawl implies a sharp imbalance between urban spatial expansion and the underlying population growth [5]. Sprawl of human settlements is a major driving force of land use and land cover changes [3], [16] with detrimental impacts on natural resources and local ecology. Sprawl process entails the growth of the urban area from the urban center towards the periphery of the city municipal jurisdiction. These small pockets in the outskirts lack basic amenities like supply of treated water, electricity and sanitation facilities. Sprawl is associated with high negative impacts and especially the increasing dependency for basic amenities [50], the need for more infrastructure [5], the loss of agricultural and natural land, higher energy consumption, the degradation of peri-urban ecosystems etc., [23], [25], [27]. Understanding the sprawl over past few decades is crucial for the regional administration to handle the population growth and provide basic amenities while ensuring the sustainable management of local natural resources.

The information about the current and historical land cover/land use plays a major role for urban planning and management [54]. Mapping landscapes on temporal scale provide an opportunity to monitor the changes, which is important for natural resource management and sustainable planning activities. In this context, “Density Gradient” with the time series spatial data analysis is potentially useful in measuring urban development [50]. This article presents the temporal land use analysis and adopts the density gradient approach to evaluate and monitor landscape dynamics and further explains the landscape pattern through use of landscape metrics.

Knowledge of the spatio-temporal pattern of the urbanization is important to understand the size and functional changes in the landscape. Spatial metrics were computed to quantify the patterns of urban dynamics, that aid in understanding spatial patterns of various land cover features in the region [33]. Quantifying the landscape pattern and its change is essential for monitoring and assessing the urbanization process and its ecological consequences [31], [20], [27], [46]. Spatial metrics have been widely used to study the structure, dynamic pattern with the underlying social, economic and political processes of urbanization [21], [22], [45], [53]. This has provided useful information for implementing holistic approaches in the regional land-use planning [48]. [1] reviews the spatial characteristics of metropolitan growth including analysis [2], [4], [14], [28] the study of urban landscapes. Applications of landscape metrics include landscape ecology (number of patches, mean patch size, total edge, total edge and mean shape), geographical applications by taking advantage of the properties of these metrics [15], [39], [44] and measurement of ecological sustainability [43].

These studies also confirmed that Spatio-temporal data along with landscape metrics would help in understanding and evaluating the spatio temporal patterns of landscape dynamics required for appropriate management measures.

According to the City Development Plan (CDP), a 20-year vision document for Mysore, there has been a 70% increase in the city’s spatial extent since 2001, resulting in the higher degree of sprawl at outskirts. Objectives of this study are to understand and interpret the evolving landscape dynamics through temporal analysis of land use land cover pattern taking 3km buffer, through spatial metrics.

II. Study Area

Mysore city in Karnataka is one of the tier II cities and the cultural capital of India with a hub of industrial activities. It is designated as the 2nd capital of Karnataka. Mysore city is 128 sq. km in area and is one of the most preferred destinations for industries including IT hubs other than Bangalore.It is a main trading centre of silk and sandalwood. Mysore district is bounded by Mandya to the northeast, Chamrajnagar to the southeast, Kerala state to the south, Kodagu to the west, and Hassan to the north. It has an area of 128.42 km² and a population of about 1 million (2001 census). The district lies in the southern Deccan plateau, within the watershed region of Kaveri River, which flows through the northern and eastern parts of the district.

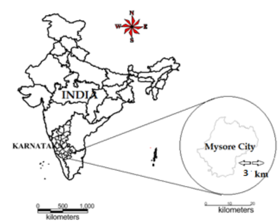


Figure 1. Study Area: Mysore city and 3 km buffer

III. Materials Used

DATA	Year	Purpose
Landsat Series MSS(57.5m)	1973	Landcover and Land use analysis
Landsat Series TM (28.5m) and ETM	1989, 1999,	Landcover and Land use analysis
IRS p6: Liss-4 MX data (5.6m)	2009	Landcover and Land use analysis
Survey of India (SOI) toposheets of 1:50000 and 1:250000 scales		To Generate boundary and Base layer maps.
Field visit data – captured using GPS		For geo-correcting and generating validation dataset

TABLE I. Materials Used In Analysis

IV. Method

A two-step approach was adopted to chart the direction of the City’s development, which includes (i) a normative approach to understand the land use and (ii) a gradient approach of 1km radius to understand the pattern of growth during the past 4 decades. Various stages in the data analysis are:

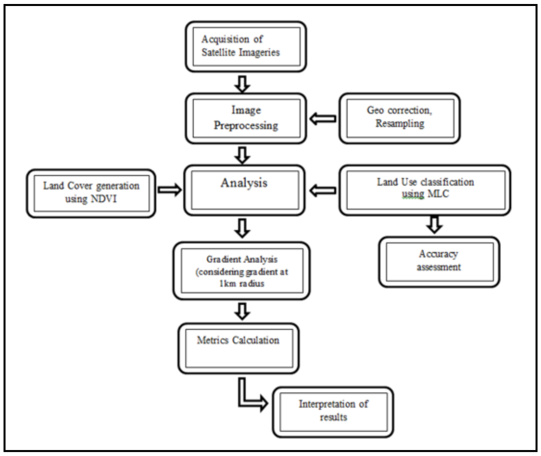


Figure 2. Procedure followed to understand the spatial pattern of landscape change

A. Preprocessing

The remote sensing data obtained were geo-referenced, rectified and cropped pertaining to the study area. The Landsat satellite 1973 images have a spatial resolution of 57.5 m x 57.5 m (nominal resolution) and 1989 - 1999 data of 28.5 m x 28.5 m (nominal resolution) were resampled to uniform 30 m for intra temporal comparisons. Latest data of IRS P6 of spatial resolution 5.6 m was procured from NRSC, Hyderabad (http://www.nrsc.gov.in).

Vegetation Cover Analysis: Normalized Difference Vegetation index (NDVI) was computed to understand the temporal dynamics of the vegetation cover. NDVI value ranges from values -1 to +1, where -0.1 and below indicate soil or barren areas of rock, sand, or urban buildup. NDVI of zero indicates the water cover. Moderate values represent low density vegetation (0.1 to 0.3) and higher values indicate thick canopy vegetation (0.6 to 0.8).

B. Land use analysis

Land use categories listed in Table 2 were classified with the training data (field data) using Gaussian maximum likelihood supervised classier. The analysis included generation of False Color Composite (bands – green, red and NIR), which helped in identifying heterogeneous area. Polygons were digitized corresponding to the heterogeneous patches covering about 40% of the study region and uniformly distributed over the study region.

These training polygons were loaded in pre-calbrated GPS (Global position System). Attribute data (land use types) were collected from the field with the help of GPS corresponding to these polygons. In addition to this, polygons were digitized from Google earth (www.googleearth.com) and Bhuvan (bhuvan.nrsc.gov.in), which were used for classifying latest IRS P6 data. These polygons were overlaid on FCC to supplement the training data for classifying landsat data.

Gaussian maximum likelihood classifier (GMLC) is applied to classify the data using the training data. GMLC uses various classification decisions using probability and cost functions [10] and is proved superior compared to other techniques. Mean and covariance matrix are computed using estimate of maximum likelihood estimator. Estimations of temporal land uses were done through open source GIS (Geographic Information System) - GRASS (Geographic Resource Analysis Support System, http://ces.iisc.ernet.in/grass).

70% of field data were used for classifying the satellite data and the balance 30% were used in validation and accuracy assessment. Thematic layers were generated of classified data corresponding to four land use categories.

Evaluation of the performance of classifiers [36], [37], [13] is done through accuracy assessment techniques of testing the statistical significance of a difference, comparison of kappa coefficients [8], [47] and proportion of correctly allocated classes [12] through computation of confusion matrix. These are most commonly used to demonstrate the effectiveness of the classifiers [8], [7], [29].

Further each zone was divided into concentric circle of incrementing radii of 1 km (figure 2) from the center of the city for visualising the changes at neighborhood levels. This also helped in identifying the causal factors and the degree of urbanization (in response to the economic, social and political forces) at local levels and visualizing the forms of urban sprawl. The temporal built up density in each circle is monitored through time series analysis.

TABLE I. a LAND USE CATEGORIES

Land use Class	Land uses included in the class
Urban	This category includes residential area, industrial area, and all paved surfaces and mixed pixels having built up area.
Water bodies	Tanks, Lakes, Reservoirs.
Vegetation	Forest, Cropland, nurseries.
Others	Rocks, quarry pits, open ground at building sites, kaccha roads.

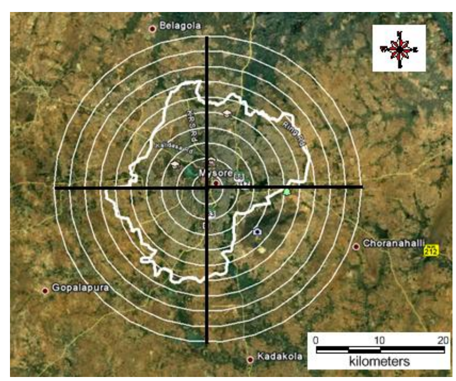


Figure 3. Google earth representation of the study region

C. Urban sprawl analysis

Direction-wise Shannon's entropy (Hn) is computed (equation 1) to understand the extent of growth: compact or divergent [26], [49], [38]. This provides an insight into the development (clumped or disaggregated) with respect to the geographical parameters across 'n' concentric regions in the respective zones.

$$Hn = -\sum_{i=1}^{n} Pi \log(Pi) \qquad \ldots\ldots \qquad (1)$$

Where Pi is the proportion of the built-up in the i^{th} concentric circle and n is the number of circles/local regions in the particular direction. Shannon's Entropy values ranges from zero (maximally concentrated) to log n (dispersed growth).

D. Spatial pattern analysis

Landscape metrics provide quantitative description of the composition and configuration of urban landscape. These metrics were computed for each circle, zonewise using classified landuse data at the landscape level with the help of FRAGSTATS [34].

Urban dynamics is characterised by 11 spatial metrics chosen based on complexity, centrality and density criteria. The metrics include the patch area, edge/border, shape, epoch/contagion/ dispersion and are listed in Table II.

V. Results & Discussion

1) Land use Land Cover analysis:

a) Vegetation cover analysis: Vegetation cover of the study area assessed through NDVI (Figure 3), shows that area under vegetation has declined to 9.24% (2009) from 51.09% (1973). Temporal NDVI values are listed in Table III.

b) Land use analysis: Land use assessed for the period 1973 to 2009 using Gaussian maximum likelihood classifier is listed Table IV and the same is depicted in figure 4. The overall accuracy of the classification ranges from 75% (1973), 79% (1989), 83% (1999) to 88% (2009) respectively. Kappa statistics and overall accuracy was calculated and is as listed in Table V.

c) There has been a significant increase in built-up area during the last decade evident from 514% increase in urban area. Other category also had an enormous increase and covers 166 % of the land use. Consequent to these, vegetation cover has declined drastically during the past four decades. The water spread area has increased due to the commissioning of waste water treatment plants (ex. Vidyaranyapura, Rayankere, Kesare) during late 90's and early 2000.

Year	Vegetation		Non vegetation	
	%	Ha	%	Ha
1973	51.09	10255.554	48.81	9583.83
1989	57.58	34921.69	42.42	8529.8
1999	44.65	8978.2	55.35	11129.77
2009	09.24	1857.92	90.76	19625.41

TABLE II. Temporal Land cover details.

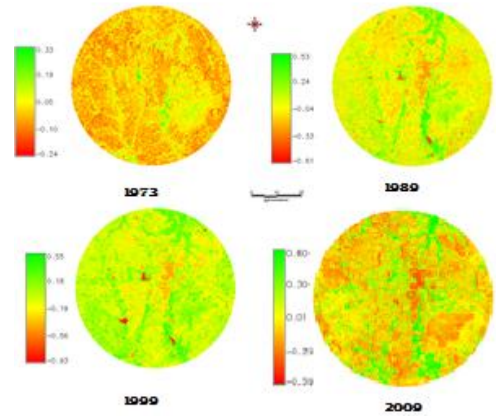


Figure 4. Temporal Land cover changes during 1973 – 2009

2) Built up Density Gradient Analysis: Built up density was minimal and the value ranges from 0.026 (considering 3km buffer) to 0.036 (without considering 3km buffer) in the North east direction (in 1973). The federal government's policy in 1990's to develop tier 2 cities led to the increase in urban area. There was a sharp growth in the region in almost all direction from 1999 till 2009, maximum value reaching 0.216 in the NE direction (considering 3km buffer) and 0.42 (without considering the buffer). This can be attributed to development of this region with the IT & BT industry which were earlier confined to Bangalore.

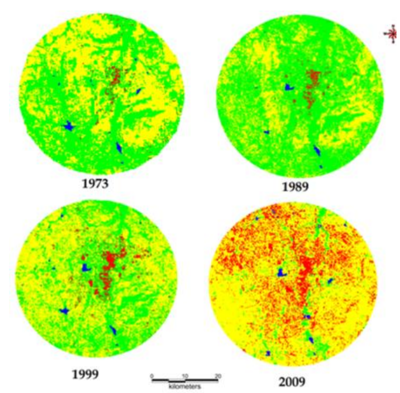


Figure 5. Classification output of Mysore

Land use / Year	Urban	Vegetation	Water	Others
1973	222.93	10705.68	124.47	9054.99
1989	229.41	13242.51	78.75	6557.4
1999	730.8	8360.1	117.9	10899.2
2009	3757.489	1159.336	142.58	15050.5
Total (Land in ha)		20108.91		

TABLE III. Temporal land use details for Mysore

	Indicators	Formula	Range
	Category : Patch area metrics		
1	Largest Patch Index(Percentage of landscape)(LPI)	$LPI = \frac{\overset{n}{j=1}\max(a_{ij})}{A}(100)$ a_{ij} = area (m^2) of patch ij A= total landscape area	$0 \leq LPI \leq 100$
2	Number of Urban Patches (NPU)	$NPU = n$ NP equals the number of patches in the landscape.	NPU>0, without limit.
3	Patch density(PD)	f(sample area) = (Patch Number/Area) * 1000000	PD>0
4	Perimeter-Area Fractal Dimension (PAFRAC)	$\frac{2}{\frac{\left[N\sum_{i=1}^{m}\sum_{j=1}^{n}(\ln P_{ij}.\ln a_{ij})\right]-\left[\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln p_{ij}\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln a_{ij}\right)\right]}{\left(N\sum_{i=1}^{m}\sum_{j=1}^{n}\ln p_{ij}^{2}\right)-\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln p_{ij}\right)}}$ Perimeter-Area Fractal Dimension a_{ij} = area (m^2) of patch ij. p_{ij}=perimeter (m) of patch ij. N= total number of patches in the landscape	$1 \leq PAFRAC \leq 2$
	Category : Shape metrics		
5	Normalized Landscape Shape Index (NLSI)	$NLSI = \frac{\sum_{i=1}^{i=N}\frac{p_i}{s_i}}{N}$ Where s_i and p_i are the area and perimeter of patch i, and N is the total number of patches.	$0 \leq NLSI < 1$
6	Landscape Shape Index (LSI)	$LSI = e_i/\min e_i$ e_i =total length of edge (or perimeter) of class i in terms of number of cell surfaces; includes all landscape boundary and background edge segments involving class i. min e_i=minimum total length of edge (or perimeter) of class i in terms of number of cell surfaces.	LSI>1, Without Limit
	Category: Compactness/ contagion / dispersion metrics		
7	Clumpiness	$CLUMPY = \begin{bmatrix} \frac{G_i - P_i}{P_i} \text{ for } G_i < P_i \;\&\; P_i < 5, else \\ \frac{G_i - P_i}{1 - P_i} \end{bmatrix}$ $G_i = \left(\frac{g_{ii}}{\left(\sum_{k=1}^{m} g_{ik}\right) - \min e_i}\right)$ g_{ii} =number of like adjacencies between pixels of patch type g_{ik} =number of adjacencies between pixels of patch types i and k. P_i =proportion of the landscape occupied by patch type (class) i.	$-1 \leq CLUMPY \leq 1.$
8	Percentage of Like Adjacencies (PLADJ)	$PLADJ = \left(\frac{g_{ii}}{\sum_{k=1}^{m} g_{ik}}\right)(100)$ g_{ii} = number of like adjacencies (joins) between pixels of patch type g_{ik} = number of adjacencies between pixels of patch types i and k	0<=PLADJ<=100
9	Cohesion	$Cohesion = \left[1 - \frac{\sum_{j=1}^{n} P_{ij}}{\sum_{j=1}^{n} P_{IJ}\sqrt[2]{a_{ij}}}\right]\left[1 - \frac{1}{\sqrt[2]{A}}\right]^{-1} * 100$	$0 \leq cohesion < 100$

10	Aggregation index(AI)	$AI = \left[\sum_{i=1}^{m} \left(\frac{g_{ii}}{\max \rightarrow g_{ii}} \right) P_i \right] (100)$ g_{ii} =number of like adjacencies between pixels of patch type P_i= proportion of landscape comprised of patch type.	1≤AI≤100
11	Interspersion and Juxtaposition(IJI)	$IJI = \frac{-\sum_{i=1}^{m} \sum_{k=i+1}^{m} \left[\left(\frac{e_{ik}}{E} \right) . \ln \left(\frac{e_{ik}}{E} \right) \right]}{\ln \left(0.5 [m(m-1)] \right)} (100)$ e_{ik} = total length (m) of edge between patch types E = total length (m) of edge in landscape, excluding background m = number of patch types (classes) present in the landscape.	0≤ IJI ≤100

TABLE IV. SPATIAL LANDSCAPE INDICES

TABLE V.

