RKE-CP: Response-based Knowledge Extraction from Collaborative Platform of Text-based Communication

Jalaja G

Research Scholar

Visvesvaraya Technological University

Belagavi, Karnataka, India

Kavitha C

Professor and Head

Department of CSE, Global Academy of Technology,

Bengaluru, India

Abstract—**With the generation of massive amount of product-centric responses from existing applications on collaborative platform, it is necessary to perform a discrete analytical operation on it. As majority of such responses are textual in nature, it increases the applicability of using text mining approaches on it. We review the existing research contribution in text mining to find that there are significant research gap. Therefore, the proposed study presents a technique called as RKE-CP i.e. Response-based Knowledge Extraction from Collaborative Platform where the term Collaborative points towards cloud environment. The proposed technique is designed using mathematical modelling where the maximum focus of design and implementation lies on accomplishing a good balance between faster response time in mining operation and higher precision/recall rate. The study outcome possess' better precision score, recall, and lowered processing time as compared with the most relevant work text mining.**

Keywords—Text Mining; Collaborative Platform; Probability Theory; Heterogeneous Domain; Precision /Recall

I. INTRODUCTION

In the area of sales and marketing, customer generated response plays a significant role in shaping the behaviour of prospective customer behaviour over e-commerce or m-commerce application [1] [2]. In existing system, majority of such responses are in terms of text on specific language which are generated every seconds in massive quantity [3] [4]. Although, there is a dedicated server or storage to save such massively generated text, but it is of no use until and unless some analysis is carried out on it. Text-mining is one such operation that applies the principle of data mining over the text of the contents in order to extract certain valuable knowledge from the text [5] [6]. However, the challenging part of the task is about the types of responses which are normally noun (for product) or adjectives (for response type). However, occurrences of such words are sometimes dynamic owing to the user's behaviour that causes the machine to hardly understand the contextual meaning [7] [8]. Somehow if the meaning or knowledge is extracted from one dataset, the problem occurs as there is an uncertainty if the same algorithm could be deployed for different text without any change. Hence, performing knowledge extraction using text-mining approach considering heterogeneity in domain is one of the challenges that still the research community is trying to deal

with. A response target could be represented as an object or product or services that the user expresses its response, usually they are noun [9] [10]. A simple example to understand this is consider i) a mobile user expressing a response "Bright Screen Resolution with responsive interface", ii) a washing machine user expressing a response as "Contrast LED buttons with responsive screen", and iii) a movie goer expresses as "good multiplex with 7 screens". A closer look into the entire three different domains has same object i.e. screen, but in every place it bears different meaning and context. Hence, it is not possible to write a single query about the object screen in this case and this is all because of the adjectives connected to it as bright, responsive, and 7 screens. A closer look will also show that an adjective responsive is used in different way in 1st and 2nd example that completely have different context. Hence, the problem becomes worst when the dataset heterogeneity is quite high. Hence, this paper presents a technique where heterogeneity of the responses is considered as a challenge in the viewpoint of text mining approach and hence is solved using a simple mathematical modelling that ensures faster response time. Section A discusses about the background of the study followed by problem identification in Section B. The highlight of proposed solution is given in Section C followed by algorithm implementation in Section II. The accomplished result of the study is discussed in Section III followed by conclusion of the paper in Section IV.

A. Background

This section discusses about the existing research contribution in text mining. Our prior study has already reviewed about effectiveness and issues in existing techniques [11]. Hence, this section will update more research work pertaining to text mining. Li et al. [12] have presented a framework for exploring the relevancy among the documents using text mining approach in order to excavate more information about document level feature extraction. A tree-based mechanism for identifying the interaction of a person was introduced by Chang et al. [13] by representing semantics, context, and syntactic data over a convolution kernel. A problem of higher dimension of text mining over cloud-based big data was introduced by Vatrapu et al. [14]. An analysis of social set is presented that uses set theory and big data in order to understand the significance in contextual terms involved in user-defined responses. Jiang et al. [15] have presented a graph-based technique applicable in biomedical sector based

