International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 5, October 2020, pp. 5526~5534
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i5.pp5526-5534

Sentimental classification analysis of polarity multi-view textual data using data mining techniques

Ali Hameed Yassir[1], Ali A. Mohammed[2], Adel Abdul-Jabbar Alkhazraji[3], Mustafa Emad Hameed[4], Mohammed Saad Talib[5], Mohanad Faeq Ali[6]
[1]College of Computer Science and Information Technology, University of Sumer, Iraq
[2,6]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
[3]Computer Science Department, College of Science, University of Diyala, Iraq
[4]Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
[5]College of Administration and Economics, University of Babylon, Iraq

Article Info

Article history:

Received Nov 28, 2019
Revised Apr 25, 2020
Accepted May 7, 2020

Keywords:

Data mining
Data science
Polarity text data
Sentiment analysis
Text data mining

ABSTRACT

The data and information available in most community environments is complex in nature. Sentimental data resources may possibly consist of textual data collected from multiple information sources with different representations and usually handled by different analytical models. These types of data resource characteristics can form multi-view polarity textual data. However, knowledge creation from this type of sentimental textual data requires considerable analytical efforts and capabilities. In particular, data mining practices can provide exceptional results in handling textual data formats. Besides, in the case of the textual data exists as multi-view or unstructured data formats, the hybrid and integrated analysis efforts of text data mining algorithms are vital to get helpful results. The objective of this research is to enhance the knowledge discovery from sentimental multi-view textual data which can be considered as unstructured data format to classify the polarity information documents in the form of two different categories or types of useful information. A proposed framework with integrated data mining algorithms has been discussed in this paper, which is achieved through the application of X-means algorithm for clustering and HotSpot algorithm of association rules. The analysis results have shown improved accuracies of classifying the sentimental multi-view textual data into two categories through the application of the proposed framework on online polarity user-reviews dataset upon a given topics.

Corresponding Author:

Ali Abdul-Jabbar Mohammed,
Faculty of Information and Communication Technology,
Universiti Teknikal Malaysia Melaka (UTeM),
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
Email: aliabduljabar22@yahoo.com

1. INTRODUCTION

In the digital age of the rapid shift to information era, enormous amounts of textual data have achieved by the participating in virtual communities on behalf of sharing data and information, social communications and daily transactions [1]. Manually processing of this huge amount of unstructured textual data and information is very costly in term of time-consuming and expensive process. Accordingly, taking advantages of this type of data and information is very hard for the reason of its availability in various forms such as unstructured or semi-structured text data forms [2, 3]. The majority of indexed resources of this textual data are mainly extracted from online reports and documents, e-mails, online textual blogs, memos, digital letters, etc [3]. Textual data resources may possibly consist of data collected from multiple

information contributors with different representations and usually handled by heterogeneous analytical models. This diversity in the textual data sources and procedures forms the sentiment multi-view textual data. This type of data comprise unique information features that involve information have the same instance but with various representations or have different instances with same representation [4]. Thus, the efficient understanding of sentiment multi-view textual data is becoming an extremely critical movement by the way of using data mining methods. The main promise of data mining tools and techniques to take advantage of multi-view textual data is the potential speed up of big data era [5]. Various methods and strategies of data analysis are proposed for delivering online and/or integrated solutions to solve complexities associated with the management of various data forms in sentiment multi-view textual data. The mixed methods and likewise the hybrid methods of text data mining technologies may be classified as the modern practice for analyzing sentiment multi-view textual data [6].

Textual Data Mining is "roughly equivalent to text analytics, refers to the process of deriving high-quality information from text" [7]. Moreover, it is as well considered as a path of data mining field when data mining algorithms and methods are applied for the discovery of knowledge in the textual data forms. Many of text data mining techniques can be individually applied or incorporation with other techniques for discovering sensible and valuable knowledge. Although, text data mining has achieved various successful applications in many areas, some deep sentiment text analysis tasks still remain as a challenge and not handled yet. The real distinction of technological efforts with text data mining from the general concept of data mining is the possibilities to deal with the special characteristics of implicit information in textual data that seems to be primarily unstructured. However, the typical practice of such technological possibilities is yet in the early stages to reveal the necessary knowledge that extracted from this unstructured textual data [8]. Therefore, the key aspect motivated this paper is that multi-view textual data can be characterized as a very complex implicit data structure. In this context, the main problem to deal with this type of data is to design hybrid techniques of data mining algorithms that able to handle implicit data and information. Additional combined efforts of knowledge discovery applications have the possibilities to reveal associating common information patterns in the multiple views or modalities of textual data formats. At Present, there is fewer addressing methods of data mining and machine learning for learning from these semi-structured or unstructured data types in the textual data formats [9].

