An Overview of Recent Machine Learning Strategies in Data Mining

Bhanu Prakash Battula
Research Scholar
Acharya Nagarjuna University
Guntur, Andhra Pradesh, India

Dr. R. Satya Prasad
Associate Professor
Acharya Nagarjuna University
Guntur, Andhra Pradesh, India

***Abstract*—Most of the existing classification techniques concentrate on learning the datasets as a single similar unit, in spite of so many differentiating attributes and complexities involved. However, traditional classification techniques, require to analysis the dataset prior to learning and for not doing so they loss their performance in terms of accuracy and AUC. To this end, many of the machine learning problems can be very easily solved just by careful observing human learning and training nature and then mimic the same in the machine learning.**

This paper presents an updated literature survey of current and novel machine learning strategies inducing models efficiently for supervised and unsupervised learning in data mining.

Keywords*—Data mining; classification; supervised learning; unsupervised learning; learning strategies.

I. INTRODUCTION

One of the research hotspots in the field of machine learning is classification. There are different types of classification models such as decision trees, SVM, neural networks, Bayesian belief networks, Genetic algorithm etc. The simple structure, the wide applicability on real time problems, the high efficiency and the high accuracy are the strengths for decision trees. In recent years, many authors proposed improvements in decision trees learning strategy. A large number of classifiers build the model of dataset for classification by using the traditional learning strategies. On the other hand, the traditional learning techniques are bottle necked the performance of the datasets. However, several investigations also suggest that there are other factors that contribute to such performance degradation, for example, size of the dataset, density of the dataset, and overall complexity of the dataset. This paper presents an updated survey of various machine learning strategies. It also describes the applicability of the algorithm on real-world data.

The rest of this paper is organized as follows. Section 2 presents the basics of data mining. Section 3 describes a generic datasets and measures used for recent learning strategies. Several recent works related to different learning strategies are reviewed in Section 4. Section 5 concludes our work by presenting future scope on the topic.

II. DATA MINING

A. Basics of Data Mining

Data Mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the owner [1]. There are many different data mining functionalities. A brief definition of each of these functionalities is now presented. The definitions are directly collated from [2]. Data characterization is the summarization of the general characteristics or features of a target class of data. Data Discrimination, on the other hand, is a comparison of the general features of target class data objects with the general features of objects from one or a set of contrasting classes. Association analysis is the discovery of association rules showing attribute value conditions that occur frequently together in a given set of data.

Classification is an important application area for data mining. Classification is the process of finding a set of models (or functions) that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The derived model can be represented in various forms, such as classification rules, decision trees, mathematical formulae, or neural networks. Unlike classification and prediction, which analyze class-labeled data objects, clustering analyzes data objects without consulting a known class label.

Outlier Analysis attempts to find outliers or anomalies in data. A detailed discussion of these various functionalities can be found in [2]. Even an overview of the representative algorithms developed for knowledge discovery is beyond the scope of this paper. The interested person is directed to the many books which amply cover this in detail [1], [2].

B. Classificationa and Clustering Tasks

Learning how to classify objects to one of a pre-specified set of categories or classes is a characteristic of intelligence that has been of keen interest to researchers in psychology and computer science. Identifying the common ―core‖ characteristics of a set of objects that are representative of their class is of enormous use in focusing the attention of a person or computer program. For example, to determine whether an animal is a zebra, people know to look for stripes rather than examine its tail or ears. Thus, stripes figure strongly in our *concept* (generalization) of zebras. Of course stripes alone are not sufficient to form a class description for zebras as tigers have them also, but they are certainly one of the important characteristics.

The ability to perform classification and to be able to *learn* to classify gives people and computer programs the power to

make decisions. The efficacy of these decisions is affected by performance on the classification task. In machine learning, the classification task described above is commonly referred to as *supervised learning*. In supervised learning there is a specified set of classes, and example objects are labeled with the appropriate class (using the example above, the program is told what a zebra is and what is not). The goal is to generalize (form class descriptions) from the training objects that will enable novel objects to be identified as belonging to one of the classes.

In contrast to supervise learning is *unsupervised learning*. In this case the program is not told which objects are zebras. Often the goal in unsupervised learning is to decide which objects should be grouped together—in other words, the learner forms the classes itself. Of course, the success of classification learning is heavily dependent on the quality of the data provided for training—a learner has only the input to learn from. If the data is inadequate or irrelevant then the concept descriptions will reflect this and misclassification will result when they are applied to new data.