on word analogy. Artificial Neural Network was used for modelling the system of mapping multiple relationships among the words. The study outcome was testified using vector length, size of corpus, and iteration. Brown [16] has carried out a study about implying potential of using text mining in order to investigate probabilities of rail accidents. Aggarwal et al. [17] have discussed over using an algorithm in order to develop a clustering mechanism. There are also studied pertaining to visual-analytics which is recently gaining a good pace. One of such work has been carried out by Liu et al. [18] using dynamic Bayesian network. The authors have designed a visualization based on sedimentation for interactive streaming of textual data with precise detailing. A mathematical modelling was introduced that uses streaming tree cut approach along with quantitative evaluation method. Using ontology over text mining was seen in the work carried out Rajpathak et al. [19] using D-matrix. The target of this study was to investigate the domain of diagnosis and its associated fault followed by implication of text mining algorithm. Ontology was introduced to check the artefacts. Chen et al. [20] have presented a text mining model that searches for shared concept with minimal ranking. The technique was found to reduce the gap of distribution between domain source and domain targeted. Ma et al. [21] have also used ontology with text mining in order to perform selection of research-based projects. The technique was implemented on both English as well as Chinese text using supervised learning algorithm as well as evolutionary learning algorithm. The study outcome was validated with respect to precision and recall rate. A generative topic framework is introduced by the author on asynchronous textual contents [22]. Zhong et al. [23] have introduced a technique of pattern discovered in order to further leverage the performance of text mining. A simple pattern-based taxonomy framework has been created. The assessment of the technique was carried out using massive dataset. Ghose and Ipeirotis [24] have presented a study that measures the impact of customer-generated reviews towards the financial aspect of product sales. Malin et al. [25] have introduced a technique where the semantic-based annotations were deployed in order to enhance the performance of knowledge discovery on text-mining approach. Usage of text mining was also seen to be applied on bioinformatics by Dai et al. [26] as well as identification of biomarkers by Li [27]. Similar direction of work is also studied by Qazi et al. [28] who have introduced a technique of feature representation using sequential mixed rule. The author has also used supervised machine learning algorithm in order to solve the problem associated with opinion mining approach. Using bag-of-words model, the proposed study has shown better text mining performance in the viewpoint of target identification and handling negation. The next section discusses about the problems that are associated with the existing system.

B. The Problem

The problems that are identified from the contribution of the existing research work are as follows:

- *Complex Modelling:* Majority of the existing technique has deployed supervised learning technique which is not only computationally complex but its accuracy in outcome highly depends on training data size. It is good

for offline analysis but not applicable for online analysis for larger size of text file.

- *Less Focus on Response:* Maximum research work has focused on accomplishing accuracy, precision, recall rate, etc., but there is less focus on accomplishing lowered algorithm processing time on massive dataset. Faster response time is an essential criterion for time-bound applications, which cannot be seen in existing studies.

- *Lesser effective mathematical modelling:* Although, there are certain degrees of mathematical-based research modelling, but the extent of such studies are pretty less. Another bigger problem in the existing system is adoption of recursive function in modelling, which is not only time consuming but also is resource hungry application.

Therefore, the proposed study has identified the above three points as open research issues towards leveraging the performance of text mining operation. The problem statement can be defined as: *It is quite a challenging task to introduce a simple and non-recursive modelling approach towards text mining that has faster response time and higher precision/recall on complex dataset.*