M. Jakubik [10] has summarized knowledge creation management with incorporation of gathering and sharing information, and effectively developing this information and then shifting the knowledge to demanders. This research is going to investigate the track outlined by M. Jakubik, [10] for managing the creation of knowledge but then will construct this research by operating the tools and techniques of data mining in multi-view textual data. Consequently, the primary purpose of conducting this research is in the direction of find out in which way can automatically classify knowledge and more accurately identified classes from unstructured textual data documents. Consequently, it will offer further knowledge that can be realized, improved, and delivered to the demanders.

2. RELATED WORK

In this section, a brief description is provided to assess the existing approaches of textual sentiment analysis and the distinction between the proposed framework and related previous works. The textual sentiment analysis is very well-known as part of data mining approaches, which can essentially classify the sentimental multi-view textual data into two categories as positive or negative [11]. In this research, the proposed framework is applied on online polarity user-reviews to be classified as 'good' or 'bad' sentiments. In actual fact, extracting knowledge from information available in textual data is increasingly gaining more consideration. This consideration by related previous works have developed a number of sentiment analysis methods to measure the opinions of users about products or services. As example of early developed methods, Reddick et al. [12] investigated features of text in which social media by some public services delivery. Likewise, Müller et al. [13] had applied a variety of text analytic approaches to understand obstacles related to customer service. Some data mining approaches have involved in the knowledge discovery from sentimental multi-view textual data, such as features extraction of online user generated contents [14], to prediction and classification [15]. In the related research of user's reviews, a sentiment classification of movie reviews has been handled by Ahuja et al. [16] using dual training and dual prediction while addressing polarity shift. Consequently, Zola et al. [17] have propose a novel sentiment classification, in which a three-step methodology is explored based on balanced training, text preprocessing and machine learning using two languages: English and Italian.

In this context, the data mining algorithms are engaged strongly for knowledge creation from sentimental textual data. As examples of using data mining algorithms, Wang et al. [18] introduced a convolutional recurrent neural network for classifying text documents, using support vector machines

(SVM) to influence of word normalization in text classification [19, 20], using naive bayes for optimal feature selection in text categorization by Tang et al. [21], and using naive bayesian algorithm by Tang et al. [22] for text unstructured classification in the Spark computing environment. Even though some of those solutions happen to be highly effective, the challenges continue to be relevant to the variation of multi-view polarity textual data that required preprocessing, anddeciding on the proper classification procedure [23]. Therefore, this research is focusing on the difficulty of providing improved accuracies of classifying the sentimental multi-view textual data into two categories through the application of the proposed framework on online polarity user-reviews dataset upon a given topics.

3. RESEARCH METHOD

The method strategy that has been designed for this paper is divided into conceptual development stage of the proposed framework which relies on the combined tools and techniques of data mining. This combination of data mining algorithms can serve the purpose of handling unstructured textual datasets. The complete procedure involved with identifying comprehending knowledge from unstructured textual data sources is simply implemented using the direction outlined in a matter of this section. The following actions had been experienced to improve and refine the development of the framework design:

- Identification of various source information regarding textual analysis data mining.
- Evaluation and analyzing available techniques and approaches of data mining for analysing textual databases.
- Determining beneficial data mining algorithms to design hybridized framework for analysing textual databases.

The next stage is the design and implementation of the framework to obtain the results and verify the objectives of this research.

3.1. Conceptual development of algorithms

3.1.1. X-means Clustering Algorithm

The X-Means clustering algorithm proposed by Pelleg and Moore [24] is a K-Means extended by an Structure Improved part. The centroid in this part of the algorithm tries to be divided automatically in its area. The join decision among the children of each centroid with the centroid itself is finalized comparing with the Bayesian Information Criterion (BIC) values. Nevertheless, the aim of X-means algorithm is to deliver a fast and effective way to cluster unstructured data.