III. Datasets and Performance Evaluation Measures

A. Benchmark Datasets

Table I summarizes the benchmark datasets used in almost all the recent studies of machine learning. The details of the datasets are given in Table I. For each data set, the number of instances, missing values, numeric attributes, nominal attributes and number of classes. The complete details regarding all the datasets can be obtained from UCI Machine Learning Repository [3].

B. Evaluation Criteria

To assess the classification results. The most commonly used performance evaluation measures in machine learning are accuracy, tree size, AUC and error rate. Let us define a few well-known and widely used measures:

The most commonly used empirical measure; accuracy

is computed by using the below equation (1),

$$ACC = \frac{TP + TN}{TP + FN + FP + FN} \qquad \text{---------- (1)}$$

Another measure for performance evaluation is AUC. A quantitative representation of a ROC curve is the area under it, which is known as AUC. When only one run is available from a classifier, the AUC can be computed as the arithmetic mean (macro-average) of TP rate and TN rate.

The Area under Curve (AUC) measure is computed by equation (2),

$$AUC = \frac{1 + TP_{RATE} - FP_{RATE}}{2} \qquad \text{---------- (2)}$$

TABLE I. Summary of benchmark datasets used in Machine Learning

S.no	Dataset	Instances	Missing Values	Numeric Attrib.	Nominal Attrib.	Classes
1.	Anneal	898	no	6	32	5
2.	Anneal.ORIG	898	yes	6	32	5
3.	Arrhythmia	452	yes	206	73	13
4.	Audiology	226	yes	0	69	24
5.	Autos	205	yes	15	10	6
6.	Balance-scale	625	no	4	0	3
7.	Breast-cancer	286	yes	0	9	2
8.	Breast-w	699	yes	9	0	2
9.	Colic-h	368	yes	7	15	2
10.	Colic-h.ORIG	368	yes	7	15	2
11.	Credit-a	690	yes	6	9	2
12.	Credit-g	1,000	no	7	13	2
13.	Pima diabetes	768	no	8	0	2
14.	Ecoli	336	no	7	0	8
15.	Glass	214	no	9	0	6
16.	Heart-c	303	yes	6	7	2
17.	Heart-h	294	yes	6	7	2
18.	Heart-statlog	270	no	13	0	2
19.	Hepatitis	155	yes	6	13	12
20.	Hypothyroid	3,772	yes	7	22	4
21.	Ionosphere	351	no	34	0	2
22.	Iris	150	no	4	0	3
23.	Kr-versus-kp	3,196	no	0	36	2
24.	Labour	57	yes	8	8	2
25.	Letter	20,000	no	16	0	26
26.	Lympho	148	no	3	15	4
27.	Mushroom	8,124	yes	0	22	2
28.	Optdigits	5,620	no	64	0	10
29.	Pendigits	10,992	no	16	0	10
30.	Primary-tumour	339	yes	0	17	21
31.	Segment	2,310	no	19	0	7
32.	Sick	3,772	yes	7	22	2
33.	Sonar	208	no	60	0	2
34.	Soybean	683	yes	0	35	19
35.	Splice	3,190	no	0	61	3
36.	Vehicle	846	no	18	0	4
37.	Vote	435	yes	0	16	2
38.	Vowel	990	no	10	3	11
39.	Waveform	5,000	no	41	0	3
40.	Zoo	101	no	1	16	7

In the machine learning experiments, the size of the tree is calculated by the depth of the tree using number of nodes and leaves. Testing errors is computed as the number of errors produced when separate training and testing set is used for training and testing.

IV. Recent Advances in Machine Learning Strategies

Serkan celik *et al.* [4] have examine vocabulary-learning strategies adopted by Turkish EFL students, specifically the frequencies and helpfulness ratings of strategy use, strategy patterns, as well as their change for students across different language levels. The study involved 95 tertiary level English as a foreign language learners.

Data were analyzed statistically and the results indicated that the participants' general use of vocabulary learning strategies was somewhat inadequate and there was a gap between their use of strategies and related perceptions of strategy usefulness. Zhou GuoDong *et al.* [5] have proposed a novel hierarchical learning strategy to deal with the data sparseness problem in semantic relation extraction by modeling the commonality among related classes. For each class in the hierarchy either manually predefined or automatically clustered, a discriminative function is determined in a top-down way. As the upper-level class normally has much more positive training examples than the lower-level class, the corresponding discriminative function can be determined more reliably and guide the discriminative function learning in the lower-level one more effectively, which otherwise might suffer from limited training data. Authors proposed, two classifier learning approaches, i.e. the simple perceptron algorithm and the state-of-the-art Support Vector Machines, are applied using the hierarchical learning strategy.