C. The Proposed Solution

The prime purpose of the proposed system is to introduce a simple and yet robust architecture of text mining in order to extract knowledge from various responses of the user pertaining to different context called as domain. The schematic architecture of proposed solution is highlighted in Figure 1.
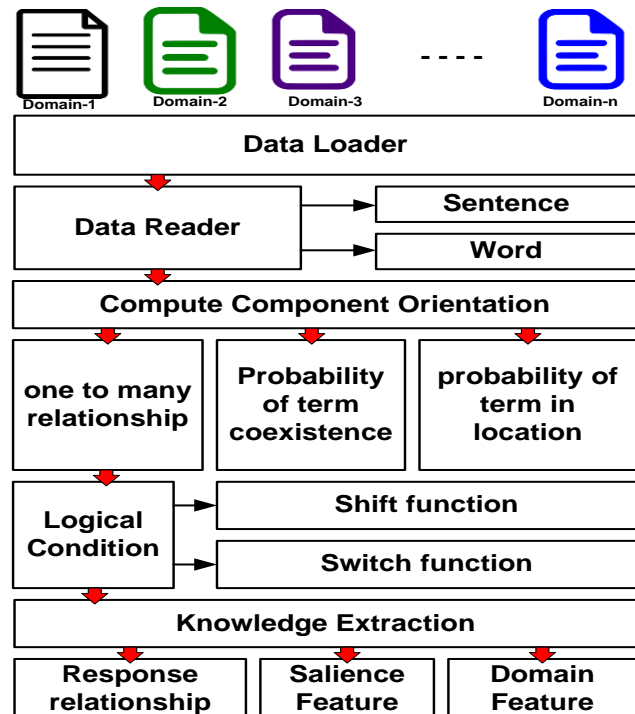

Fig. 1. Schematic Diagram of Proposed RKE-CP

The proposed system takes the input of various responses from multiple domain and feeds to the data loader which is responsible for converting the text into machine readable strings. The mechanism than subjects the strings to data reader performs segregation of sentences and words. The next step is to perform computation of the component (or word) orientation, which is a mechanism of exploring the context and location of various significant terms. Applying undirected graph, $G=\{\chi_1, \chi_2, \chi_3\}$, the study constructs graph for response relationship. χ_1 represents vertices with component targeted response and words of response, χ_2 represents edges meaning that there is connectivity between two different response residing in two different vertices, and χ_3 represents weight assigned to the edge signifying response that is associated with these two response vertices. The next part of the study implement the concept of word orientation which is basically a mathematical modelling integrating three components: (1) one-to-many relationship, (2) probability of term existence, and (3) probability of term in particular location. A logical condition is then constructed that assess the limits of shift and switch function responsible for rectifying the context of the domain. This function will eventually eliminate any possibility of false positive among the different words from different domain bearing similar individual meaning but different contextual meaning as a whole. Finally, the extraction of the knowledge is carried out using further three empirical functions: (1) response relationship, (2) salience feature, and (3) domain feature. The next section discusses about the algorithm implementation followed by results obtained by implementing the algorithm.

II. ALGORITHM IMPLEMENTATION

The algorithm targets to apply a novel text mining approach in order to perform extract knowledge of diversified user's responses on a collaborative platform. The algorithm takes the input of ρ (component orientation), σ (complete sentence), η (number of words), α_i (total words that maps with δ_i), δ_i (individual i^{th} word), β (coexisting data attribute), τ (shift function), and γ (switch function), which after processing yields an output of K (Knowledge Discovery). For the proposed algorithm to be functional, it is necessary to keep its textual data to be highly domain specific. It will mean that each text file will be pertaining one specific and dominant domain. The algorithms take the input of text and organize a matrix form of it in order to apply mining approach. It then computes the cumulative possible word orientation ρ (Line 1), where, $\rho=\{(i, a_i)|i\epsilon[1, \eta], a_i\epsilon[1, \eta]\}$ and σ empirically represents a set equivalent to $\{\delta_1, \delta_2, ..., \delta_\eta\}$. The variable a_i will represent the noun word at i^{th} position of the text. A novel function (Line 2) is used for exploring the unique information from the data captured from the collaborative platform that uses three different components in order to perform modelling of potential relationship among the textual responses. The first component $\eta (\alpha_i | \delta_i)$ represents a set of one to too many relationship that will mean that any single word can be used to amend the meaning of other words. The variable α_i will represent total terms that are mapped with δ_i. The second component $\beta (\delta_i | \delta_{aij})$ empirically represents the probability that the term δ_i could possible co-exist with δ_{aij} for a given corpora. This will also mean that if there is more probability of any specific word in order to change the meaning in another word

i.e. noun. Then that particular term δ_i will have maximum value of $\beta (\delta_i | \delta_{aij})$. The steps involved in the proposed study are as follows:

Algorithm for RKE-CP
Input: $\rho, \sigma, \eta, \alpha_i, \delta_i, \beta, \tau, \gamma$.
Output: K
Start
1. $\rho^*=\arg_{max} f(\rho|\sigma)$
2. Apply function for word orientation for a given text

$$f(\rho | \sigma) \propto \sum_{i=1}^{\eta} \eta(\alpha_i | \delta_i).\sum_{j=1}^{\eta} \beta(\delta_j | \delta_{a_j})\theta(j | a_j, \eta)$$

3. **While** $\tau_{i1, j1}>1$ or $\gamma_{j1,j2}>1$ **do**
4.   **If** $(j_1, a_2)\neq \rho$ or $(j_2, a_{j1}) \neq \rho$ **then**
5. $\sigma=-1$
6.   **End**
7.   **If** $\tau_{i1, j1}>\gamma_{j1,j2}$ **then**
8.   Update τ, γ
9. **End**
10. **End**
11. Compute $prob(\delta_t|\delta_o)=card(\delta_t, \delta_o)/card(\delta_o)$
12. $K= [R_{rel}, S_F, D_F]$
End

The last component $\theta(j | a_j, \eta)$ basically represents the data about term location for a given text in order to represent a probability of a particular term residing in location a_j considered to be mapped with the location j of another word. We also introduce a variable called as τ and γ representing shift and switch function that will be computed as follows:

$$\tau_{i,j} = (1-\lambda(a_j,i)).\frac{P(t_{i,j}(a) | \phi,\upsilon)}{P(a | \phi,\upsilon)} \qquad \text{and}$$

$$\gamma_{j_1,j_2} = \begin{cases} (1-\lambda(a_{j_1}, a_{j_2})).\dfrac{P(\varepsilon_{j_1,j_2}(a)|\phi,\upsilon)}{P(a|\phi,\upsilon)} & a_{j_1} \leq a_{j2} \\ 0 & otherwise \end{cases} \qquad (1)$$

The above mathematical equation (1) shows the technique of computing τ and γ that is used over logical condition in Line 4 of algorithm. In the above equation, the variable P represents probability, λ represents statistical significant factor, while the variable ϕ and ϑ represents lower and higher significance factor. The variable ε is a subset of a sentence. Basically, Line 7-9 in the algorithm assists in filtering the unnecessary data using statistical approach and considers only the necessary information, which will be only considered in knowledge extraction process. Finally, a probability for ultimate text orientation between the two terms is computed as the cardinality (*card* in Line 11). The algorithm also computes the response relation attributed R_{rel} as equivalent to $[h. P (\delta_t|\delta_o) + (1-h)P(\delta_o|\delta_t))^{-1}$. In this expression, the variable h $(=0.05)$ represents progression coefficient in order to integrate the two orientation probabilities. Along with R_{rel}, S_F and D_F are also computed. S_F represents feature for score of prominent condition computed using TF-IDF while D_F represents domain feature computed using the technique discussed by Hai et al. [29]. The next section discusses about the results being accomplished after implementing the proposed system.

III. RESULT ANALYSIS

This section discusses about the outcomes being accomplished from the proposed study where the assessment was carried out over normal 32 bit windows machine. The study also uses standard lexical database of Word Net. A synthetic dataset is built up considering a text file with different types of domain with sentences size ranging between 5000 and 10,000. All the dataset used are large in size and hence a crawling is done arbitrarily over such larger sizes of sentences. The targets and the terms of the response are subjected to manual annotation. Table 1 highlight the data used adoption for carrying out evaluation.

TABLE I. DATA SET CONSIDERED FOR ANALYSIS

Domain	Sentence	#Response Term	#Response Target
Mobile	5000	555	959
Hotel	5000	576	983
Temple	4500	329	897
Song	6000	654	965
Mall	7000	123	934
Vehicle	7500	675	955
School	8000	455	989
Hospital	8500	541	928
Computer	10000	512	915