Year	Kappa coefficient	Overall accuracy (%)
1973	0.76	75.04
1989	0.72	79.52
1999	0.82	78.46
2009	0.86	84.58

TABLE VI. KAPPA STATISTICS AND OVERALL ACCURACY

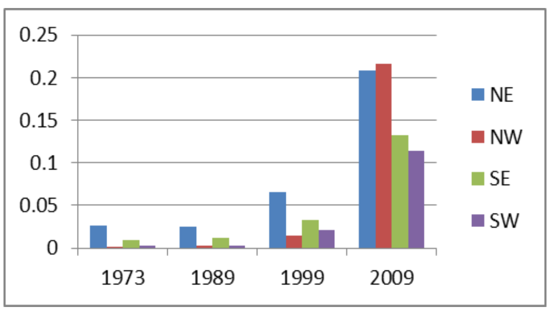


Figure 6. Urban density analysis of Mysore

3) *Urban sprawl analysis:* Shannon entropy computed using temporal data are listed in Table VI. Mysore is experiencing the sprawl in all directions as entropy values are closer to the threshold value (log (8) = 0.9). Lower entropy values of 0.007 (NW), 0.008 (SW) during 70's shows an aggregated growth as most of urbanization were concentrated at city centre. However, the region experienced dispersed growth in 90's reaching higher values of 0.452 (NE), 0.441 (NW) in 2009 during post 2000's.

The entropy computed for the city (without buffer regions) shows the sprawl phenomenon at outskirts. However, entropy values are comparatively lower when buffer region is considered. Shannon's entropy values of recent time confirms of minimal fragmented dispersed urban growth in the city. This also illustrates and establishes the influence of drivers of urbanization in various directions.

	NE	NW	SE	SW
2009	0.452	0.441	0.346	0.305
1999	0.139	0.043	0.0711	0.050
1992	0.060	0.010	0.0292	0.007
1973	0.067	0.007	0.0265	0.008

TABLE VII. SHANNON ENTROPY INDEX

4) *Spatial patterns of urbanisation:* In order to understand the spatial pattern of urbanization, eleven landscape level metrics were computed zonewise for each circle. These metrics are discussed below: Number of Urban Patch (Np) is a landscape metric indicates the level of fragmentation and ranges from 0 (fragment) to 100 (clumpiness).

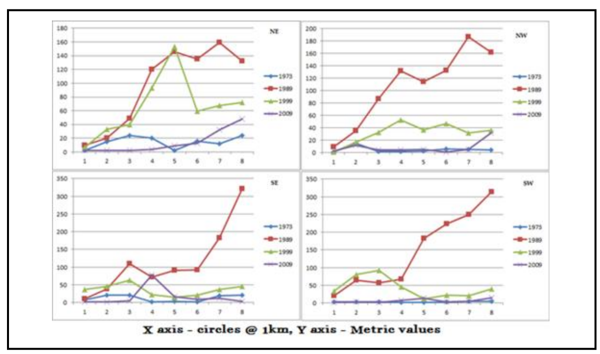


Figure 8.a Number of urban patches (zonewise, circlewise)

Figure 8a illustrates that the city is becoming clumped patch at the center, while outskirts are relatively fragmented. Clumped patches are more prominent in NE and NW directions and patches is agglomerating to a single urban patch. Largest patch index (Fig 8b) highlights that the city's landscape is fragmented in all direction (in 1973) due to heterogeneous landscapes, transformed a homogeneous single patch in 2009.The patch sizes given in figure 8c highlights that there were small urban patches in all directions (till 1999) and the increase in the LPI values implies increased urban patches during 2009 in the NE and SW. Higher values at the center indicates the aggregation at the center and in the verge of forming a single urban patch largest patches were found in NE and SW direction (2009).

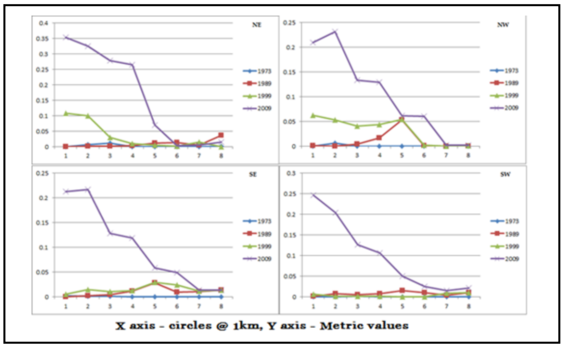


Figure 8.b Largest Patch – zonewise, circlewise

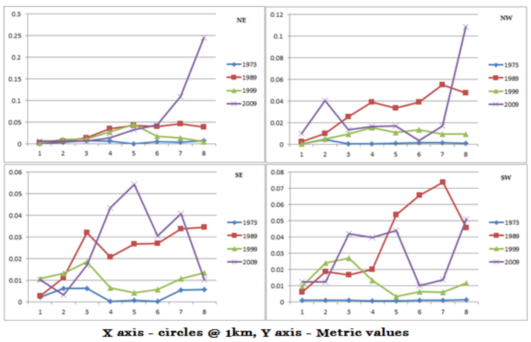


Figure 8.c Patch density – zonewise, circle wise

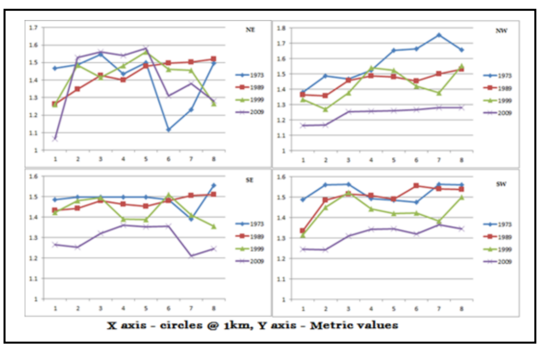


Figure 8.d PAFRAC – zonewise, circle wise

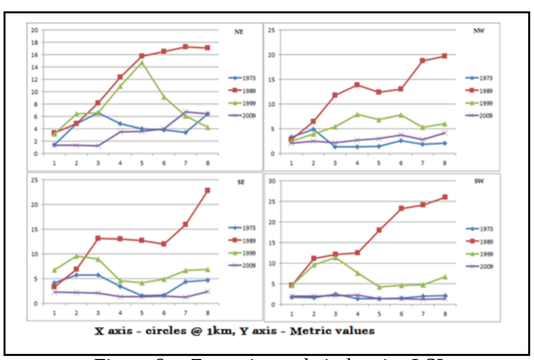


Figure 8.e Zonewise and circle wise LSI

The patch density (Fig 8c) is calculated on a raster map, using a 4 neighbor algorithm. Patch density increases with a greater number of patches within a reference area. Patch density was higher in 1973 as the number of patches is higher in all directions and gradients due to presence of diverse land use, which remarkably increased post 1989(NW) and subsequently reduced in 1999, indicating the sprawl in the region in in early 90's and started to clump during 2009, which was even confirmed by number of patches.

PAFRAC approaches 1 for shapes with very simple perimeters such as squares (indicating clumping of specific classes), and approaches 2 for shapes with highly convoluted, perimeters. PAFRAC requires patches to vary in size. Results (Fig 8d) indicate of dispersed development during 70's and 80's as PAFRAC highly convoluted. The value approaches 1 in 1990's and 2000's indicating aggregation leading to clumped region of urban land use.

Landscape Shape Index (LSI): LSI equals to 1 when the landscape consists of a single square or maximally compact (i.e., almost square) patch of the corresponding type and LSI increases without limit as the patch type becomes more disaggregated. Results (Fig 8e) indicate that there were low LSI values in 1973 as there was minimal urban areas which were aggregated at the centre. Since 1990's the city has been experiencing dispersed growth in all direction and circles, towards 2009 it shows a aggregating trend as the value reaches 1.Normalized Landscape Shape Index (NLSI): NLSI is 0 when the landscape consists of single square or maximally compact almost square, it increases as patch types becomes increasingly disaggregated and is 1 when the patch type is maximally disaggregated. Results (Fig 8f) indicates that the landscape had a highly fragmented urban class, which became further fragmented during 80's and started clumping to form a single square in late 90's especially in NE and NW direction in all circle and few inner circles in SE and SW directions, conforming with the other landscape metrics.

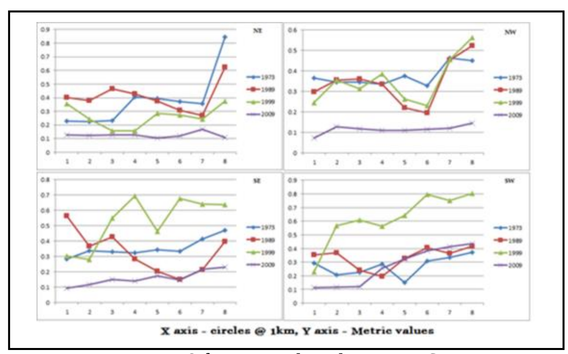


Figure 8.f Zone and circlewise NLSI

Clumpiness index equals 0 when the patches are distributed randomly, and approaches 1 when the patch type is maximally aggregated. Aggregation index equals 0 when the patches are maximally disaggregated and equals 100 when the patches are maximally aggregated into a single compact patch. IJI approaches 0 when distribution of adjacencies among unique patch types becomes increasingly uneven; is equal to 100 when all patch types are equally adjacent to all other patch types.

Clumpiness index, Aggregation index, Interspersion and Juxtaposition Index highlights that the center of the city is more compact in 2009 with more clumpiness and aggregation in NW and NE directions. In 1973 the results indicate that there were a small number of urban patches existing in all direction and in every circle and due to which disaggregation is more. Post 1999 and in 2009 it is observed that large urban patches are located closely almost forming a single patch especially at the center and in NW direction in different gradients (Fig 8g, Fig 8h and Fig 8i).

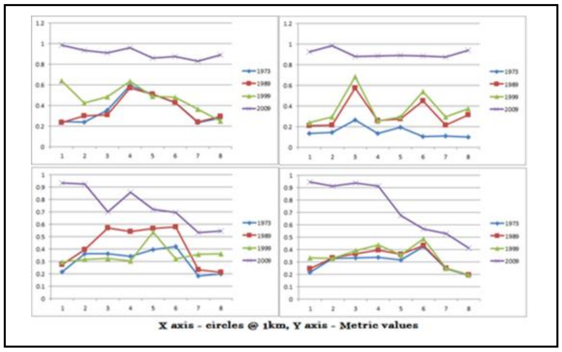
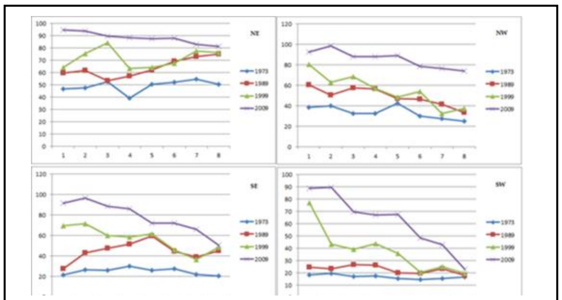


Figure 8.g Clumpiness – zonewise, circle wise



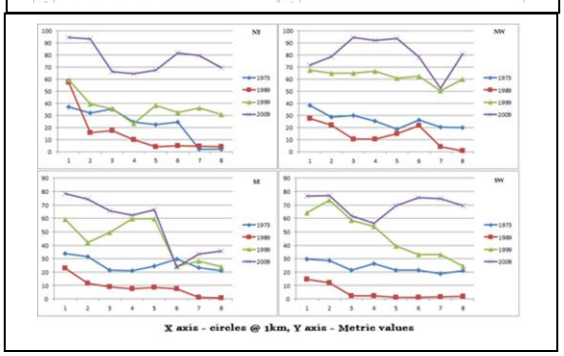


Figure 8.i Zone and circle wise - IJI

Percentage of Like Adjacencies (Pladj) is the percentage of cell adjacencies involving the corresponding patch type those are like adjacent. Cell adjacencies are tallied using the double-count method in which pixel order is preserved, at least for all internal adjacencies. This metrics also indicates the city center is getting more and more clumped with similar class (Urban) and outskirts are relatively sharing different internal adjacencies.

Patch cohesion index measures the physical connectedness of the corresponding patch type. This is sensitive to the aggregation of the focal class below the percolation threshold. Patch cohesion increases as the patch type becomes more clumped or aggregated in its distribution; hence, more physically connected. Above the percolation threshold, patch cohesion is not sensitive to patch configuration [18]. Figure 8k indicate of physical connectedness of the urban patch with the higher cohesion value (in 2009). Lower values in 1973 illustrate that the patches were rare in the landscape.