The X-Means clustering algorithm allows to specify the range number of clusters. It starts to run with the minimum number of clusters (K) of the given range and assigns a BIC score and it continues to add centroids where they are needed until the maximum number of clusters (Kmax) range is reached. Throughout the mentioned process, the centroid set will be recorded when it is achieving the best score, and this becomes the final centroid set output. Each model has (K) to (Kmax)+1 numbers of centroids and a BIC score which will decides the fit number of clusters. Thus, it is automatically selecting the optimal resultant of clusters number. The steps flow of the X-Means clustering algorithm is given in the Figure 1, as defined in [24].

The X-means Clustering functions is associated with strengths for further improve the knowledge discovery within multi-view textual dataset in terms of the proposed framework implementation. The clustering task process is repeated with updating the best BIC score till the maximum number of clusters (Kmax) range is reached which is provides a typical number of clusters. Also, this will affects positively the required processing time. Furthermore, the use of concurrency speeds up the clustering process and the use of the BIC gives a mathematically complete measure of quality.

3.1.2. HotSpot algorithm of association rules

HotSpot algorithm of association rules finds out a set of rules presented in structure similar to a tree, which is maximize or minimize a target value of interest [25, 26]. The HotSpot algorithm with a nominal target looks for segments of the information where there is a high chance of a minority value occurring by given the constraint of a minimum support, and with a numeric target could be attracted in discovery of segments where they are higher than average in the whole dataset. For example, in the sentiment multi-view textual data scenario, it attempts to find in which positive or negative textual review groups are at the highest emotional claim ratio, or in which positive or negative textual review groups have the highest average sequence of sensitive issues in sentimental reviews dataset. This algorithm is close to the PRIM bump hunting algorithm described by Friedman and Fisher, [27].

The Single-key Information sequences were recently linked with identified clusters using X-means clustering algorithm, which is used in this research for separation to a sentimental review categories. The goal here is to identify characteristics of each textual document as positive or negative user-review

categories, where performing HotSpot association rules later for multi-view textual data is significantly helpful in terms of maximizing the target value of interest in each cluster to obtain the maximum concentration improvement by recognizing the frequent individual terms in each cluster and combines them to a larger term sets as long as those term sets appear appropriately frequented in the database. The whole algorithm can be divided into four steps as the Figure 2.

The automatic discovery of frequent term sets or multi-key frequent information rules (MKFIR) is a very promising area of data mining applications in the field of multi-view textual data. A frequent term set is described as being a collection of key phrases which usually takes place inside the text document set ın excess of the minimum support times.