The study outcome of proposed system is compared with that of the Lin et al. [30] and Yano et al. [31] who have carried out similar sort of research on text mining. Lin et al. [30] have introduced an entity recognition system carried out on massive data of social networking sites. The authors have used it for extracting a term for medical significance using unique representation techniques of word that consists of methods for embedding words, normalization of tokens, and usage of global vectors. Usage of n-gram tokenisation technique was seen in the work of Yano et al. [31] who have carried out text mining over similar datasets as that of Lin et al. [30] for extracting significant behaviour. The words that are specific to domain were extracted using Bayes classifier and n-gram tokenise. The study outcome of both the techniques has been mainly assessed using precision and recall factor and hence, we choose to retain the same for comparative analysis. The outcome of the study in Figure 2 highlights the precision accomplished from the different forms of techniques of text mining. The technique of Lin et al. [30] has usage of extraction of knowledge using much number of features but in including any feature that could minimise the data redundancies over different corpus. This leads to lowered precision of Lin et al. [30] model. Similarly, Yano et al. [31] model has implemented a naïve Bayes Classifier using heavy logs of behaviour history. The process ensure better mapping with behaviour but also yields false positive when a discrete behaviour is not found to be matching with trained database
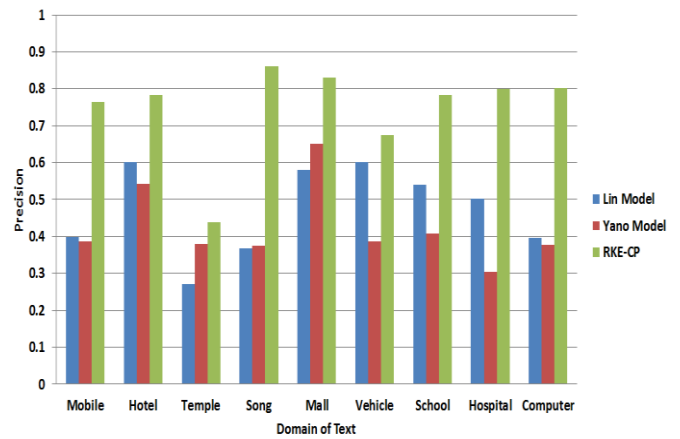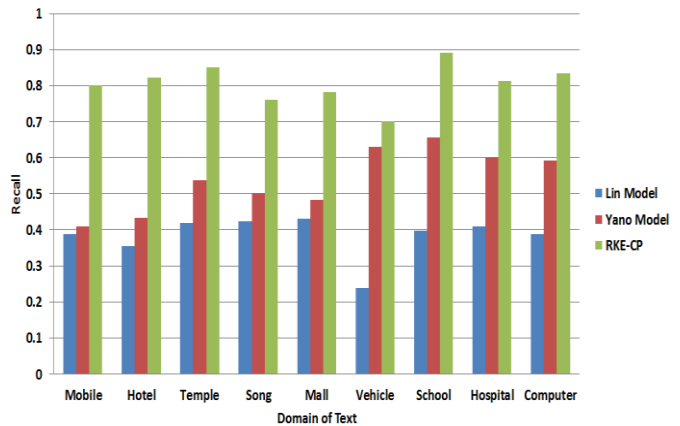

Fig. 2. Comparative Analysis of Precision

Fig. 3. Comparative Analysis of Recall

Figure 3 highlights the recall scores of the proposed study as well as existing system. The outcome shows that RKE-CP has better recall performance as compared to Lin et al. [30] and Yano et al. [31]. The proposed system has the benefit of better decision making on various domains applicable for larger size of text document. Once the data is subjected to data reader, the complete analysis starts by extracting sentence followed by extraction of words. The algorithm works after that. This phenomenon causes enough reduction of redundant data once it is again subjected for similarity check followed by operations involving TF-IDF causing further narrow down of higher dimensionality of the data. Therefore, the algorithm has dual advantages i.e. (1) reduction of storage or memory as dimensionality of the data reduces and increases accuracy indirectly, and (2) decrease of processing time.
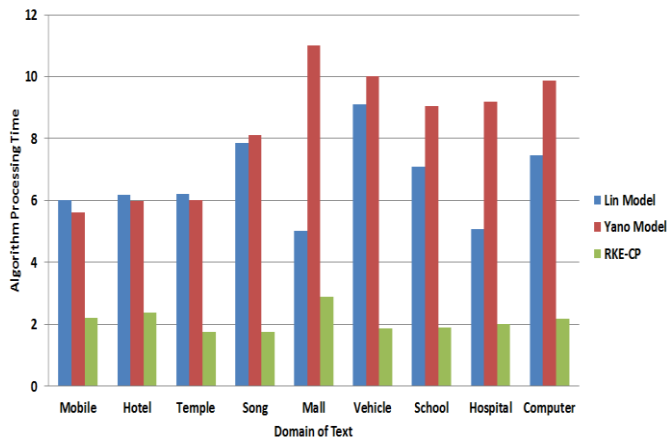
Fig. 4. Comparative Analysis of Algorithm Processing Time