Edwin Lughofer [6] has proposed a novel active learning strategy for data-driven classifiers, which is based on unsupervised criterion during off-line training phase, followed by a supervised certainty-based criterion during incremental on-line training. In this sense, they call the new strategy hybrid active learning. Sample selection in the first phase is conducted from scratch (i.e.no initial labels/learners are needed) based on purely unsupervised criteria obtained from clusters: samples lying near cluster centers and near the borders of clusters are expected to represent the most informative ones regarding the distribution characteristics of the classes. In the second phase, the task is to update already trained classifiers during on-line mode with the most important samples in order to dynamically guide the classifier to more predictive power.

Both strategies are essential for reducing the annotation and supervision effort of operators in off-line and on-line classification systems, as operators only have to label an exquisite subset of the off-line training data representation give feedback only on specific occasions during on-line phase.

Kevin Duh *et al.* [7] have proposed a flexible transfer learning strategy based on sample selection. Source domain training samples are selected if the functional relationship between features and labels do not deviate much from that of the target domain. This is achieved through a novel application of recent advances from density ratio estimation. The approach is flexible, scalable, and modular. It allows many existing supervised rankers to be adapted to the transfer learning setting. Xiaodong Yu *et al.* [8] have proposed a novel updating algorithm based on iterative learning strategy for delayed coking unit (DCU), which contains both continuous and discrete characteristics. Daily DCU operations under different conditions are modeled by a belief rule-base (BRB), which is then, updated using iterative learning methodology, based on a novel statistical utility for every belief rule. Compared with the other learning algorithms, their methodology can lead to a more optimal compact final BRB. With the help of this expert system, a feed forward compensation strategy is introduced to eliminate the disturbance caused by the drum-switching operations.

R.J. Gil *et al.* [9] have proposed a novel model of an Ontology-Learning Knowledge Support System (OLeKSS) is proposed to keep these KSSs updated. The proposal applies concepts and methodologies of system modeling as well as a wide selection of OL processes from heterogeneous knowledge sources (ontologies, texts, and databases), in order to improve KSS's semantic product through a process of periodic knowledge updating. An application of a Systemic Methodology for OL (SMOL) in an academic Case Study illustrates the enhancement of the associated ontologies through process of population and enrichment.

Md Nasir *et al.* [10] have proposed a variant of single-objective PSO called Dynamic Neighborhood Learning Particle Swarm Optimizer (DNLPSO), which uses learning strategy whereby all other particles' historical best information is used to update a particle's velocity as in Comprehensive Learning Particle Swarm Optimizer (CLPSO). But in contrast to CLPSO, in DNLPSO, the exemplar particle is selected from a neighborhood. This strategy enables the learner particle to learn from the historical information of its neighborhood or sometimes from that of its own.

Moreover, the neighborhoods are made dynamic in nature i.e. they are reformed after certain intervals. This helps the diversity of the swarm to be preserved in order to discourage premature convergence. Biao Niu *et al.* [11] have proposed a novel batch mode active learning scheme for informative sample selection. Inspired by the method of graph propagation, we not only take the correlation between labeled samples and unlabeled samples, but the correlation among unlabeled samples taken into account as well. Especially, considering the unbalanced distribution of samples and the personalized feedback of human we propose an asymmetric propagation scheme to unify the various criteria including uncertainty, diversity and density into batch mode active learning in relevance feedback.

Ching-Hung Lee *et al.* [12] have proposed a hybrid of algorithms for electromagnetism-like mechanisms (EM) and particle swarm optimization (PSO), called HEMPSO, for use in designing a functional-link-based Petri recurrent fuzzy neural system (FLPRFNS) for nonlinear system control. The FLPRFNS has a functional link-based orthogonal basis function fuzzy consequent and a Petri layer to eliminate the redundant fuzzy rule for each input calculation. In addition, the FLPRFNS is trained by the proposed hybrid algorithm. The main innovation is that the random-neighbourhood local search is replaced by a PSO algorithm with an instant-update strategy for particle information. Each particle updates its information instantaneously and in this way receives the best current information. Thus, HEMPSO combines the advantages of multiple-agent-based searching, global optimization, and rapid convergence. Gwénolé Quellec *et al.* [13] have proposed a novel multiple-instance learning framework, for automated image classification. Given reference images marked by clinicians as relevant or irrelevant, the image classifier is trained to detect patterns, of arbitrary size, that only appear in relevant images. After training, similar patterns are sought in

new images in order to classify them as either relevant or irrelevant images. Therefore, no manual segmentations are required. As a consequence, large image datasets are available for training.