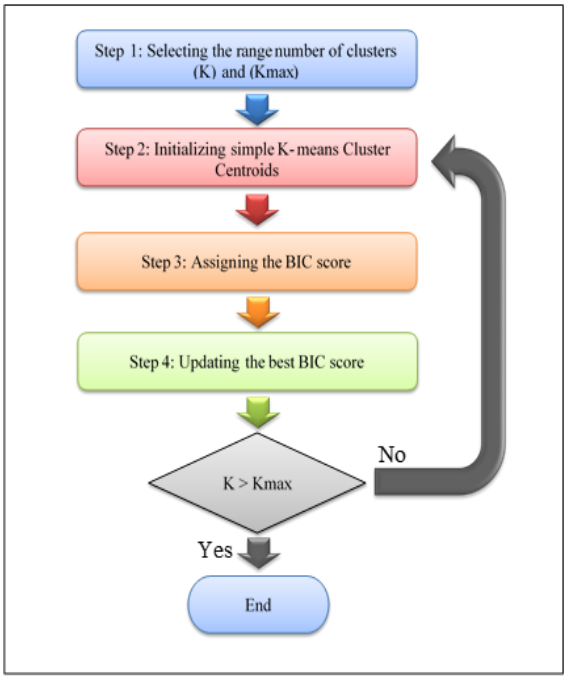

Figure 1. Flow of information in x-means clustering algorithm

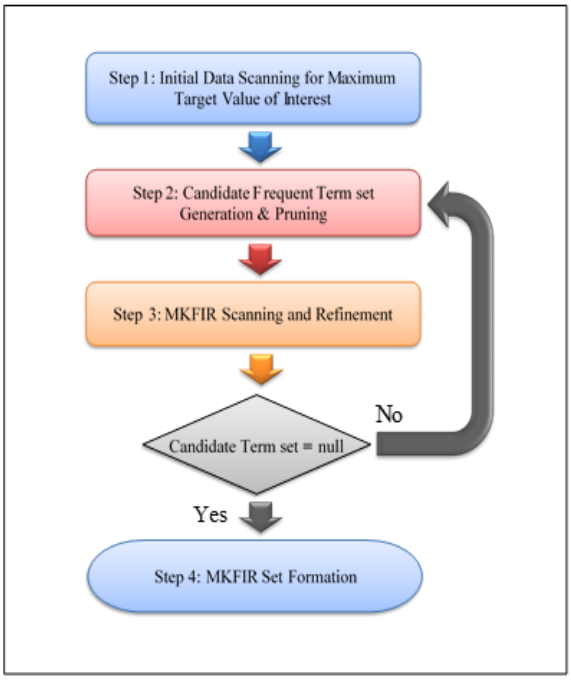

Figure 2. Flow diagram for hotspot algorithm for MKFIR formation

3.1.3. Naive bayes and decision trees (J48) algorithms for classification

In this stage, two different classifiers, in particular decision trees (J48) and naive bayes classification algorithms are compared and applied in this research. The purpose of analyzing each of these distinct classification algorithms lies in the fact that each uses different selection criterias to the critical information variables. The J48 classification algorithm is based on decision tree. Naïve bayes classification algorithm is based on probability [28]. Successively, they extend right from the simple distance measure to other path of probabilistic distance measures to search the similarity criteria among a set of documents and then classify them directly based on their particular category. In this manner, the main objective of such experiments is to verify the proposed framework on such basis as maximizes the target value of interest by discovering multi-key frequent information rules, which delivers considerably better accuracy when compared to the simple term-based classification techniques. The Decision Trees and Naive Bayes classification algorithms has been examinedwith regards to classifying textual data in accordance with two different classes.

4. THE PROPOSED FRAMEWORK

In this section the major functions of knowledge discovery defined above are integrated to develop the proposed framework for classifying the sentimental multi-view textual data. This can be done through maximizing the target value of interest by discovering multi-key frequent information rules. The initial stages of knowledge discovery are realized by applying the X-means clustering algorithm to identify the Single-key Information sequences. Afterward, the identified Single-key Information sequences utilized by the HotSpot algorithm of association rules to generate multi-key frequent information rules.

The integrated application of these data mining algorithms which generate the multi-key frequent information rules can be utilized to shape a knowledge-based relational database for improved analyzing and handling of the sentiment multi-view textual data. The benefit with discovering the key information in terms of multi-key frequent information rules serves to automate the process of categorizing the sentimental multi-view textual documents within two predefined categories (i.e. positive or negative information classes).

Furthermore, the proposed framework is outlined to analyze sentiment multi-view textual dataset, which lies to four main data processing stages, as shown in Figure 3.

- Data preparation and preprocessing stage.
- Data clustering stage.
- Maximizing target value of interest stage.
- Classification stage.

The knowledge discovery and text classification from these four main parts is influenced by the discovery of multi-key frequent information rules. When the information is already has been structured by the data preparation and preprocessing stage, it is then passed to the data clustering stage for arranging textual data into a number of clusters and identify the Single-key Information sequences. Identifying the sequence of key terms helps ensuring some key information rules in the multi-view textual documents. In the context of current research, each document available in the sentimental multi-view textual dataset comprises a combination of terms to explicate impression. Maximizing target value of interest by utilizing the multi-key frequent information rules enables summarizing the key information within each particular document for further processing or classification.