Figure 4 showcases the algorithm processing time of the proposed system and existing system. The prime reason of the proposed system to have lowered processing time even without using any form of conventional learning algorithm is its simple mathematical modelling using orientation of the words. Majority of the problems in handling complexity of text mining was eliminated by applying function of word orientation as well as the logical conditions specified in the steps of algorithm. The algorithm processing time was computed in terms of seconds in normal core i5 processor and 4 GB RAM with same data sets on both proposed and existing system.

IV. CONCLUSION & FUTURE SCOPE

With the increase of cloud adoption and mobile networks, text-based response will always keep on increasing and a new challenge will be met in order to extract a significant level of knowledge from it. Response and their respective analysis with respect to text-mining turns up into a fascinated research domain because of an accessibility of a vast amount of user commented contents in survived sites, discussions and an online journals. Various text-based responses have a wide range of different fields signifying different meaning and context to each sentence. In the proposed system, emphasis has been laid to the fact that when a specific word adds to any noun there are chances that it may change its meaning and complete context too sometimes. This is really confusing from machine-based adaption viewpoint leading to a problem in text-mining. Therefore, the proposed research work introduces a technique called as RKE-CP which is supported by a simple mathematical modelling with its steps carefully controlled by using probability theory. Usage of probability theory has its own advantage in controlling the uncertain over controlling the precision rate of proposed study. One of the significant contribution/novelty of proposed study is that i) it offers a text-mining algorithm with faster response time, ii) its mathematical modelling leads to elimination of various redundancies and complexities that are highly suitable in larger and complex text dataset, and iii) the nature of algorithm processing is quite responsible for dimensionality reduction of the data to be mined and thereby it leads to faster algorithm processing time.

Our future work will be in the direction of further optimizing the mining performance. This can be carried out by designing a novel framework using keyword based text mining

as well as using multivariate analysis methodology. Our first initiative will be to develop a mechanism of heterogeneous extraction of keywords using keyword clustering process as the first step of optimisation. This work will be targeting to accomplish reduced processing time. The next consecutive work will be to use semantic-based concept using context-based annotation over heterogeneous domains over contextual document. The primary target will be to achieve higher accuracy and minimised processing time.

REFERENCES

[1] S. Kudyba, "Big Data, Mining, and Analytics: Components of Strategic Decision Making", CRC Press, pp. 325, 2014

[2] M. Chen, S. Mao, Y. Zhang, V.C.M. Leung, "Big Data: Related Technologies, Challenges and Future Prospects", Springer, pp. 89, 2014

[3] R. Zhang, A. Zhou, W. Yu, Y. Gao, P. Chao, "Review Comment Analysis for E-commerce", World Scientific, pp. 172, 2016

[4] Z. Lu, "Information Retrieval Methods for Multidisciplinary Applications", Idea Group Inc (IGI), pp. 325, 2013

[5] S.M. Weiss, N. Indurkhya, T. Zhang, "Fundamentals of Predictive Text Mining", Springer, pp. 239, 2015

[6] G. Wiedemann, "Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany", Springer, pp. 294, 2016

[7] M.W. Berry, "Survey of Text Mining: Clustering, Classification, and Retrieval", Springer Science & Business Media, pp. 244, 2013

[8] Bhatnagar, Vishal, "Data Mining in Dynamic Social Networks and Fuzzy Systems", IGI Global, pp. 412, 2013

[9] G.Miner, J. Elder, A.Fast, T. Hill, Robert, "Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications", Academic Press Mathematics, pp. 1000, 2012

[10] G.Chakraborty, M. Pagolu, S. Garla, "Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS", SAS Institute Mathematics, pp. 340, 2014

[11] G. Jalaja, N. Sajitha, K.R.U.Kumar, "Reviewing the Pathway of Text-Mining Approaches to Gauge the Applicability in Data Analysis", International Journal of Computer Applications, Vol. 125, No. 3, 2015

[12] Y. Li, A. Algarni, M. Albathan, Y. Shen and M. A. Bijaksana, "Relevance Feature Discovery for Text Mining," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1656-1669, June 1 2015.