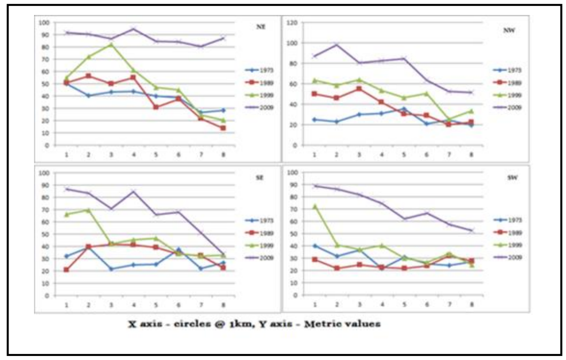


Figure 8.j Zone and circlewise Pladj

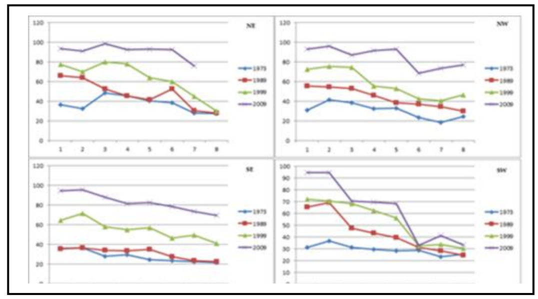


Figure 8.k Cohesion Index

VI. Conclusion

Karnataka government's current focus to develop tier 2 cities in order to decongest major cities, has posed a challenge as unplanned developmental activities is leading to urban sprawl impinging basic amenities to the common man in the outskirts. Availability of spatial data since 1970's has aided in the temporal land use dynamics. Spatial metrics in conjunction with the density gradient approach have been effective in capturing the patterns of urbanization at local levels. The techniques would aid as decision-support tools for unraveling the impacts of classical urban sprawl patterns in Mysore. A set of spatial metrics describing the morphology of unplanned areas have been extracted along with temporal land uses. The extracted indices have indicated the areas of high likelihood of 'unplannedness' considering the three dimensions (size/density/pattern).

Land use assessed for the period 1973 to 2009 using Gaussian maximum likelihood classifier highlight that there has been a significant increase (514%) in urban area, with consequent reduction in vegetation cover. Built up density was minimal and the value ranges from 0.026 (considering 3km buffer) to 0.036 (without considering 3km buffer) in the North east direction (in 1973). Shannon entropy computed using temporal data illustrates that Mysore city is experiencing the sprawl in all directions as entropy values are closer to the threshold. Spatial metrics at landscape level reveal that the landscape had a highly fragmented urban class and started clumping to form a single square in late 90's especially in NE and NW direction in all circle and few inner circles in SE and SW directions, conforming to the other landscape metrics.

Local urban and rural planners need to put forward effective implementable adaptive plans to improve basic amenities in the sprawl localities. Temporal land use analysis along with urban density gradient across four directions has helped in visualizing the growth along with the cultural and industrial evolution.

ACKNOWLEDGEMENT

We are grateful to NRDMS Division, The Ministry of Science and Technology (DST), Government of India and Centre for infrastructure, Sustainable Transportation and Urban Planning (CiSTUP), Indian Institute of Science for the financial and infrastructure support.

REFERENCES

[1] F. Aguilera, R. Talavera. (2009). “Valoración de escenarios futuros a través de la conectividad del Paisaje (Future scenarios assessment through landscape connectivity).” Observatorio Medioambiental 12, pp. 17–36. (in Spanish, abstract version in English)

[2] M. Alberti, P. Waddell. (2000). “An integrated urban development and ecological simulation model.” Int. Assess. 1, pp. 215–227.

[3] N. Batisani, B. Yarnal. (2009). “Urban expansion in Centre County, Pennsylvania: Spatial dynamics and landscape transformations. Applied Geography.” 29(2), pp. 235–249.

[4] S. Berling-Wolf, J. Wu. (2004). “Modelling urban landscape dynamics: a case study in Phoenix, USA.” Urban Ecosyst. 7, pp. 215–240.

[5] J. Brueckner. (2001). “Urban Sprawl: Lessons from Urban Economics,” Unpublished Manuscript, (Department of Economics and Institute of Government and Public Affairs, University of Illinois at Urbana-Champaign, 2000).

[6] M.S. Calvo-Iglesias, U. Fra-Paleo, R.A. Dias-Varela, (2008). “Changes in farming system and population as drivers of land cover and landscape dynamics: the case of enclosed and semi openfield systems in Northern Galicia (Spain)”. Landscape & Urban Plan. 90, pp. 168–177.

[7] R. G. Congalton, (1991). “A review of assessing the accuracy of classifications of remotely sensed data”. Remote Sensing of Environment, 37, pp. 35-46.

[8] R. G. Congalton, R. G. Oderwald, & R. A. Mead, (1983). “Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques.” Photogrammetric Engineering and Remote Sensing, 49, pp. 1671-1678.

[9] N. Currit, W. Easterling, (2009). “Globalization and population drivers of rural-urban land-use change in Chihauhua, Mexico.” Land Use Policy, 26, pp. 535–544.

[10] R.O. Duda, P.E. Hart, D.G. Stork, (2000), “Pattern Classification.” A Wiley-Interscience Publication, Second Edition, ISBN 9814-12-602-0.

[11] A. Firmino, (1999). “Agriculture and landscape in Portugal”. Landscape and Urban Planning. 46, pp. 83–91.

[12] Food and Agriculture Organization, (1995). “Planning for Sustainable Use of Land Resources: Towards a New Approach.” Land and Water Development Division, Rome.

[13] Y. Gao, Liu, (2008). “Mapping land degradation from space: A comparative study of Landsat ETM+ and ASTER data.” International Journal of Remote Sensing, 29, pp. 4029–4043.

[14] J. Geoghegan, L.A. Wainger, N.E. Bockstael, (1997). “Spatial landscape índices in a hedonic framework: an ecological economics analysis using GIS.” Ecol. Econ. 23, pp. 251–264.

[15] K. Gibert, M. Sànchez-Marrè, (2011). “Outcomes from the iEMSs data mining in the environmental sciences work shop series.” Environmental Modelling & Software, 26 (7), pp. 983 -985.

[16] C. E. Gonzalez-Abraham, V. C. Radeloff, R. B. Hammer, T. J. Hawbaker, S. I. Stewart, & M. K. Clayton, (2007). “Building patterns and landscape fragmentation in northern Wisconsin, USA.” Landscape Ecology, 22, pp. 217–230.

[17] B.H. Green, (1990). “Agricultural intensification and the loss of habitat, species and anemity in British grasslands: a review of historical change and assessment of future prospects.” Grass Forage Sci. 45, pp. 365-372.

[18] E.J. Gustafson, (1998). “Quantifying landscape spatial pattern: what is the state of the art?,” Ecosystems 1, pp.143–156.

[19] W.B. Harms, A.H.F. Stottelder, W. Vos. (1984). “Effects of intensification of agriculture on nature and landscape in the netherlands.” Ekologia (CSSR) 3, pp. 281-304.

[20] M. Herold, J. Scepan, K.C. Clarke, (2002). “The use of remote sensing and landscape metrics to describe structures and changes in urban land uses.” Environ. Plann. A 34, pp.1443–1458.

[21] G.D.Jenerette, D. Potere. (2010). “Global analysis and simulation of land-use change associated with urbanization.” Landsc. Ecol. 25 (5), pp. 657–670.

[22] G.D. Jenerette, J. Wu, (2001). “Analysis and simulation of land-use change in the Central Arizona – Phoenix region, USA.” Landsc. Ecol. 16, pp. 611–626.

[23] M. Johnson, (2001). “Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda.” Environment and Planning A, 33, pp. 717-735.

[24] J. Kubes, (1994). “Bohemian agricultural landscape and villages, 1950 and 1990: Landuse, landcover and other characteristics.” Ekologia (Bratislava) 13, pp. 187-l98.

[25] A. Lagarias, 2011. “Land use planning for sustainable development in periurban ecosystems: the case of Lake Koroneia in Thessaloniki, Greece. “In Proceedings of the 2nd international exergy, life cycle assessment and sustainability workshop & symposium, ELCAS 2, Nisyros, Greece.

[26] K.M., Lata, C.H.S. Rao, V.K. Prasad, K.V.S. Badarianth, V. Raghavasamy, (2001). Measuring Urban sprawl A case study of Hyderabad, gis@development, 5(12), pp. 26-29.

[27] Y. Li, S. Zhao, K. Zhao, P. Xie, J. Fang, (2006). “Land-cover changes in an urban lake watershed in a mega-city, central China.” Environmental Monitoring and Assessment, 115, pp. 349-359.

[28] X. Li, Q. Yang, X. Liu, (2008). “Discovering and evaluating urban signatures for simulating compact development using cellular automata.” Landsc. Urban Plann. 86, pp. 177–186.

[29] T. M. Lillesand, R. W. Kiefer, (2005). “Remote sensing and image interpretation.” New York: John Wiley & Sons.

[30] Z. Lipsky, (1992). “Landscape structure changes in the Czech rural landscape.” In: Cultural aspects of landscape. 2nd Int. Con., 1992. IALE, pp. 80-86.

[31] M. Luck, J.G. Wu, (2002). “A gradient analysis of urban landscape pattern: a case study from the Phoenix metropolitan region, Arizona, USA.” Landsc. Ecol. 17, pp. 327–339.

[32] U. Mander, H. Palang, (1994). “Change on landscape structure in Estonia during the Soviet period.” GeoJournal 33, pp.55–62.

[33] K. McGarigal, B. J. Marks, 1995. “FRAGSTATS: spatial pattern analysis program for quantifying landscape structure.” USDA Forest Service General Technical Report PNW-351. www.umass.edu/landeco/research/fragstats/fragstats.html.

[34] J.H.A. Meeus, (1993). “The transformation of agricultural landscapes in Western Europe.” The Science of the Total Environment 129. pp. 171-190.

[35] A. Melluma, (1994). “Metamorphoses of Latvian landscapes during fifty years of Soviet rule.” GeoJournal 33, pp. 55–62

[36] N. E. Mitrakis, C. A. Topalogou, T. K. Alexandridis, J. B. Theocharis, G. C. Zalidis, (2008). “A novel self organising neuro-fuzzy multilayered classifier for land cover classification of a VHR image.” International Journal of Remote Sensing, 29, pp. 4061− 4087.

[37] T. G. Ngigi, R. Tateishi, A. Shalaby, N. Soliman, , M. Ghar, (2008). “Comparison of a new classifier, The mix unmix classifier, with conventional hard and soft classifiers.” International Journal of Remote Sensing, 29, pp. 4111 −4128.

[38] S.K. Pathan, J G. Patel, R.J. Bhanderi. Ajai, Kulshrestha, V.P. Goyal, D.L. Banarjee. S. Marthe, V.K. Nagar, V. Katare, (2004). “Urban planning with specific reference to master plan of lndore city using RS and GIS techniques.” Proc. of GSDI-7 International Conference on

Spatial Data Infrastructure tbr Sustainable Development, held at Bangalore from Feb, 2-6, 2004.

[39] M. Pereira dos Santos Silva, G. Camara , M. Sobral I. Escada, R.C. Modesto de Souza, (2008). “Remote sensing image mining: detecting agents of land-use change in tropical forest areas.” International Journal of Remote Sensing 29 (16), pp. 4803-4822.

[40] M. Phipps, J. Baudry, F. Burel, (1986). “Dynamique de I’organisation ecologique d’un paysage rural: Modalites de la desorganisation dans une zone peri-urbaine.” C.R. Acad. SC. Paris. 303. pp.263-268.

[41] C. Potter, M. Lobley, (1996). “Unbroken threads? Succession and its effects on family farms in Britain.” Sociologia Ruralis 36, pp. 286–306.

[42] T. V. Ramachandra, U. Kumar, (2008). “Wetlands of Greater Bangalore, India: Automatic Delineation through Pattern Classifiers.” Electronic Green Journal, Issue 26, Spring 2008 ISSN: 1076-7975.

[43] C. Renetzedera, S. Schindlera, J. Peterseila, M.A. Prinza, S. Mücherc, T. Wrbka, (2010). “Can we measure ecological sustainability? Landscape pattern as an indicator for naturalness and land use intensity at regional, national and European level.” Ecol. Indic. 10, pp. 39–48.

[44] J.P. Rossi, I. van Halder, (2010). “Towards indicators of butterfly biodiversity based on a multiscale landscape description.” Ecol. Indic. 10, pp. 452–458.

[45] A. Schneider, K.C. Seto, D. Webster, (2005). “Urban growth in Chengdu, Western China: application of remote sensing to assess planning and policy outcomes.” Environ. Plann. B 32 (3), pp. 323–345.

[46] K.C. Seto, M. Fragkias, (2005). “Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics.” Landsc. Ecol. 20, pp. 871–888.

[47] Z. Sha, Y. Bai, Y. Xie, M. Yu, L. Zhang, (2008). “Using a hybrid fuzzy classifier (HFC) to map typical grassland vegetation in Xilin River Basin, Inner Mongolia, China.” International Journal of Remote Sensing, 29, pp. 2317–2337.

[48] T. Simoniello, M.T. Carone, R. Coppola, A. Grippa, M. Lanfredi, M. Liberti, M. Macchiato, (2006). “Preliminary study to monitor land degradation phenomena through landscape metrics. In: 2nd Workshop of the EARSeL SIG on Land Use and Land Cover.” Center for Remote Sensing of Land Surfaces, Bonn, pp. 408 - 414.

[49] H.S. Sudhira, T.V. Ramachandra, K.S. Jagadish, (2004). “Urban sprawl: metrics, dynamics and modelling using G1S.” Int. J Applied Earth Observation and Geoinformation,' 5, pp. 29-39.

[50] P. Torrens, M. Alberti, (2000). “Measuring sprawl. CASA working paper series 27.” UCL. http://www.casa.ucl.ac.uk/publications/workingpapers.asp.

[51] P. M. Vitousek, (1994). “Beyond global warming: ecology and global change.” Ecology 75, pp. 1861-1876.

[52] R. Wood, J. Handley, (2001). “Landscape dynamics and the management of change.” Landscape Research, 26 (1), pp. 45-54.

[53] X. J. Yu, C.N. Ng, (2007). “Spatial and temporal dynamics of urban sprawl along two urban-rural transects: a case study of Guangzhou, China.” Landsc. Urban Plann. 79, pp. 96–109.

[54] Q. Zhang, J. Wang, X. Peng, R. Gang, P. Shi, (2002). “Urban built-up land change detection with road density and spectral information from multi- temporal Landsat TM data.” Int. J. Remote Sensing, 23(15), pp. 3057-3078.

AUTHORS PROFILE

Dr. Ramachandra has Ph.D. in energy and environment from Indian Institute of Science. At present, Coordinator of Energy and Wetlands Research Group (EWRG), Convener of Environmental Information System (ENVIS) at Centre for Ecological Sciences (CES), Indian Institute of Science (IISc). He has made significant contributions in the area of energy and environment. His research area includes wetlands, conservation, restoration and management of ecosystems, environmental management, GIS, remote sensing, regional planning and decision support systems. During the past ten years he has established an active school of research in the area of energy and environment. He teaches principles of remote sensing, digital image processing and Natural resources management. He has published over 184 research papers in reputed peer reviewed international and national journals, 118 papers in the international and national symposiums as well as 14 books. In addition, he has delivered a number of plenary lectures at national and international conferences. He is a fellow of Institution of Engineers (India), IEE (UK), Senior member, IEEE (USA) and many similar institutions. Details of his research and copies of publications are available at
http://ces.iisc.ernet.in/energy/
http://ces.iisc.ernet.in/grass

Bharath H Aithal, Electrical and Electronics Engineering graduate from Bangalore University. Currently, he is pursuing Ph.D at Indian Institute of Science. His area of interest are spatial pattern analysis, Urban growth modeling, natural disasters, geoinformatics, landscape modeling and urban planning, open source GIS, digital image processing.

Sreekantha S has M Sc in Remote Sensing and GIS and currently working as project trainee at Energy and Wetlands Research Group (EWRG), CES, IISc. His area of interest are urban sprawl analysis, seismic zonation, remote sensing.