TABLE II. RECENT ADVANCES IN LEARNING STRATEGY

ALGORITHM	DESCRIPTION	REFERENECE
Hi-SVM Hi-PA	Hierarchical learning Strategy On SVM and Perceptron Algorithm.	[5]
IL-RIMMER	Iterative learning-RIMMER Algorithm.	[8]
OLeKSS	Ontology-Learning Knowledge Support System	[9]
DNLPSO	Dynamic Neighborhood Learning Particle Swarm Optimizer	[10]
APAL	Asymmetric Propagation based Active Learning algorithm	[11]
HEMPSO	Hybridization of ElectroMagnetism like and Particle Swarm Optimization Algorithm	[12]
MIL	Multiple Instance Learning Framework	[13]
MGDT	Maximum Gain Decision Tree for OR-Decision Tables	[14]
S2D	Simple-to-Complex Human Learning Strategy	[16]
GCSDT	Genetically optimized Cluster Oriented Soft Decision Trees	[17]

Costantino Grana *et al.* [14] have proposed a novel algorithm to synthesize an optimal decision tree from OR-decision tables, an extension of standard decision tables, complete with the formal proof of optimality and computational cost analysis. As many problems which require recognizing particular patterns can be modeled

with this formalism, They select two common binary image processing algorithms, namely connected components labeling and thinning, to show how these can be represented with decision tables, and the benefits of their implementation as optimal decision trees in terms of reduced memory accesses.

Joel E. Denny *et al.* [15] have demonstrate that a well-known algorithm described by David Pager and implemented in Menhir, the most robust minimal LR(1) implementation they have discovered that, it does not always achieve the full power of canonical LR(1) when the given grammar is non-LR(1) coupled with a specification for resolving conflicts. They also detail an original minimal LR(1) algorithm, IELR(1) (Inadequacy Elimination LR(1)), which they have implemented as an extension of GNU Bison and which does not exhibit this deficiency.

Eileen A. Ni*et al.* [16] have proposed a novel, simple and effective machine learning paradigm that explicitly exploits this important simple-to-complex (S2C) human learning strategy, and implement it based on C4.5 efficiently. Sanjay Kumar Shukla *et al.* [17] have developed a novel methodology, genetically optimized cluster oriented soft decision trees (GCSDT), to glean vital information imbedded in the large databases. In contrast to the standard C-fuzzy decision trees, where granules are developed through fuzzy (soft) clustering, in the proposed architecture granules are developed by means of genetically optimized soft clustering. In the GCSDT architecture, GA ameliorates the difficulty of choosing an initialization for the fuzzy clustering algorithm and always avoids degenerate partitions. This provides an effective means for the optimization of clustering criterion, where an objective function can be illustrated in terms of cluster's center. Growth of the GCSDT is realized by expanding nodes of the tree, characterized by the highest inconsistency index of the information granules.

Sanjay Jain *et al.* [18] have a present study aims at insights into the nature of incremental learning in the context of Gold's model of identification in the limit. With a focus on natural requirements such as consistency and conservativeness, incremental learning is analyzed both for learning from positive examples and for learning from positive and negative examples. In [19] authors introduced a novel form of decision tables, namely OR-Decision Tables, which allow including the representation of equivalent actions for a single rule. An heuristic to derive a decision tree for such decision tables was given, without guarantees on how good the derived tree was. In [20], authors presented a preliminary version of a bottom-up dynamic programming proposed by Schumacher et al. [21] which guarantees to find the optimal decision tree given an expanded limited entry (binary) decision table, in which each row contains only one non zero value.

V. CONCLUSION

Traditional classification techniques build the model for the datasets by following traditional and old strategy. New and novel learning strategies which mimic human learning can of great use to improve the process of model building for the datasets. In this paper we first investigate the state of the art methodologies for machine learning. This issue hinders the performance of standard classifier learning algorithms that assume relatively balanced class distributions, and classic ensemble learning algorithms are not an exception.