[13] Y. C. Chang, C. C. Chen and W. L. Hsu, "SPIRIT: A Tree Kernel-Based Method for Topic Person Interaction Detection," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2494-2507, Sept. 1 2016.

[14] R. Vatrapu, R. R. Mukkamala, A. Hussain and B. Flesch, "Social Set Analysis: A Set Theoretical Approach to Big Data Analytics," in *IEEE Access*, vol. 4, pp.2542-2571, 2016.

[15] Z. Jiang, L. Li and D. Huang, "An Unsupervised Graph Based Continuous Word Representation Method for Biomedical Text Mining," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 13, no. 4, pp. 634-642, July-Aug. 1 2016.

[16] D. E. Brown, "Text Mining the Contributors to Rail Accidents," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 346-355, Feb. 2016.

[17] C. C. Aggarwal, Y. Zhao and P. S. Yu, "On the Use of Side Information for Mining Text Data," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 6, pp. 1415-1429, June 2014.

[18] S. Liu, J. Yin, X. Wang, W. Cui, K. Cao and J. Pei, "Online Visual Analytics of Text Streams," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 11, pp. 2451-2466, Nov. 1 2016.

[19] D. G. Rajpathak and S. Singh, "An Ontology-Based Text Mining Method to Develop D-Matrix From Unstructured Text," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 7, pp. 966-977, July 2014.

[20] B. Chen, W. Lam, I. W. Tsang and T. L. Wong, "Discovering Low-Rank Shared Concept Space for Adapting Text Mining Models," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1284-1297, June 2013.

[21] J. Ma, W. Xu, Y. h. Sun, E. Turban, S. Wang and O. Liu, "An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 42, no. 3, pp. 784-790, May 2012.

[22] X. Wang, X. Jin, M. E. Chen, K. Zhang and D. Shen, "Topic Mining over Asynchronous Text Sequences," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 1, pp. 156-169, Jan. 2012

[23] N. Zhong, Y. Li and S. T. Wu, "Effective Pattern Discovery for Text Mining," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 1, pp. 30-44, Jan. 2012.

[24] A. Ghose and P. G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1498-1512, Oct. 2011.

[25] J. Malin, C. Millward, F. Gomez and D. Throop, "Semantic Annotation of Aerospace Problem Reports to Support Text Mining," in *IEEE Intelligent Systems*, vol. 25, no. 5, pp. 20-26, Sept.-Oct. 2010.

[26] H-J. Dai, C-H.Wei, H-Y.Kao, R-L.Liu, R.T-H.Tsai, and Z.Lu, "Text Mining for Translational Bioinformatics", Hindawi Publishing Corporation, pp. 2, 2015

[27] H. Li and C. Liu, "Biomarker Identification Using TextMining", Hindawi Publishing Corporation, pp. 4, 2012

[28] A. Qazi, R. G. Raj, M. Tahir, E. Cambria, and K. B. S. Syed, "Research Article Enhancing Business Intelligence by Means of Suggestive Reviews", Hindawi Publishing Corporation, pp. 11, 2014

[29] Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, "Identifying features in opinion mining via intrinsic and extrinsic domain relevance," IEEE Trans. Knowledge Data Eng., vol. 26, no. 3, p. 623–634, 2014.

[30] W. S. Lin, H-J. Dai, J. Jonnagaddala, N-W. Chang, T.R. Jue, "Utilizing different word representation methods for twitter data in adverse drug reactions extraction", *Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, Tainan, pp. 260-265, 2015

[31] Y. Yano, T. Hashiyama, J. Ichino, and S. Tano, "Behavior extraction from tweets using character N-gram models", In IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1273-1280, 2014