ELECTRE-Entropy method in Group Decision Support System Modelto Gene Mutation Detection

Ermatita
Department of Information System , Computer Science Faculty of SriwijayaUniversity,Palembang, Indonesia
(Doctoral Program Student Computer Science GadjahMadaUniversity)

Sri Hartati
Departement of Computer Science and Electronics Mathematics and Natural Sciences Faculty GadjahMada University, Indonesia

Retantyo Wardoyo
Departement of Computer Science and Electronics Mathematics and Natural Sciences Faculty GadjahMada University, Indonesia

Agus Harjoko
Departement of Computer Science and Electronics Mathematics and Natural Sciences Faculty GadjahMada University, Indonesia

***Abstract*—Application of Group Decision Support System (GDSS) can assist for delivering the decision of various opinions (preference) cancer detection based on the preferences of various expertise. In this paper we propose ELECTRE-Entropy for GDSS Modeling. We propose entropy weighting for each criteria under ELECTRE Method.ELECTRE is one method in Multi-Attribute Decision Making (MADM). Modeling of Group Decision Support Sytemapplyfor multi-criteria which the simulation data mutated genes that can cause cancer and solution recommended.**

Keywords-component; Group decision support system(GDSS); Multi Atributte Decision making(MADM); Electre-entropy; preference.

I. INTRODUCTION

To Determine wheather a person has abnormal gene of cancer-causing can be done from different areas of expertise in medical science, such as pathologist, oncology or other disciplines in medicine. The opinions in various expert in medical science requires decision that could provide treatman provision against person alleged having abnormal genes as the cause of cancer. Group Decision Support System is one of application in information technology that can assist in delivering the decision from various opinion (preferences) for detecting person having mutated gene that causes cancer based on the preferences of various expertise.

The decision making process requires the aggregation method to get single value of each alternative from variety of criteria. In the decision-making system, this problem can be solved by Multiple Criteria Decision Making (MCDM). The study of MCDM has begun to emerge in the late 19th century. But the very rapid development of new beginning to be felt since the 1970s, especially in the field of operations research (Suiran, et al; 2001; Sage, 1991).

This study establish Clinical Model Group Decision Support System (CGDSS) where knowledge base is built based on preferences that differ from the experts of different expertise from the classification genes with SVM method uses the concept of multi-attribute decision making (MADM).

II. BACKGROUND THEORIES

A. Group decision support system (GDSS)

Group decision support system (GDSS) is an interactive computer-based system that facilities the solution of semistructured or unstructured problems by a group of decision maker.[11]

GDSS has several major characteristics,i.e.:

- GDSS has goal to support the process of group decision makers by providing automation of subprocesses, using information technology tools.
- GDSS is specially designed information system, not merely a configuration of already existing system components. It can be designed tobe address one type of problem or a variety of group-level organizational decisions.
- GDSS encourages generation of ideas, resolution of conflicts, and freedom of expression. It contains built-in mechanismes that discourage development of negative group behaviors, such as destructive conflict, miscommunication, and groupthink.(Turban,2011)

ELECTRE (Elimination EtChoixTraduisant la realite) is one method in MADM based on the concept of ranking through pairwise comparisons between alternatives on the appropriate criteria. An alternative is said to dominate another alternative if one or more of the criteria are exceeded (compared with the other criteria of alternative) and the same with the remaining criteria. The relationship between the two alternatives Ak ranking of the A1 (Roy, 1973)in [4]

MADM is evaluated against the alternative m Ai (i = 1.2,...., m) of a group of attributes or criteria c, (j = 1.2,..., n) where each of the attributes are not mutually dependent on one other. Decision matrix of each alternative on each attribute, X is given as:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \qquad (1)$$

The Procedure of using ELECTRE is shown as follows :

1) In ELECTRE. First the weightted normalized impact matrix of v_{ij} is computed to provide a dimensionless environment. This step is the same as the first two steps describes in:

$$V_{ij} = w_j x_{ij} \qquad (2)$$

2) The concordance set cij, is determined for each pair of alternatives j an j*i.e the setr of criteria for attributes in which the impact of alternative j is better than or equal to that of alternative j*. Similarly, a discordance set Dij is define which compares two alternatives in which alternative i perform worse tha alternative i:

$C_{kl} = \{ j \mid v_{kj} \geq v_{ij} \}$ untuk j = 1,2,...,n (3)

And $d_{kl} = \{ j \mid v_{kj} < v_{ij} \}$ untuk j = 1,2,...,n

Where vkj and vij are impact values with the ith criterion and l is the set of atributes

3) Once the concordance and disconcordance sets are found, concordonce (c_{ij}) and disconcordance (d_{ij}) indices can be calculated respectively. The concordance index is equal to the sum of the weights associated with the ith attribute which are contained in the concordance set. Hence, the formula is shown as follows :

$$C_{kl} = \sum_{j \in c_{kl}} w_j \qquad (4)$$

Where w_i is the weight of the i^{th} attribute and 0<cij<1 .the concordance index reflects the relative importance of alternative j*. A higher value of c_{ij}, indicates that alternatives j is preferred to j* as far as the concordance attributes are concerned. In addition, disconcordanceindex(dij)can be calculated such that:

$$d_{kl} = \frac{\max \{ v_{kj} - v_{ij} \}_{j \in d_{kl}}}{\max \{ |v_{kj} - v_{ij}| \}_{j \in v_j}} \quad (5)$$

Where vkj and vij are the data in normalized impact matrix and 1 is the set of attributes.[19]

Concordance matrix calculated based on the dominant [10]

$$f_{kl} = \begin{cases} 1, & jika \ c_{kl} \geq \underline{c} \\ 0, & jika \ c_{kl} < \underline{c} \end{cases} \qquad (6)$$

elements of the matrix F is determined as the dominant discordance:

$$g_{kl} = \begin{cases} 1, & jika \ d_{kl} \geq \underline{d} \\ 0, & jika \ d_{kl} < \underline{d} \end{cases} \qquad (7)$$

Aggregation of the dominant matrix (E) showing a partial preference order of alternatives, obtained with the formula in mathlab:

(1.1)

$e_{kl} = f_{kl} . g_{kl}$ (8)[4]

B. Entropy

Entropymethodcan be usedtodeterminetheweights. Entropyconsistencyininvestigatingdiscriminationamonga set of data.Alternativeset of datavaluesoncertaincriteriadescribedin theDecisionMatrix(DM). Using theentropymethod, thevariationvaluecriterionwill be canhighestweights. [12]

The measuresusedin this methodare asfollows:

- Creating criterion data table The criteriacan beidentifiedqualitative andquantitativecriteria, but allmustbemeasurable.
- Normalizationof datatableof criteria

$$d_i^k = \frac{x_i^k}{x_{i\ maks}^k} \qquad d_i = d^1_i, \ldots, d^m_i$$

$$D_i = \sum_{k=1}^{m} d_i^k \quad i = 1,2,\ldots,n \qquad (9)$$

After getting Entropy weighting for each criterion, if it has no initial weight or weight of a predetermined weight of Entropy then the truth to each criterion will be obtained with the following calculation

$$\lambda_i = \frac{\overline{\lambda_i} \times w_i}{\sum_{i=1}^{n} \overline{\lambda_i} \times w_i} \qquad i = 1, \ldots, n \qquad (10)$$

C. Copeland score

The Copeland Score is more likely to produce ties, since it does not take into account the margin of victory, or the magnitude of support. In some contests, a Copeland Score will not identify a clear winner and provide only a limited differentiation between the options. In a three-way contest with no Condorcet winner and no ties in the binary contests, all three candidates will have the same Copeland score (each will have 1 win, 1 loss). In a four-way contest without a Condorcet winner, there will be at best a two-way tie (2 wins, 1 loss each). Consequently, many contests will need some secondary mechanism to resolve contests which end in a tie. [28]

Population	Preferences
45%	a d b c
40%	b a d c
15%	c b a d

Preference profiles

Contest	Winner
a vs. b	b
a vs. c	a
a vs. d	a
b vs. c	b
b vs. d	b
c vs. d	d

Pair-wise contests

Alternative	Copeland score
a	2 − 1 = 1
b	3 − 0 = 3 *
c	0 − 3 = −3
d	1 − 2 = −1

Voting Results

Figure 1. Copeland Score

III. ELECTRE-ENTROPY METHOD FOR GROUP DECISION SUPPORT SYSTEM MODEL GENE MUTATION DETECTION

Process in group decision support system begins with this clinical classification process, the process of ranking the decision, then made a decision ranking of the ranking recommendations of each expert. This system has three engine block in the decision-making systems like the picture below

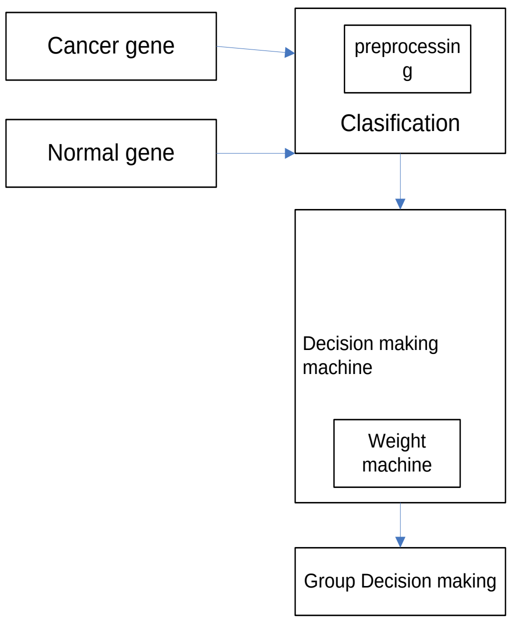


Figure 2. Architecture model of GDSS (Group Decision Support System) for Decision Making.

Materials are processed in the architecture of this system is the data model of cancer gene available on the Internet.

The data that exists is shaped gene sequences that would later be classified into training data and testing data.

Components in the engine builders are composed of:

1) Classification engine. This component is assigned toperform the classification of the existing gene data. In thismachine used the method of classification with SupportVector Machine method.

2) Machine Builders Decision. These components perform ranking and data processing as preferences given by the experts. The model built in the engine builders these decisions using Multi Attribute Decision Making with ELECTRE methods.

3) Weighting machine. Decision builder method that does the determination of the ideal weight in development decisions. The resulting weights will be used in the ranking of each expert's decision.

4) Decision maker. This component is the decision maker has been obtained from each expert. Recommendations in order to get better decisions. In this machine used method of Copeland score.

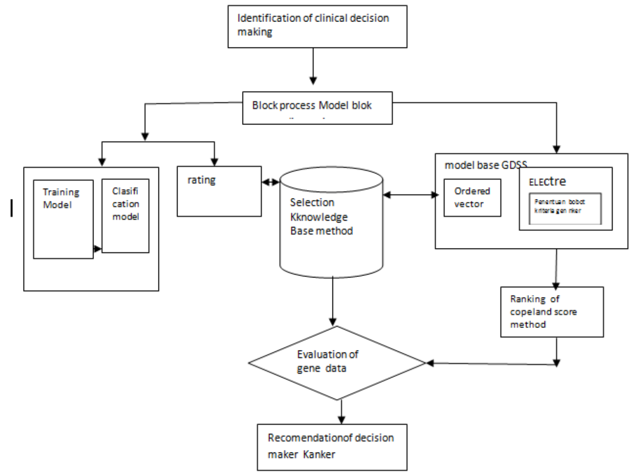


Figure 3. CGDSS Model: Modeling Preferences and skills base on Classification as Model in Group Decision Support System for Decision Making (adapted from Deng, 2008)

A. *Classification*

Protein sequence data from the example above would convert the system into the strands of the DNA sequence consisting of A, G, C and T. DNA strands will be made in two dimensions in the number of A and T number of G and C. Then do the process of classification with Support Vector Machine method. The conversion process is done by creating a numerical code for each letter in the DNA sequence code. Values that have been in this conversion will be processed by the method of classification by support vector machine, getting class genes of normal and abnormal.

In Support Vector Machine, best Hyperplane / clasifier is located in the middle between the two groups of objects from two classes, namely class of normal genes and gene abnormalities. In the classification process to maximize this margin, the system will determine the class of data to be ranked in the training class, after being found clasifier. Performance is good, It will be set classroom testing. If the class already exists then this testing will be used other data to be processed classification. Looking for the best hyperplane is to maximize the margin or distance between the two groups of objects from different classes of genes.

B. *Model*

Group Decision Support System base with ELECTREmethod-Entropy Classification results will be analyzed by specialists. This analysis will refer to the provision of suitability rating on each criterion for the count with ELECTRE method. This modeling begins with the preparation component of alternatrive the situation in the identification. The objectives of the component preparation is to construct a table of estimated components of the situation and identification of alternatives, specification of objectives, criteria and attributes. This model is used to evaluate alternative m Ai (i = 1,2 ,......, m) against a set of attributes or criteria Cj (j = 1,2 ,....., n) and attributes are not mutually dependent each other.

Decision matrix of is built each alternative on each attribute, X. Preparation of the components on this modeling is used to detect gene mutations in humans to determine whether there is through virus, nutrition or foreign object. So, the model can detect whether a person is identified to have cancer cells or not.

In this model simulation there are three alternatives that can be set to identify cancer cells in human genes, namely:

A1 = mutation because the virus
A2 = Mutations for Foreign Objects
A3 = Mutations for Nutrition
There are 3 that a reference in making decisions which are:
C1 = protein bound to viral
C2 = Expression HSP
C3 = proteisn expression containing CH_3

Rating the suitability of each alternative on each criterion, the value of 1 to 5, namely:1 = very bad,2 = poor,3 = quite,4 = Good and 5 = Very good. Level of importance of each criterion in value by 1 to 5, namely:
1 = very low,2 = Low,3 = quite,4 = High and 5 = very high The values given by experts in each alternative on each criterion is the value of a match. Suitability value is simulated as in the following table:

Our model begins by establishing a paired comparison of each alternative on each criterion (xij) which are being formulated in a matrix X as a decision matrix. Xij is a performance rating of alternative i-th j-th attribute. Then we use Copeland score method in making final decision.

Classification block. This component is assigned to perform the classification of the existing gene data. In this block we use Support Vector Machine method. The classification system machine data in the form of the gene will be read from the database. The data in the form of the gene sequences of DNA will be in the formula calculating the number of A and T, as well as the number of G and C. A and T will be in put in one dimension, then G and C are grouped in one dimension.