In recent years, several methodologies integrating solutions to enhance the induced classifiers in the presence of learning strategies by the usage of evolutionary techniques have been presented. However, there was a lack of framework where each one of them could be classified; for this reason, a taxonomy where they can be placed has been taken as our future work. Finally, we have concluded that intelligence based algorithms are the need of the hour for improving the results that are obtained by the usage of data preprocessing techniques and training a single classifier.

In our future work, we will apply our proposed method for learning wide range of tasks, especially for high dimensional feature learning tasks.

REFERENCES

[1] David Hand, Heikki Mannila, and Padhraic Smyth. Principles of Data Mining. MIT Press, August 2001.

[2] Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, April 2000.

[3] A. Asuncion D. Newman. (2007). *UCI Repository of Machine Learning Database* (School of Information and Computer Science), Irvine, CA: Univ. of California [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.htm

[4] Serkan ÇELİK, Veli TOPTAŞ. Vocabulary learning strategy use of Turkish EFL learners, Procedia Social and Behavioral Sciences 3 (2010) 62–71.

[5] Zhou GuoDong, Zhang Min, Ji DongHong, Zhu QiaoMing. Hierarchical learning strategy in semantic relation extraction, Information Processing and Management 44 (2008) 1008–1021.

[6] Edwin Lughofer. Hybrid active learning for reducing the annotation effort of operators in classification systems, Pattern Recognition 45 (2012) 884–896.

[7] Kevin Duh, Akinori Fujino. Flexible sample selection strategies for transfer learning in ranking, Information Processing and Management 48 (2012) 502–512.

[8] Xiaodong Yu , DexianHuang, YonghengJiang, YihuiJin. Iterative learning belief rule-base inference methodology using evidential reasoning for delayed coking unit, Control Engineering Practice 20 (2012) 1005–1015.

[9] R.J. Gil, M.J. Martin-Bautista. A novel integrated knowledge support system based on ontology learning: Model specification and a case study, Knowledge-Based Systems 36 (2012) 340–352.

[10] Md Nasir, Swagatam Das, Dipankar Maity, Soumyadip Sengupta, Udit Halder, P.N. Suganthan. A dynamic neighborhood learning based particle swarm optimizer for global numerical optimization, Information Sciences 209 (2012) 16–36.

[11] Biao Niu, JianCheng, XiaoBai, HanqingLu . Asymmetric propagation based batch mode active learning for image retrieval, Signal Processing , Article in Press.

[12] Ching-Hung Lee, Yu-Chia Lee. Nonlinear systems design by a novel fuzzy neural system via hybridization of electromagnetism-like mechanism and particle swarm optimisation algorithms, Information Sciences 186 (2012) 59–72.

[13] Gwénolé Quellec, Mathieu Lamard, Michael D. Abràmoff, Etienne Decencière, Bruno Lay, Ali Erginay, Béatrice Cochener, Guy Cazuguel. A multiple-instance learning framework for diabetic retinopathy screening, Medical Image Analysis 16 (2012) 1228–1240.

[14] Grana, C., Montangero, M., Borghesani, D., Optimal Decision Trees for Local Image Processing Algorithms, Pattern Recognition Letters (2012), doi: http://dx.doi.org/10.1016/j.patrec.2012.08.015.

[15] Joel E. Denny, Brian A. Malloy, "The IELR(1) algorithm for generating minimal LR(1) parser tables for non-LR(1) grammars with conflict resolution", Science of Computer Programming 75 (2010) 943_979.

[16] Eileen A. Ni and Charles X. Ling, "Supervised Learning with Minimal Effort", M.J. Zaki et al. (Eds.): PAKDD 2010, Part II, LNAI 6119, pp. 476–487, 2010.

[17] Sanjay Kumar Shukla a, M.K. Tiwari," Soft decision trees: A genetically optimized cluster oriented approach", Expert Systems with Applications 36 (2009) 551–563.

[18] Sanjay Jain a,1, Steffen Lange b, Sandra Zilles, "Some natural conditions on incremental learning", Information and Computation 205 (2007) 1671–1684.

[19] C. Grana, D. Borghesani, R. Cucchiara, Optimized Block-based Connected Components Labeling with Decision Trees, IEEE T Image Process 19 (2010) 1596–1609.

[20] C. Grana, M. Montangero, D. Borghesani, R. Cucchiara, Optimal decision trees generation from or-decision tables, in: Image Analysis and Processing - ICIAP 2011, volume 6978, Ravenna, Italy, pp. 443–452.

[21] H. Schumacher, K. C. Sevcik, The Synthetic Approach to Decision TableConversion, Commun ACM 19 (1976) 343–351.