Data in this dimension has been produced by the process of determining the classification engine to perform training classes. The process will continue until the ideal performance close to 100% using classification performance. If this figure is already approaching the ideal performance, then the testing class will be performed. Testing this class that will test the incoming data so it will be grouped into normal and abnormal classes. The figure above is result classification of SVM

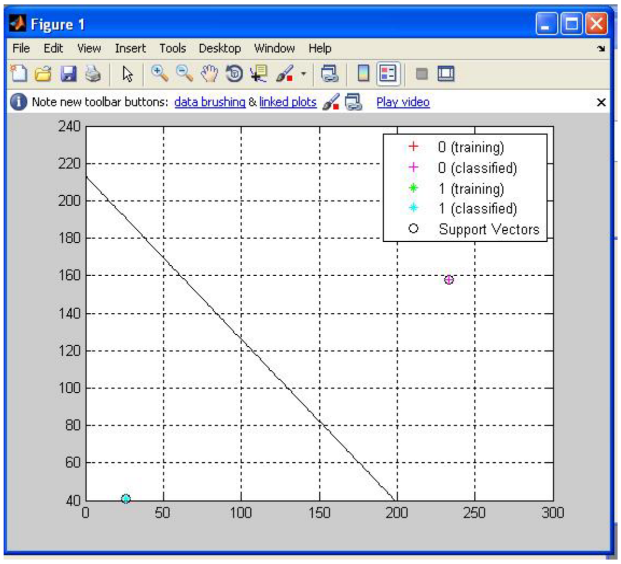


Figure 4. result classification in SVM method

```
global data1 CP
[a b c d]=mysql('select * from gen');f=[];e=[];
fori=1:length(b)
e(i)=sum(ismember(c{i},'T'))+sum(ismember(c{i},'A'));
f(i)=sum(ismember(c{i},'G'))+sum(ismember(c{i},'C'));
end
species=d;e=e';f=f';
data=[e f];
groups = ismember(species,'Y');
[train, test] = crossvalind('holdOut',groups);
cp = classperf(groups);
figure(1);
svmStruct=svmtrain(data(train,:),groups(train),'showplot',true)
;
gridon;
classes = svmclassify(svmStruct,data(test,:),'showplot',true);
classperf(cp,classes,test);
CP=cp.CorrectRate;
figure(cp1);
data1=[b(test,:) c(test,:) d(test,:) num2cell(e(test,:))
num2cell(f(test,:))];
figure(lihat);
```

Block Decision. These components perform ranking and processing of data is a preference that is given by the experts.

The model built in the engine builders these decisions using Multi Attribute Decision Making with ELECTRE methods. The system in this machine will accept input in the form of rating the suitability and weights based on the interests of the experts with an ordered vector format. Each expert provide the weight of each of the criteria based on interests. The ideal weight then calculated with entropy in ELECTRE method. We then can obtain alternative ranking in table below:

TABLE I. SUITABILITY OF EACH ALTERNATIVE ON EACH CRITERION

Alternative	criteria		
	C1	*C2*	*C3*
A1	4	4	5
A2	4	5	4
A3	4	3	5

TABLE II. SUITABILITY OF EACH ALTERNATIVE ON EACH EXPERT

Alternative	criteria		
	C1	*C2*	*C3*
P1	4	1	1
P2	3	1	1
P3	1	3	3
P4	1	1	1

Result of calculation data for normalization is :

$$R= \begin{matrix} 0.4575 & 0.6860 & 0.7071 \\ 0.7625 & 0.5145 & 0.5657 \\ 0.4575 & 0.5145 & 0.4243 \end{matrix}$$

And result of entropy calculation is :

$$W= \begin{matrix} 0.3734 & 0.3133 & 0.3133 \\ 0.3734 & 0.3133 & 0.3133 \\ 0.3734 & 0.3133 & 0.3133 \end{matrix}$$

Implementation on matlab the method like follow :

```
%matriks_c = matrix-c
c_gabungan = c_gabungan.*w;
c_gabungan = sum(c_gabungan');
c_gabungan = (reshape(c_gabungan,m,m))';
matriks_c = c_gabungan;
rata_c = sum(sum(matriks_c)) / (m*(m-1));
rata_c = repmat(rata_c,m,m);
matriks_fkl = matriks_c>= rata_c;

%matriks_d
matriks_d = zeros(m,m);
[p,q] = size(d_gabungan);
fori=1:p,
    d = d_gabungan(i,:);
pos_d = find(d==1);
if (isempty(pos_d) == 0),
        [r,s] = size(pos_d);
pos_x = ceil(i./m);
pos_y = mod(i,m);
if (pos_y==0),
pos_y = m;
end;
for j=1:r,
            s = V(pos_x,pos_d(j,1)) -
V(pos_y,pos_d(j));
            s = abs(s);
matriks_d(pos_x,pos_y) =
max([matriks_d(pos_x,pos_y) s]);
end;
        s = max(abs(V(pos_x,:)-
V(pos_y,:)));
matriks_d(pos_x,pos_y) =
matriks_d(pos_x,pos_y)./s;
end;
end;
savedvfkl.matmatriks_dVmatriks_fkl
%matriks_d
rata_d = sum(sum(matriks_d)) / (m*(m-1));
rata_d = repmat(rata_d,m,m);
matriks_gkl = matriks_d>= rata_d;
matriks_e=matriks_fkl.*matriks_gkl;
hasil = matriks_e;
savegkl.matmatriks_gkl
```

Result of calculation ELECTRE-Entropy method and voting in copeland score as follow:

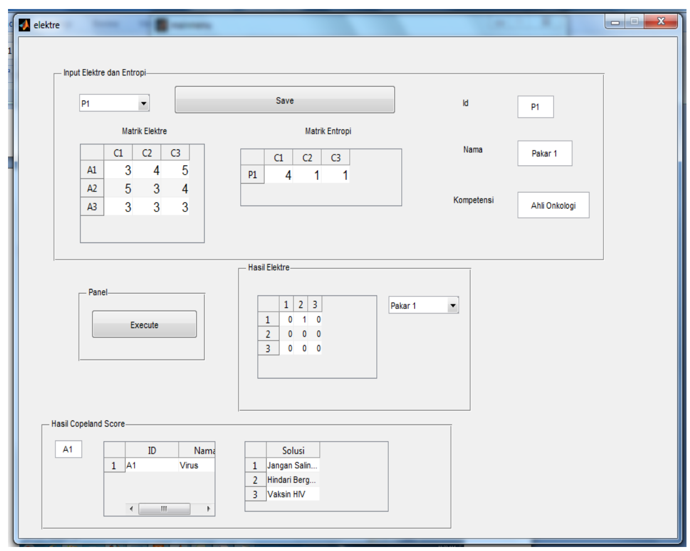


Figure 5. Result of ELECTRE-entropy method and copeland score voting.

The results of calculations with ELECTRE-Entropy modeling with voting Copeland scorec an result in a vote with a value of A1. Figure. 5 copeland score results show tha talternative 1 (A1) is dominant. The result show that the Alternative A1 is recommended as a result of group decisions and the solution recommended.

IV. CONCLUSIONS

ELECTRE-Entropy modeling method can be helpful in determining alternatives rank.. This modeling can be applied to several other conditions of a similar case.

Calculation of weighting with Entropy method using the preference ofe ach expert candetermine the ideal value of modeling so that can provide recommendations for producing better decisions.

ACKNOWLEDGMENT

This work is a continuation of the paper title "Implementation of MADM Methods in Solving Group Decision Support System on Gene Mutations Detection Simulation" Presented at ICCMS at Mumbai and ELECTRE solving Methods in Bioinformatics Group decision support system on gene mutation detection simulation, published on International Journa Computer Science and InformationTechnology.

REFERENCES

[1] Amiri, "Developing a New ELECTRE Method with Interval Data in MultipleAttribute Decision Making Problems" 2008

[2] Bashiri, "An Extension of Multi-Response Optimization In MADM view", Journal of applied Sciences 9(9); page 1695-1702, 2009

[3] R. Fitriadi, "Pendekatan Compromise Programming dengan memperhitungkan Faktor lingkungan (Studi Kasus Industri Otomotif PT"XX" Jawa Tengah)", Jurnal Ilmiah Teknik Industri Vol. 5 No. 2 Des 2006, hal 72 – 81, 2006

[4] S. Kusumadewi, S. Hartati, A. Harjoko, dan R. Wardoyo, "Fuzzy Multi-AttributeDecision Making (FUZZY MADM)", Yogyakarta: Penerbit Graha Ilmu, 2006

[5] S. Opricovic, G.H. Tzeng, "Extended VIKOR method in comparison with outranking methods" , European Journal of Operational Research 178 (2007) 514–529, 2007

[6] A. Prayitno, et All, "The expression of p53, Rb, and c-myc protein in cervical cancer by immunohistochemistry stain", Biodiversitas ISSN: 1412-033X Volume 6, Nomor 3 Juli 2005 Halaman: 157-159, 2005

[7] Soltanmohammadi, "Achieving to some outranking relationships between post mining land uses through mined land suitability analysis", Int. J. Environ. Sci. Tech., 5 (4), 535-546, Autumn 2008 ISSN: 1735-1472© Irsen, Ceers, IAU, 2008

[8] E. Turban, "Dicision Support and Expert Systems: Management Support Systems", Fourth Edition, Prentice-Hall,Inc., United State, 2005

[9] Q. Zhang and J. Ma, "Determining Weights of Criteria Based on Multiple Preference Formats", online pada http://www.is.cityu.edu.hk/Research /WorkingPapers/paper/0102.pdf12 Oktober 2004, 2004

[10] Ermatita, et All, MADM Methods in Solving Group Decission support sytem on Gene Mutation Detection simulation, Proceedings The second International Conference on DfmA 2010, ISBN:978-602-9747-9-0-4, August 2010, UGM, Yogyakarta

[11] Turban ,E,et All,2011, Decision Support and Business intelligence Systems, ninth Edition, Pearson education, Inc., publishing as Prentice Hall, new jersey, USA

[12] Vivi Triyanti dan M.T Gadis, 2008, Pemilihan Supplier Untuk Industri Makanan Menggunakan Metode Promethee, Journal of Logistics and Supply Chain Management, Vol 1, No. 2, June 2008: 83-92,jurnal.pdii.lipi.go.id/admin/jurnal/12088392.pdf, di akses 24 Mei 2011

[13] Lontar, 2011 Fakultas kedokteran Gigi Universitas Indonesia, www.lontar.ui.ac.id/file?=file digital/125719-Rv-om, Jakarta

[14] Guan.P, Huang.D, He. M and Zhou.B, 2009, Lung cancer gene expression database analysis incorporating prior knowledge with support vector machine-based classification method, Journal of Experimental & Clinical Cancer Research 2009, 28:103 doi:10.1186/1756-9966-28-103, http://www.jeccr.com/content/28/1/103

[15] Lee .MC, Jui-Fang Chang2, Jung-Fang Chen .JF,2011, An Entropy Decision Model for Selection of Enterprise Resource Planning System, International Journal of Computer Trends and Technology- March to April Issue 2011, ISSN:2231-2803, IJCTT,pp1-9

[16] Gong.J , Xu.J, 2006, The integration of valued outranking relations in ELECTRE methods for ranking problem, ISSN 1 746-7233, England, UK

[17] World Journal of Modelling and Simulation.Vol. 2 (2006) No. 1, pp 3-14, di akses 12 September 2011 citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.7500&rep...

[18] Karacu.M, Arslan.T, 2010, Electre Approach for modeling Public Decision making behavior on Transportation Project Selection Process, 12th WCTR, July 11-15, 2010 – Lisbon, Portugal , intranet.imet.gr/Portals/0/UsefulDocuments/documents/03182.pdf

[19] Sevkli.M, 2010, An application of the fuzzy ELECTRE method for supplier selection, International Journal of Production Research, Vol. 48, No. 12, 15 June 2010, 3393–3405

[20] Cheng and regina, 2000, Development of a Fuzzy Multi criteria Decision support system for manucipal Solid Waste management, A thesis submitted to the faculty of graduate studies and research for degree of master of Applied Science in advance manufacturing and Production systems University of Regina, Kanada

[21] Zandi,F , Tavana.M and Martin.D, 2011,A fuzzy group Electre method for electronic supply chain management, International Journal of Logistics Research and Applications framework selection, International Journal of Logistics: Research and Applications Vol. 14, No. 1, February 2011, 35–60 Publication details, including instructions for authors and subscription information:

[22] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[23] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[24] K. Elissa, "Title of paper if known," unpublished.

[25] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[26] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[27] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[28] Gavish, B. and Gerdes, J.H., 1997, Voting Mechanisms and Their Implications in A Gdss Environment, Annals of Operations ResearchScience Publisher

[29] Deng H dan Wibowo, S, 2008, A Rule-Based Decision Support System for Evaluating and Selecting IS Projects ,Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol II IMECS 2008, 19-21 March, 2008, Hong Kong

AUTHORS PROFILE

Ermatita. Lecturer Sriwijaya University Palembang Indonesia.

Sri Hartati, Lecturer Gadjah Mada Univeersity Yogyakarta Indonesia.

Retantyo Wardoyo, lecturer Gadjah Mada University Yogyakarta Indonesia.

Agus harjoko Lecturer Gadjah Mada University Yogyakarta Indonesia.

The Solution of Machines' Time Scheduling Problem Using Artificial Intelligence Approaches

Ghoniemy S.
Computer Engineering Dept.,
Faculty of Computers and Information Technology, Taif University
Taif, KSA.

El-sawy A. A.
Computer Science Dept.,
Faculty of Computers and Information Technology, Taif University
Taif, KSA.

Shohla M. A.
Computer Engineering Dept.,
Faculty of Computers and Information Technology, Taif University
Taif, KSA.

Gihan E. H. Ali
Mathematics Dept.,
Faculty of Science and Education, Khurma Branch, Taif University,
Taif, KSA.

***Abstract*—The solution of the Machines' Time Scheduling Problem (MTSP) is a hot point of research that is not yet matured, and needs further work. This paper presents two algorithms for the solution of the Machines' Time Scheduling Problem that leads to the best starting time for each machine in each cycle. The first algorithm is genetic-based (GA) (with non-uniform mutation), and the second one is based on particle swarm optimization (PSO) (with constriction factor). A comparative analysis between both algorithms is carried out. It was found that particle swarm optimization gives better penalty cost than GA algorithm and max-separable technique, regarding best starting time for each machine in each cycle.**

Keywords- Machine Time Scheduling; Particle swarm optimization; Genetic Algorithm; Time Window.

I. Introduction

A great deal of research has been focused on solving scheduling problems. One of the most important scheduling problems is the Machine Time Scheduling Problem (MTSP). This problem was investigated in [1] as a parameterized version of the MTSP, which was defined in [2], with penalized earliness in starting and lateness in the completion of the operation. The authors in [1] applied the optimal choice concept which is given in [3] and some theoretical results from [4] to obtain the optimal values of the given parameters.

In [5] the authors investigated two cycles MTSP and introduced an algorithm to find the optimal choice of parameters, which represent the earliest possible starting time for the second cycle. In [6] an algorithm was developed (MTSP Algorithm (MTSPA)) for multi-cycle MTSP which found the starting time for each machine in each cycle by using the max-separable technique.

The processing times in the previous researches were deterministic. The authors in [7] discussed how to solve the MTSP when the processing time for each machine is stochastic. To solve this problem, the Monte Carlo simulation was suggested to handle the given stochastic processing times. A generalization was introduced in [8] to overstep the cases at which an empty feasible set of solutions is described by the system.

This paper examines two approaches for solving the MTSP; PSO (with constriction factor), and GA (with non-uniform mutation). A comparative analysis between both algorithms is to be carried out for the solution that minimizes the penalty cost regarding the best starting time for each machine in each cycle.

The paper is organized as follows. Part 2 formulates the problem. Part 3 introduces the proposed GA, as well as its implementation. Part 4 presents the PSO Algorithm, and its implementation. Both algorithms, in addition to the max-separable technique, are applied for a specific scenario. Obtained results are investigated in Part 5. The paper is terminated by conclusions and proposals for future work.

II. Problem Formulation

In machines' time scheduling problem there are n machines, each machine carries out one operation j with processing time p_j for $j \in N = \{1,\ldots,n\}$ and the machines work in k cycles.

Let xjr represent starting time of the jth machine in cycle r for all $j \in N$, $r \in K = \{1,\ldots,k\}$ (k number of cycles). Machine j can start its work in cycle r only after the machines in a given set $N^{(j)}, N^{(j)} \subset$ N ($N^{(j)}$ is the set of precedence machines) had finished their work in the (r-1)th cycle, so we can define the starting time in the (r+1)th cycle as follows:

$$x_{ir+1} \geq \max_{j \in N^{(i)}} (x_{jr} + p_{jr}) \quad \forall i \in N, \forall r \in K$$

Assuming that the starting time x_{jr} is constrained by a time interval $[l_{jr}, L_{jr}]$ for each $j \in N$, $r \in K$, then the set of feasible starting times x_{jr} is described by the following system for each $r \in K$:

$$\max_{j \in N^{(i)}} (x_{jr} + p_{jr}) \le x_{ir+1} \quad \forall i \in N,$$

$$l_{jr} \le x_{jr} \le L_{jr} \qquad \forall j \in N \qquad (1)$$

Assume also that for some echological reasons, there are a given recommended time interval *$[a_{jr}, b_{jr}]$*, $\forall i \in N$, $\forall r \in K$ such that:

$$[x_{jr}, x_{jr} + p_j] \subset [a_{jr}, b_{jr}], \qquad (2)$$

The violation of (2) will be penalized by the following penalty function

$$f(x) = \max_{j \in N} f_{jr}(x_{jr}) \ \to \min \qquad r \in K$$

where the penalty function in a certain cycle r is given by:

$$f_{jr}(x_{jr}) = \max\{ f_{jr}^{(1)}(x_{jr}), f_{jr}^{(2)}(x_{jr} + p_{jr}), 0\} \ \forall j \in N$$

where $f_{jr}^{(1)}$: R→ R is a decreasing continuous function such that $f_{jr}^{(1)}(a_{jr}) = 0$,

and $f_{jr}^{(2)}$: R→ R is an increasing continuous function such that $f_{jr}^{(2)}(b_{jr}) = 0$

To minimize the maximum penalty in each cycle r, we should solve the following problem:

$$f(x) \to \min$$

$$subject\ \ to:$$

$$\max_{j \in N^{(i)}} (x_{jr} + p_{jr}) \le x_{ir+1} \ \forall i \in N \qquad (3)$$

$$l_{jr} \le x_{jr} \le L_{jr} \qquad \forall j \in N$$

III. PROPOSED APPROACHES FOR SOLVING THE MTSP

A. *Using Genetic Algorithm (GA):*

GA maintains a set of candidate solutions called population and repeatedly modifies them. At each step, the GA selects individuals from the current population to be parents and uses them to produce the children for the next generation. Candidate solutions are usually represented as strings of fixed length, called chromosomes. A fitness or objective function is used to reflect the goodness of each member of population [9]. The principle of genetic algorithms is simple [10]:

- Encoding of the problem in a binary string.
- Random generation of a population. This one includes a genetic pool representing a group of possible solutions.
- Reckoning of a fitness value for each subject. It will directly depend on the distance to the optimum.
- Selection of the subjects that will mate according to their share in the population global fitness.
- Genomes crossover and mutations.
- And then start again from point 3.

This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied.

Non-uniform mutation has been used to reduce the disadvantage of random mutation in the real-coded GA [11]. This new operator is defined as follows. For each individual X_i^t in a population of tth generation, create an offspring X_i^{t+1} through non-uniform mutation as follows: if $X_i^t = \{x_1, x_2, ..., x_m\}$ is a chromosome (t is the generation number) and the element xk is selected for this mutation, the result is a vector $X_i^{t+1} = \{x_1', x_2', ..., x_m'\}$ where

$$x_k' = \begin{cases} x_k + \Delta(t, UB - x_k) & if\ a\ random\ \xi\ is\ 0. \\ x_k - \Delta(t, x_k - LB) & if\ a\ random\ \xi\ is\ 1. \end{cases} \qquad (4)$$

and LB and UB are the lower and upper bounds of the variables x_k. The function $\Delta(t, y)$ returns a value in the range [0,y] such that $\Delta(t, y)$ approaches to zero as t increases. This property causes this operator to search the space uniformly initially (when t is small), and very locally at later stages. This strategy increases the probability of generating a new number close to its successor than a random choice. We use the following function:

$$\Delta(t, y) = y.(1 - r^{(1-\frac{t}{T})^b}). \qquad (5)$$

Where r is a uniform random number from [0, 1], T is the maximal generation number, and b is a system parameter determining the degree of dependency on the iteration number.

In [12], the authors introduced a new mutation operator characterized by its non-uniformness. The operator has been used in the parameter optimization of the controllers of a supply ship. Four different kinds of controllers have been considered and optimized, providing a wide range of optimization problems with their own unique search spaces to test the mutation operator.

The main steps for solving the MTSP using GA are as follows:

1) *Reformulation:*

Each machine boundaries will be reformulated (calculate the new boundaries) based on its' successors machines boundaries. For each machine, the new lower boundary is called h and the new upper boundary is called H.

2) *Initial population:*

First, the chromosome is defined as a set of starting times for the machines in all cycles. So, S is the size of the chromosome is equal to n multiplied by k (n number of machines and k number of cycles). The gene x_{ircp} is the starting times for machine i in cycle r in chromosome c in population p

(Q number of chromosomes and W number of population) which satisfy the constraint in (P_r). The value of x_{ircp} is generated randomly where $h_{ir} \le x_{ircp} \le H_{ir}$.

3) Other generation:

The value of fitness for each chromosome in the previous population had been calculated. The next generation is created by selecting the best chromosomes which have the greatest value in the objective function. T is a percent of the previous population which determines the number of best chromosomes that transfer to the next generation.

4) Crossover:

The remaining chromosomes are divided as a pair. Each chromosome in each pair is divided in certain cycle v then swap between v to k cycle in first chromosome and v to k cycles in the second chromosome.

5) Mutation:

The chromosome will be mutated based on the non-uniform mutation; equation (4), and equation (5).

The steps form 3 to 5 will be repeated until the last number of population. The solution is the first chromosome of the last population.

6) GAnuM-MTSP Algorithm:

GA1: Reformulate the boundaries for each machine in each cycle as follows:

$$\text{Put } H_{ik} = L_{ik} \quad \forall i \in N, \; h_{jr} = l_{jr} \quad \forall j \in N,$$

$$H_{jr} = \min(L_{jr}, (\min_{i \in U_j} H_{ir+1} - p_j))$$

$$\text{where } U_j = \{ i \in N : j \in N^{(i)} \} \quad r = k-1, \ldots, 2, 1$$

GA2: Put p = 1.
GA3: Put c = 1.
GA4: Put r = 1.
GA5: Put i = 1.
GA6: If $r \ne 1$ then $h_{ir+1} = \max_{j \in N^{(i)}} (x_{ircp} + p_{jr}) \quad \forall i \in N$
GA7: Generate random number for x_{ircp} where $h_{ir} \le x_{ircp} \le H_{ir}$.
GA8: If i < n then i = i + 1 go to GA6.
GA9: If r < k then r = r + 1 go to GA5.
GA10: If c < Q then c = c + 1 go to GA4.
GA11: $x_{irc(p+1)} = x_{ircp} \quad \forall i \in n, \forall r \in k, \forall c \in Q.$
GA12: Sort descending $f_{cp}(x_{cp})$ if $p = W$ then go to GA26.
GA13: Put $c = Q - (T*Q)$.
GA14: Put $r = v$.
GA15: Put i = 1.
GA16: Swap between x_{ircp} and $x_{ir(c+1)p}$.
GA17: If $i < n$ then $i = i + 1$ go to GA16.
GA18: If $r < k$ then $r = r + 1$ go to GA15.
GA19: If $c < Q$ then c = c + 2 go to GA14.
GA20: Put c = 1.
GA21: Put r = 1.
GA22: Put i = 1.
GA23: Generate random number *rand* between 0 and 1.
GA24: $\text{if } rand \ge 0.5 \text{ then } x_{irap} = x_{irap} + (H_{ir} - x_{irap}).(1 - r^{(1-\frac{p}{w})^b})$
GA25: $\text{if } rand < 0.5 \text{ then } x_{irat} = x_{irat} - (x_{irat} - h_{ir}).(1 - r^{(1-\frac{t}{T})^b})$
GA26: If $i < n$ then $i = i + 1$ go to GA23.
GA27: If $r < k$ then $r = r + 1$ go to GA22.
GA28: If $c < Q$ then c = c + 1 go to GA21.
GA29: If $p < W$ then $p = p + 1$ go to GA3.
GA30: The solution is $x_{ir1p} \quad \forall i \in n, \forall r \in k.$

7) Simulation Results:

Consider the MTSP with the following parameters:

n = 5, i.e. N = {1,2,3,4,5}, p = {2,4.5,6.25,4,5}, the machines boundaries are as shown in Table I, and the machines relations are as shown in Table II. Assume further that $f_{jr}(x_{jr}) = \max(a_{jr} - x_{jr}, x_{jr} + p_{jr} - b_{jr}, 0) \quad \forall j \in N$

where a_j, b_j are for all $j \in N$ given constants so that we have in our case for all $j \in N$

$$f_{jr}^{(1)}(x_{jr}) = a_{jr} - x_{jr} \qquad f_{jr}^{(2)}(x_{jr} + p_{jr}) = x_{jr} + p_{jr} - b_{jr}$$

Input values of a_{ir} and b_{ir} for each cycle are as shown in Table I.

TABLE I. MACHINE BOUNDARIES

Cycle (r)	r = 1	r = 2	r = 3
l_{ir} i=1,2,…,5	{1,0,0,3,1}	{4,6,6,5,6}	{10,11,12,9,11.5}
L_{ir} i=1,2,…,5	{5,4,3,5,6}	{6.5,7,7.5,7.25,6.5}	{13,12,15,12,14}

TABLE II. MACHINE RELATIONS

i	1	2	3	4	5
$N^{(i)}$	{1,2,3}	{2}	{2,3}	{1,4,5}	{1,3,5}
U_j	{1,4,5}	{1,2,3}	{1,3,5}	{4}	{4,5}

TABLE III. MACHINES PENALTY BOUNDARIES

Cycle (r)	r=1	r=2	r=3
a_{ir} i=1,2,…,5	{1,1,1,3,3}	{5,7,6,5,7}	{11,12,11,10,13}
b_{ir} i=1,2,…,5	{4,6,8,5,5}	{8,9,8,6.5,8}	{13,15,14,12,14}

The GA-MTSP algorithm is implemented and run for solving the allocated problem on a computer with processor Intel Centrino 1.6 GHz with 215 MB RAM. The population size has been tested by 40, 60, 80 and 100 chromosomes. The result shows that, the best population size is 100 chromosome as shown in figure (1-a). The number of chromosomes that will be kept in the next population has been tested by 30%, 20%, 10%, 5%, 2.5% and 1.25% of population size. It was found that, the best number of chromosomes that will be kept in the next population equals 1.25% from population size as shown in figure (1-b). This means that the best chromosome, which has the best fitness function, will be transferred to the next population. The rest of chromosomes will be crossovered together to generate the rest of next population. It means that the probability of crossover is 98.75%, which is another parameter of GA. Using single point crossover, it was found that the best position of cutting point is 33% of the chromosome size as shown in figure (1-c). Finally, the Genetic algorithm parameters that give the best starting times have been determined:

TABLE IV. MACHINE STARTING TIME BY GANUM-MTSP

	M1	M2	M3	M4	M5
C1	1.18	0.85	0.07	3.1	1.36
C2	6.49	6.07	6.67	7.3	6.48
C3	13	11.27	13.13	11.5	13.3

The best value for the objective function f is 35.26. Assuming that α is the cost, and then the penalty cost equals 35.26 α

B. Using Particle Swarm Optimization (PSO)

The PSO method is a member of wide category of Swarm Intelligence methods for solving the optimization problems. It is a population based search algorithm where each individual is referred to as particle and represents a candidate solution. Each particle in PSO flies through the search space with an adaptable velocity that is dynamically modified according to its own flying experience and also the flying experience of the other particles. Further, each particle has a memory and hence it is capable of remembering the best position in the search space ever visited by it. The position corresponding to the best fitness is known as pbest and the overall best out of all the particles in the population is called gbest [9].

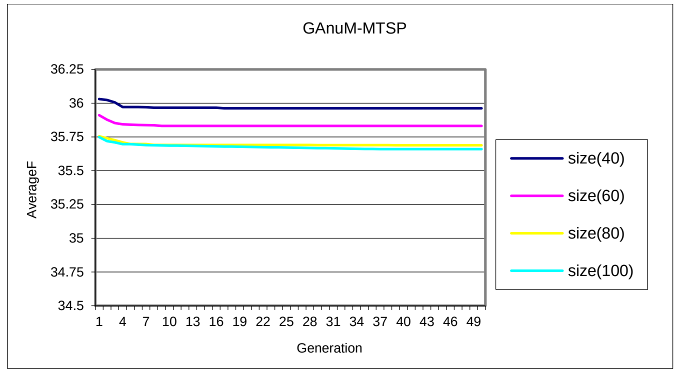
The modified velocity and position of each particle can be calculated using the current velocity and the distance from the $pbest_j$ to gbest as shown in the following formulas:

$$v_{j,g}^{(t+1)} = w * v_{j,g}^{(t)} + c_1 * r_1 * (pbest_{j,g} - x_{j,g}^{(t)}) + c_2 * r_2 * (gbest_{j,g} - x_{j,g}^{(t)})$$

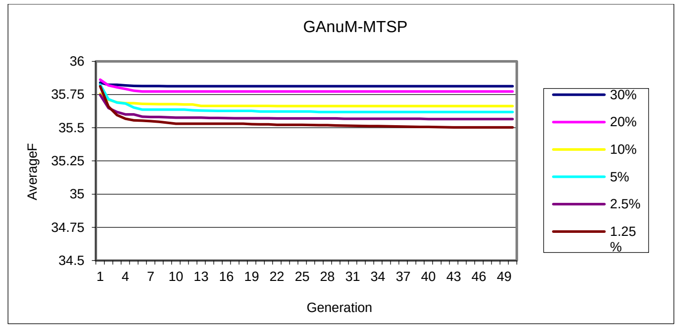
$$x_{j,g}^{(t+1)} = x_{j,g}^{(t)} + v_{j,g}^{(t+1)}$$

With $j=1, 2, \ldots, n$ and $g=1, 2, \ldots, m$
n =number of particles in a group;

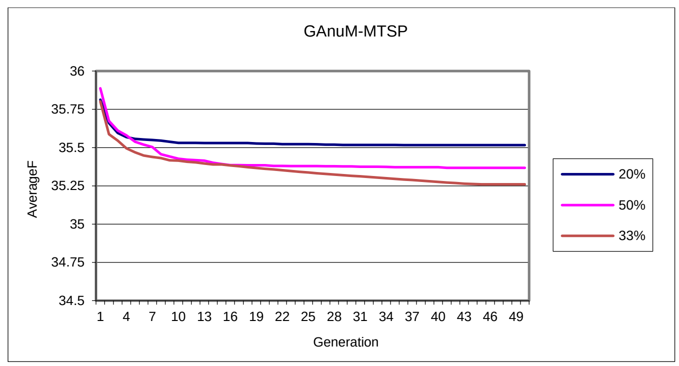
m = number of members in a particle;
t = number of iterations (generations);
$v_{j,g}^{(t)}$ =velocity of particle j at iteration t,
w = inertia weight factor;
c_1, c_2 = cognitive and social acceleration factors, respectively;
r_1, r_2 = random numbers uniformly distributed in the range (0, 1);
$x_{j,g}^{(t)}$ = current position of j at iteration t;
$pbest_j$ = *pbest* of particle j;
gbest = *gbest* of the group.



a) Population size



b) No. of chromosomes kept in next generation



c) Position of the cutting point crossover

Figure 1. Genetic algorithm parameters

The index of best particle among all of the particles in the group is represented by the gbest. In PSO, each particle moves in the search space with a velocity according to its own previous best solution and its group's previous best solution. The velocity update in a PSO consists of three parts; namely momentum, cognitive and social parts.

The balance among these parts determines the performance of a PSO algorithm. The parameters c_1 & c_2 determine the relative pull of pbest and gbest and the parameters r_1 & r_2 help in stochastically varying these pulls [9]. [13] Showed that combining them by setting the inertia weight, χ, to the constriction factor, v, improved performance across a wide range of problems as follows:

$$v_{j,g}^{(t+1)} = \chi\{w * v_{j,g}^{(t)} + c_1 * r_1 * (pbest_{j,g} - x_{j,g}^{(t)}) + c_2 * r_2 * (gbest_{j,g} - x_{j,g}^{(t)})\}$$

$$\chi = \frac{2}{\left|2 - c - \sqrt{c^2 - 4c}\right|} \quad where\, c = c_1 + c_2, \; c < 4.$$

In [14] PSO is combined with the Lagrangian Relaxation (LR) framework to solve a power-generator scheduling problem known as the unit commitment problem (UCP). In terms of the solution quality, the PSO-LR provided a "best solution" with a lower cost than GA for problem sizes larger than 20 units, and than LR for problem sizes 20 and 80 units. PSO-LR also provided a "best solution", for the problem size of 10 units, with a much lower cost than using PSO alone.

A novel approach based on Particle Swarm Optimization (PSO) for scheduling jobs on computational grids is introduced in [15]. The proposed approach is to dynamically generate an optimal schedule so as to complete the tasks within a minimum period of time as well as utilizing the resources in an efficient way. When compared to genetic algorithm (GA) and simulating Annealing (SA), an important advantage of the PSO algorithm is its speed of convergence and the ability to obtain faster and feasible schedules.

Application and performance comparison of PSO and GA optimization techniques were presented in [9], for Thyristor Controlled Series Compensator (TCSC)-based controller design. Results indicate that in terms of computational time, the GA approach is faster. The computational time increases linearly with the number of generations for GA, whereas for PSO the computational time increases almost exponentially with the number of generations. The higher computational time for PSO is due to the communication between the particles after each generation. However, the PSO seems to arrive at its final parameter values in fewer generations than the GA.

The main steps for solving the MTSP by PSO are as follows:

1) Reformulation:

Each machine boundaries will be reformulated (calculate the new boundaries) based on its' successors machines boundaries. For each machine the new lower boundary is called h and the new upper boundary is called H.

2) Initial iteration:

First, the particle is defined as a set of starting times for the machines in all cycles. The particle is represented by *D*-dimensional, where *D* is equal to *N* multiplied by *K* (where *N* number of machines and *K* number of cycles). The x_{irpt} is the starting time for machine *i* in cycle *r* in particle *p*, $p = 1,2,..,Q$ in iteration *t*, $t = 1,2,..,T$ (where *Q* is number of particles in the SWARM and *T* is number of iterations) which satisfy the constraints in (P). The value of x_{irpt} is generated randomly where $h_{ir} \le x_{irpt} \le H_{ir}$.

The x_{irpt} must satisfy the second constrain which is $x_{irpt} \ge \max_{j \in N^{(i)}} (x_{j(r-1)pt} + p_j)$. Determine the $pbest_p$ which is the best position of particle *p* that makes the best value of the objective function. Then determine the *gbest* which is the best particle that make the best value of the objective function in all iterations.

3) Other generation:

The next iteration created by modifying the velocity of each particle by the following equation:

$$v_{irp(t+1)} = \chi\{w * v_{irpt} + c_1 * r_1 * (pbest_{irp} - x_{irpt}) + c_2 * r_2 * (gbest_{ir} - x_{irpt})\}$$

Then the particle position will be update by the following equation:

$$x_{irp(t+1)} = x_{irpt} + v_{irp(t+1)}$$

The new iteration has been created with new position of SWARM. Calculate the objective function then find the $pbest_p$ and gbest. Repeat this step until last iterations. The solution is the gbest in the last iteration.

4) PSOc-MTSP Algorithm:

A1: Reformulate the boundaries for each machine in each cycle as follows:

$$\text{Put} \quad H_{ik} = L_{ik} \quad \forall i \in N, \; h_{jr} = l_{jr} \quad \forall j \in N,$$

$$H_{jr} = \min(L_{jr}, (\min_{i \in U_j} H_{ir+1} - p_j)) \quad \text{Where}$$

$$U_j = \{ i \in N : j \in N^{(i)} \} \quad r = k-1, \ldots, 2, 1$$

A2: Put $t = 1$.
A3: Put $p = 1$.
A4: Put $r = 1$.
A5: Put $i = 1$.
A6: If $r \ne 1$ then $h_{ir} = \max_{j \in N^{(i)}} (x_{j(r-1)pt} + p_j)$.
A7: Generate random number for x_{ircp} where $h_{ir} \le x_{ircp} \le H_{ir}$.
A8: If $i < n$ then $i = i + 1$ go to A6.
A9: If $r < k$ then $r = r + 1$ go to A5.
A10: $pbest_p = f(x_{irpt})_{pt}$ $\exists i = 1,.., N, \exists r = 1,.., K$.
A11: If $p < Q$ then $p = p + 1$ go to A4.
A12: find max $(f(pbest_p)_{pt})$ $\exists p = 1,.., Q$.
A13: $gbest = pbest_{p_{\max}}$
A14: $t = t + 1$.
A15: Put $p = 1$.
A16: $v_{irpt} = \chi\{w * v_{irp(t-1)} + c_1 * r_1 * (pbest_p - x_{irp(t-1)}) + c_2 * r_2 * (gbest - x_{irp(t-1)})\}$
.
A17: $x_{irpt} = x_{irp(t-1)} + v_{irpt}$.
A18: if x_{irpt} is not feasible then go to A20.
A19: if $f(x_{irpt})_{pt} > f(x_{irpt})_{p(t-1)}$ then $pbest_p = x_{irpt}$
A20: $gbest = pbest_{p_{\max}}$.
A21: If $p < Q$ then $p = p + 1$ go to A16.
A22: If $t < T$ then go to A15.
A23: The solution is *gbest.*

5) Simulation Results

The PSOc-MTSP algorithm has been implemented and run for the same scenario used in Part 3.1.2. The program is run 100 times for determining the suitable parameters. Using test sizes of 5, 20, 40 and 50 particle in the swarm, we found that, the best swarm size equals 20 as show in figure (2-a). After testing the w value by 0.1, 0.5, 0.9 and 0.9→ 0.1 (decreasing value) we found that the best value of w equals 0.5 as show in figure (2-b). The last parameters, we need to determine, are c_1, c_2. Using the values 0.5, 1, 1.5, 1.7, 1.9 and 2, for c1 and c2 we found that the best value for c_1, c_2 equal 1.5 as shown in figure (2-c). Finally, the swarm parameters that give the best starting times have been determined is shown below.

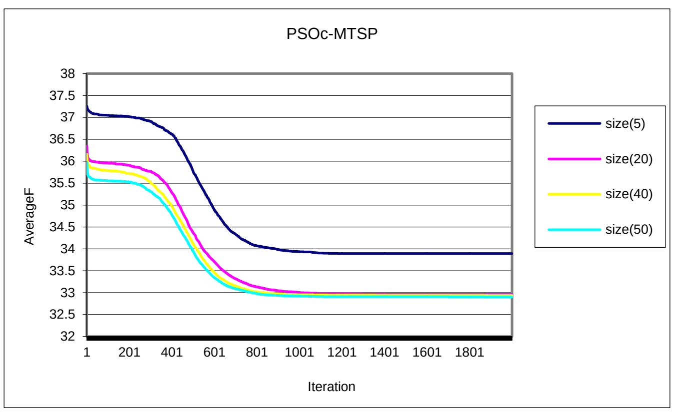
TABLE V. MACHINE STARTING TIME BY PSOC-MTSP

	M1	M2	M3	M4	M5
C1	1.95	1.25	0.0	3	1.26
C2	6.24	6	6.26	7	6.23
C3	12.5	11.26	12.5	11.27	12.5

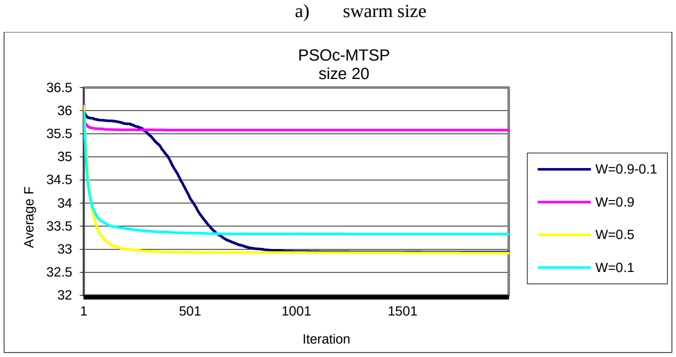
The best value for the objective function f equals 32.9. Assuming that the cost equal α. Then the penalty cost equals 32.9 α.

C. Discussion

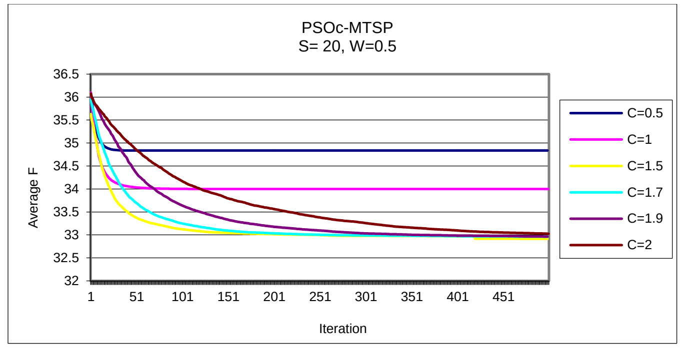
From previous experimental results we found that, solving the MTSP using particle swarm optimization algorithm (PSOc-MTSP), leads to 32.9α penalty cost in 420 iterations that took 11 seconds. When solving the MTSP using genetic algorithm (GAnuM-MTSP), the penalty cost was 35.25 α, reached in generation 41 that took 2 seconds. But when the MTSP was solved using max-separable algorithm in [2], the penalty cost was 35.75 α in less than 0.5 seconds, (Figure 3).



a) swarm size



b) *w* value



c) c_1, c_2 value

Figure 2. Determination of the swarm parameters

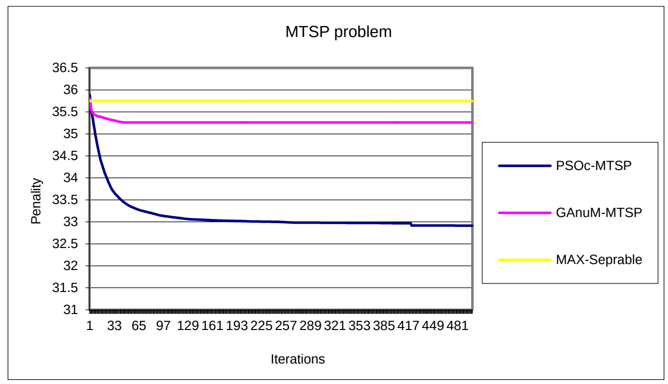


Figure 3. The result of solving MTSP problem by SWARM, GA and Max-separable

D. Conclusion

The machine time scheduling problem (MTSP) was solved using particle swarm optimization (with constriction factor), genetic algorithm (GA) (with non-uniform mutation), and max-separable technique. We found that, particle SWARM optimization gives the lowest penalty cost of the MTSP problem, followed by GA algorithm. The max-separable technique gives the highest penalty cost. That means that particle swarm optimization algorithm is the most suitable for solving the MTSP problem, giving the best starting time for each machine in each cycle.

REFERENCES

[1] [1] Y. Sok and K. Zimmermann: "Optimal Choice of Parameters in Machine Time Scheduling Problems with Penalized Earliness in Starting Time and Lateness", AUC-Mathematica et Physica, V33, No 1, pp.53-61. 1992.

[2] M. Vlach and K. Zimmermann, "Machine Time Scheduling Synchronization of Starting Times", the Proceeding of the MME'99 Conference, Prague, Czech Republic, 1999.

[3] S. Zlobec: "Input Optimization I: Optimal Realizations of Mathematical Models," Mathematical Programming, V 31, pp.245-268, 1985.

[4] S. Zlobec: "Input Optimization III: Optimal Realizations of Mathematical Models," Mathematical Programming, V 17, No 4, pp.429-445, 1986.

[5] A. Tharwat and K. Zimmermann: "Optimal Choice of Parameters in Machine Time Scheduling Problems Case I," Conference MMEI, Liberc, Czech Republic, 1998.

[6] A. Tharwat and A. Abuel-Yazid: “Multi-Cycles Machine Time Scheduling Problem”, First International Conference on Informatics and Systems, Cairo, Egypt, 2002.

[7] S. A. Hassan*, A.A. Tharwat*, I.A. El-Khodary*, A. A. El-Sawy "Using Monte Carlo Simulation to Solve Machine Time Scheduling Problems With Stochastic Processing Time" Mathematical methods in Economics conference: MME’2003, Prague, Czech Republic.

[8] A. Tharwat and A. Abuel-Yazid: "Generalized Algorithm For Muulti-Cycle Machine Time Scheduling", Proceeding of Meatip3 Conference, Assuit, Egypt, 2002.

[9] S. Panda and N. P. Padhy,"Comparison of Particle Swarm Optimization and Genetic Algorithm for TCSC-based Controller Design", International Journal of Computer Science and Engineering, Volume 1 Number 1, 2007.

[10] Jean-Philippe Rennard, "Genetic Algorithm Viewer: Demonstration of a Genetic Algorithm", Ph.D. May 2000.

[11] Zhao X. and Gao X-S., "Evolutionary Programming Based on Non-Uniform Mutation", MM Research Preprints, MMRC, AMSS, Academia Sinica, No. 23, December 2004, Pages: 352-374.

[12] Alfaro-Cid E., McGookin E. W., Murray-Smith D. J., "A Novel Non-uniform Mutation Operator and its Application to the Problem of Optimizing Controller Parameters", IEEE, Vol. 2, 2005,pages: 1555 - 1562.

[13] Alec Banks, Jonathan Vincent, Chukwudi Anyakoha “A review of particle swarm optimization. Part II: hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications” NATURAL COMPUTING Volume 7, Number 1, 109-124.

[14] H. H. BALCI and J. F. VALENZUELA,"scheduling electric power generators using particle swarm optimization combined with the lagrangian relaxation method", Int. J. Appl. Math. Comput. Sci., Vol. 14, No. 3, 411–421, 2004.

[15] H. Liu, A. Abraham and C. Grosan, "A Novel Variable Neighborhood Particle Swarm Optimization for Multi-objective Flexible Job-shop Scheduling Problems", IEEE International Conference on Digital Information Management, Lyon, France, IEEE Press, USA, ISBN 1-4244-1476-8, pp. 138-145, 2007.

The study of prescriptive and descriptive models of decision making

Prof. Ashok A. Divekar
Indira College of Engineering and Management,
PUNE, INDIA

Prof. Sunita Bangal
Indira Institute of Management,
PUNE, INDIA

Prof. Sumangala D.
Indira Institute of Management,
PUNE, INDIA

Abstract—**The field of decision making can be loosely divided into two parts: the study of prescriptive models and the study of descriptive models. Prescriptive decision scientists are concerned with prescribing methods for making optimal decisions. Descriptive decision researchers are concerned with the bounded way in which the decisions are actually made.**

The statistics courses treat risk from a prescriptive, by suggesting rational methods. This paper brings out the work done by many researchers by examining the psychological factors that explain how managers deviate from rationality in responding to uncertainty.

Keywords- Expected Value; Prospect; expected-value rule; risk-averse.

I. Introduction

Most management students are formally taught the concept of risk and uncertainty (or probability) in microeconomics and for statistics courses. These courses typically treat risk by describing rational methods for making decisions involving risk i. e. probability of gain. This paper extends this perspective by examining the psychological factors that tell us how managers deviate from rationality in responding to uncertainty. Max Bazerman has examined how subtle aspects in the presentation of information, called as the “framing” of information, can significantly impact decision making [1]. Throughout the literature on decision making, it is many times suggested that intuitions about risk routinely take over the minds of decision-makers, leading to decisions which deviate from rationality. This is because managers do not understand the nature of uncertainty and the effects of framing. However, uncertainty is a managerial fact of life.

Let us consider two situations which are very common and representative of managers’ dilemma.

A. Situation 1:

A company is faced with a suit from a hostile social group of consumers. The organization believes that it is innocent but it also realizes that the court may not have the same opinion.

1) *Option A:* Settle out of the court and accept a sure loss of Rupees 25,000,000

2) *Option B:* Go to the court expecting a fifty percent probability of a Rs. 50,000,000 loss. Alternatively, the same situation can be defined or framed in this way.

3) *Option C:* Settle out of court and save Rs. 25,000,000 that could be lost in court, or

4) *Option D:* Go to court expecting a fifty percent probability of saving Rs.50, 000,000.

5) *Options A&B are framed negatively*- that is, in terms of possible losses. Option C&D are framed positively that is, in term of possible gains.

A is equivalent to C, and B is equivalent to D. A consistent decision maker would choose A&C, or B&D. choosing A&D, or B&C would be inconsistent.

B. Situation 2 :

When a test was conducted on MBA students, surprisingly, over eighty percent selected option B in the first situations, and option C in the second situations [1]. Research has shown that individuals tend to take risks concerning choices framed in terms of possible losses- even though both sets of choice are objectively the same. Thus, decision makers are systematically affected by the way in which information is presented. Dawes argues that humans want to “know now” in situations containing inherent uncertainty [2]. He claims that the need to do away with uncertainty frequently leads people to take too much credit for successes and too much blame for failures.

In real world, most people are not consistent in their approach to risk. When Slavic compared the scores of eighty-two people on nine different measures of risk taking, he found no evidence to suggest that a generalizable risk taking trait exits [3]. People who are aggressive in one situation may be conservative in another. The conservative financial analyst may let himself break loose on weekends. The habitual gambler may be conservative in caring for his health or making rules for his children. Slovic found that intelligence was not related to a risk – taking level [3].

When a test was conducted on MBA students, surprisingly, over eighty percent selected option B in the first situations, and option C in the second situations [1]. Research has shown that individuals tend to take risks concerning choices framed in terms of possible losses- even though both sets of choice are objectively the same. Thus, decision makers are systematically affected by the way in which information is presented. Dawes argues that humans want to “know now” in situations containing inherent uncertainty [2]. He claims that the need to do away with uncertainty frequently leads people to take too much credit for successes and too much blame for failures.

In real world, most people are not consistent in their approach to risk. When Slavic compared the scores of eighty-two people on nine different measures of risk taking, he found no evidence to suggest that a generalizable risk taking trait exits [3]. People who are aggressive in one situation may be conservative in another. The conservative financial analyst may let himself break loose on weekends. The habitual gambler may be conservative in caring for his health or making rules for his children. Slovic found that intelligence was not related to a risk – taking level [3].

As managers, we are constantly faced with decisions that will lead to uncertain outcomes [4]. Many of this risky decision are crucial, involving issues like jobs, safety, product reliability, and sometimes organizational existence. As such, managers should understand risk, which will enhance their ability to make and evaluate decision in uncertain situations. Managers will make better decisions by accepting that uncertainty exists and learning how to think systematically in risky situations. Dawes describes,

"Imagine a life without uncertainty - - - - - - - -Imagine how dull life would be if everything can be predicted with great accuracy. Life would be intolerable - - - - - - No hope no challenge!" [5].

We know that the individuals are typically neither rational nor consistent in making judgments under uncertainty. We all know that the rational decision-making process includes

- Specifying the problem.
- Identifying all factors.
- Weighting Factors.
- Identifying all alternatives.
- Rating alternatives on each factor.
- Choosing the optimal alternative.

The rational model does not tell us how to rate alternatives when the outcome on a particular factor is uncertain.

Let us understand two concepts: Probability and Expected value.

A. Probability:

The concept of probability tells us the likelihood that any particular outcome will occur. A probability of 1.0 represents certainty that an event will occur. A probability of zero represents certainty that an event will not occur.

B. Expected Value:

Calculating the expected value of any alternative involves weighting all potential outcomes associated with that alternative by their probabilities and summing them. One simple rule for making decision is to always select the alternative with the highest expected value. Consider following scenarios:

You can

a) Have Rs. 10,000,000 for sure (expected value = Rs. 10,000,000) or

b) Flip an honest coin where you get Rs. 24,000,000, if a heads occurs but get nothing if a tail occurs (expected value = Rs. 12,000,000).

An expected value decision rule would pick. (b)

What would you do?

You are being sued for Rs. 100,000 and estimate a 50 % chance of losing the case (expected value = Rs. - 50,000).

However, the other side is willing to accept an out-of-court settlement of Rs. 40,000 (expected value = - 40,000), ignoring attorney fees, court costs, and so on.

a) Would you fight the case?

b) Settle out of court?

An expected value decision rule would tell you to settle out of court. Most people would take a) in both cases. This clearly indicates that a number of situations exist in which people do not follow an expected-value rule.

II. Risk Considerations

Let us try to understand the concept of a "certainty equivalent". A certainty equivalent establishes the certain value that would make a decision maker indifferent between an uncertain event and that certain value. For example if you had an opportunity to accept a 50 % chance of obtaining Rs. 100,000, what would be the certain amount that would make you indifferent between the 50 % chance of Rs. 100,000 and that amount. Rs. 10,000? Rs. 40,000? This is your certainty equivalent to a 50 % chance of Rs. 100,000. For most people, this amount is far less than the expected value of the bet of Rs. 50,000.

An individual who has a certainty equivalent for an uncertain event that is equal to the expected value of the uncertain payoff is "risk neutral", with regard to that decision.

For example, if your certainty equivalent in the previous paragraph was Rs. 50,000, you would be risk neutral, concerning that choice thus, risk neutrality is synonymous with using an expected –value decision rule. An individual with a certainty equivalent for an uncertain event that is less than the expected value of that uncertain payoff is "risk averse", with regard to that decision. If your certainty equivalent in the previous paragraph was Rs. 40,000, then you are risk averse, since you are willing to take an expected value reduction of Rs. 10,000 to avoid the risk associated the uncertain event. Although it is unlikely in the situation described in the previous paragraph, an individual with a certainty equivalent for an uncertain event that is more than the expected value for that uncertain payoff is "risk seeking" with regard to that decision. If your certainty equivalent was Rs. 70,000, you are "risk seeking", since you are demanding extra expected value of Rs. 20,000 to forego the risk. Thus you seek risk, holding the expected value constant.

When decision makes act in a risk-averse or risk seeking manner, they make decisions that often exclude the maximizing of expected value. To explain departures from the expected value decision rule, Daniel Bernoulli first suggested replacing the criterion of expected monetary value with the criterion of expected utility [1]. Expected utility theory suggests that an each level of an outcome is associated with some degree of pleasure, or net benefit, called utility. The expected utility of an uncertain choice is the weighted sum of the utilities of its outcomes, each multiplied by its probability while an expected value approach to decision-making would treat Rs. 100,000 as being worth twice as much as Rs. 50,000, a gain of Rs. 100,000 does not always create twice as much expected utility as a gain of Rs. 50,000. Most individuals do not obtain as much utility from the second Rs. 50,000 as they did from the first Rs. 50,000. Thus under expected utility theory, the decision maker is predicted to select the option with the highest expected utility, regardless of whether that choice has the highest expected value.

Further, according to expected- utility theory, individuals identity outcomes in terms their overall wealth and the additional wealth they would have as a result of each alternative outcome. Kahneman and Tversky's prospect theory refutes this aspect of expected-utility theory [6]. They argue that each decision is approached independently.

III. THE FRAMING OF INFORMATION

Consider the following problem [7].

A large car manufacturer has recently been hit with a number of economic difficulties, and it appears as if three plants need to be closed and 6000 employees laid off. The vice president of production has been exploring alternative ways to avoid this crisis. He has developed two plans:

A. *Plan A:*

This plan will save one of the three plants, and 2000 jobs.

B. *Plan B:*

This plan has a 1/3 probability of saving all three plants and all 6000 jobs, but has a 2/3 probability of saving no plants and no jobs.

Which plan would you select?

There are number of things which we might consider in evaluating these options like impact on union, moral of retained employees, and the values of the corporation. While all these questions are important, a more fundamental question underlies the subjective situation and the resulting decision.

Now consider this problem, replacing the choices with the following choices:

C. *Plan C:*

This plan will result in the loss of two of three plants and 4000 jobs.

D. *Plan D:*

This plan has a 2/3 probability of resulting in the loss of all three plants and all 6000 jobs, but has a 1/3 probability of loosing no plants and no jobs.

Which plan would you select?

Close examination of the two sets of alternative plans finds them to be objectively the same. Plan A and Plan C offer the same objective outcome. Plans B and D are objectively identical.

However, the investigation demonstrates that most management graduates choose Plan A in the first set and Plan D in the second set [7]. Changing the description of the outcomes from jobs and plants saved to jobs and plants lost is responsible to shift choice from risk-averse to risk seeking behavior.

This shift is consistent with the available literature that show that individuals treat risks concerning perceived gains (saving plants, jobs) differently from risks concerning perceived losses (loosing plants, jobs) [8][9][10]. In an attempt to explain these common and systematic deviations from rationality, Kahneman and Tversky developed "Prospect Theory" [8]. This theory suggests the following:

1) *Rewards and losses are evaluated relative to a neutral reference point.*
2) *Potential outcomes are expressed as gains (such as jobs and plants saved) or losses (jobs and plants lost) relative to this fixed, neutral reference point.*
3) *The choices that people make are formed based on the resulting change in asset position as assessed by S – shaped value function. (See Fig. 1).*

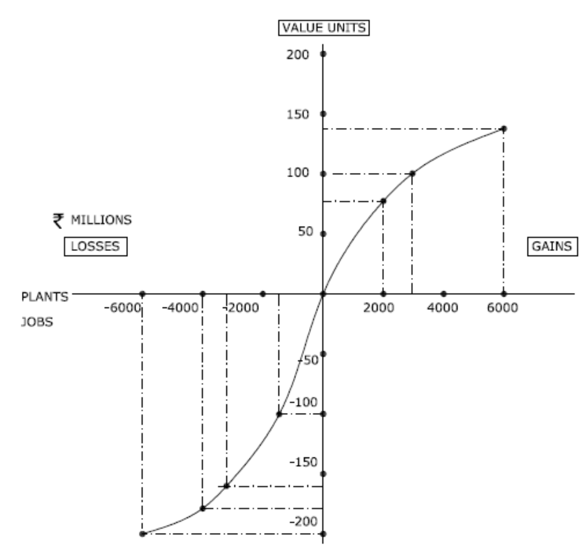


Fig 1. Hypothetical value function accounting for framing

Figure 1. Hypothetical Value function

On this graph, the X – axis represents the nominal units gained or lost, and the Y – axis represents the units of utility associated with varying levels of gain or loss. This figure shows that decision makers tends to avoid risk concerning gains and seek risk concerning losses. For example, the S – shaped value function implies that most individuals would choose a Rs. 10,000,000 sure gain over a 50% chance of getting a Rs. 20,000,000 gain, since the utility placed on Rs. 20.000,000 is not twice as great as the value placed on Rs. 10,000,000, but that most individuals would choose a 50% chance of a Rs. 20,000,000 loss over a sure loss of Rs. 10,000,000, since the negative value placed on Rs. 20,000,000 is not twice as great as the negative value placed on Rs. 10,000,000.

IV. Conclusion

An important outcome of this theory is that the way in which the problem is "framed" or presented, can dramatically change the perceived neutral point of the question. In the above problem, if the problem is framed in terms of losing jobs and plants (Plan C and Plan D), the current position of having three plants open forms the neutral reference point, and risk – seeking behavior results. The negative value placed on the loss of three plants and 6000 jobs is usually perceived as not being three times as negative as losing one plant and 2000 jobs. However, if the problem is framed in terms of saving jobs and plants (Plan A and Plan B) the potential disaster of losing everything becomes the neutral reference point, leading to risk-averse behavior.

A second characteristic of our decision making processes identified by prospect theory is that our response to loss is more extreme than our response to gain. According to Fig. 1, the pain associated with loosing Rs. X is generally greater than the pleasure associated with winning the same amount.

Prospect theory identifies a third way in which our decision making processes deviate from expected utility theory. It states that we tend to overweight the probability of two probability events and underweight the probability of moderate and high probability events.

Prospect theory represents the most important advance in our understanding of behavioral decision-making processes. Apart from "framing" gains and losses, the context in which outcomes occur also affects our consistencies in interpreting outcomes. Similarly, the summing, or aggregation, of a number of outcomes evokes different responses.

We put a higher value on loosing personal time than we do on gaining more personal time. We want good outcomes now, but bad outcomes later. Postponing a good outcome results in a greater loss of perceived utility than not postponing an equivalent bad outcome.

We should now be able to identify situations in which we currently adopt a particular frame. If we understand and apply this knowledge, the consistency and quality of our decision will improve.

References

[1] Bazerman, Max H: Judgment in managerial decision making: John Wiley & Sons.

[2] Dawes, R.M. Rational choice in an uncertain world: New York: Harcourt Brace Jovanovich.

[3] Slovic, P.: Information processing, situation specificity, and the generality of risk-taking behavior: Journal of personality & social psychology.

[4] MacCrimmon, K.R. and Wehrung D. A. Taking risks: New York Free Press.

[5] Dawes, R. M.: Rational choice in an uncertain world: New York: Harcourt Brace Jovanovich.

[6] Kahneman D., and Tversky A.: Prospect theory: An analysis of decision under risk: Econometrica.

[7] Miller, P. M. and Fagley, N. S.: The effects of framing, problem variations, and providing rationale on choice.

[8] Bazerman, Max H.: The relevance of Kahneman & Tversky's concept of framing to organizational behavior: Journal of management.

[9] Tversky, A. and Kahneman, D.: Rational choice and the framing of decision: Journal of business.

[10] Thaler, R.: Toward a positive theory of consumer choice: Journal of economic behavior & organization.