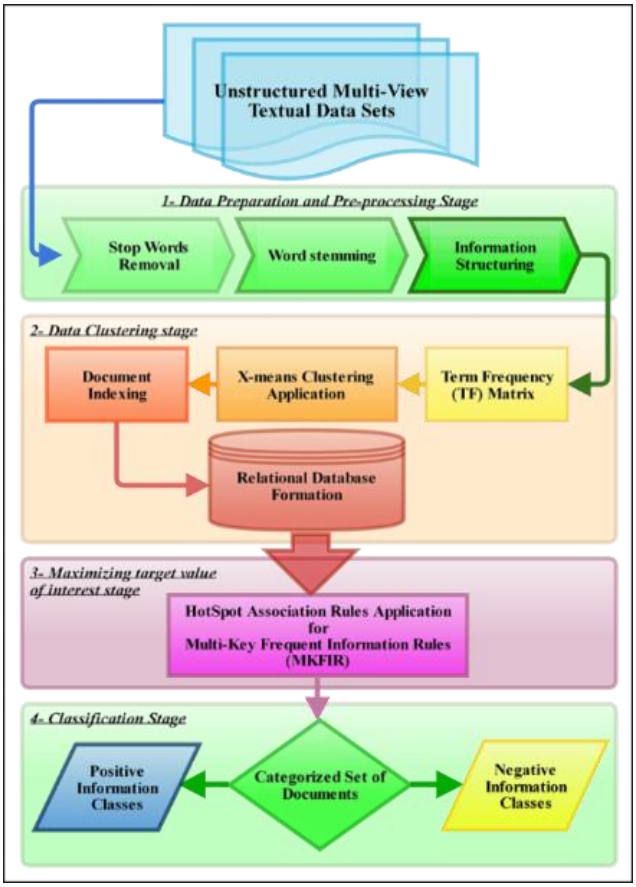

Figure 3. Multi-view textual classification system

The illustrated framework above with integration of different efforts of data mining algorithms is proposed to analyze free formatted form of multi-view textual data. The proposed framework is achieved through the application of X-means algorithm for clustering and HotSpot algorithm of association rules to discover useful knowledge in terms of multi-key frequent information rules.

5. DATA ANALYSIS

The analysis of classifying the sentimental multi-view textual data into two categories through the application of the proposed framework is discussed in the scenario of online polarity user-reviews dataset.

4.1. Data collection

In order to perform the experimentation of this research study, the performances of particular test and analysis with the proposed framework is conducted using sentiment polarity reviews dataset. The sentiment polarity reviews data provided by [29] is one of the commonly used sentiment analysis datasets. The sentiment polarity reviews dataset involves 2000 online user-created reviews on a given topic archived from the cs.cornell web portal and it is available in (http://www.cs.cornell.edu/people/pabo/movie-review-data).

4.2. Functional requirements

Weka 3.7.12 software is used to support the activities defined during the experimentation of the proposed framework, starting with the pre-processing stage toward the application of knowledge discovery level, which is used to classify the sentiment polarity reviews dataset.

4.3. Implementation of framework functionalities

The validation of the proposed framework is performed in this section in order to ensure better performance than the simple term-based classification models, through identifying key knowledge areas within sentiment polarity reviews dataset.

a. Preparation and Pre-processing of Text Documents

The data preparation and pre-processing tasks are being used to remove un-necessary terms from unstructured text data formats and prepare the textual data for further analysis of various data mining algorithms for subsequent analysis of the text. The data preparation and pre-processing tasks are including the following steps:

- Remove the stop words or terms that can be less effective in knowledge discovery from sentimental multi-view textual data.
- Perform word stemming process by way of reducing derived words or terms to its actual stem or the root form.
- Structure the information into vector form in which every term is considered a word vector.
- Represent text in a form of term frequency (TF) matrix by counting the words and their corresponding frequencies, as shown in Figure 4(a).
- Determine the similarities by calculating the Euclidean Distance method.
- Weight terms in the set of documents through Inverse document frequency (IDF) and term frequency (TF) by using its matrices (TF*IDF).
- The output is then saved in attribute relation file format (.arff) file that will be used later for further data mining analytics.

b. Application of X-Means clustering algorithm

The X-means algorithm is applied on the (.arff) file format which was obtained as a result from the preparation and pre-processing stage. This algorithm is very supportive to calculate automatically the appropriate number of generated clusters that capturing related information by using the Single-key Information sequences. The implementation experiment of this research has been shown that X-means algorithm adopts the total clusters for the used sentiment multi-view textual dataset to be two clusters as shown in the Figure 4(b). It gave substance to be very advantageous clustering algorithm in preserving valuable information structures that identified in the sentiment multi-view textual dataset. Thus, the X-means algorithm has been utilized as an efficient tool to capture valuable key Information within every single cluster, as shown in the Figure 4(c).

c. Application of HotSpot association rules algorithm

The discovery of maximum target value of interest by determining multi-key frequent information rules in multi-view textual datasets helps with finding terms that appear most frequently with each other. Knowledge discovered with the help of rules for maximum target value of interest is employed for mapping the positive or negative information document categories, as shown in the Figure 4(d).

d. Application of classification algorithms

The output of (.arff) file based on Single-key Information sequences obtained by X-means clustering and maximum target value of interest achieved by HotSpot association rules, which is then loaded into Weka 3.7.12. Two different classifier algorithms were tested to classify the sentiment multi-view textual data into positive or negative information document categories and the result accuracies that achieved by both classification algorithms are described and discussed in next (Section 5).

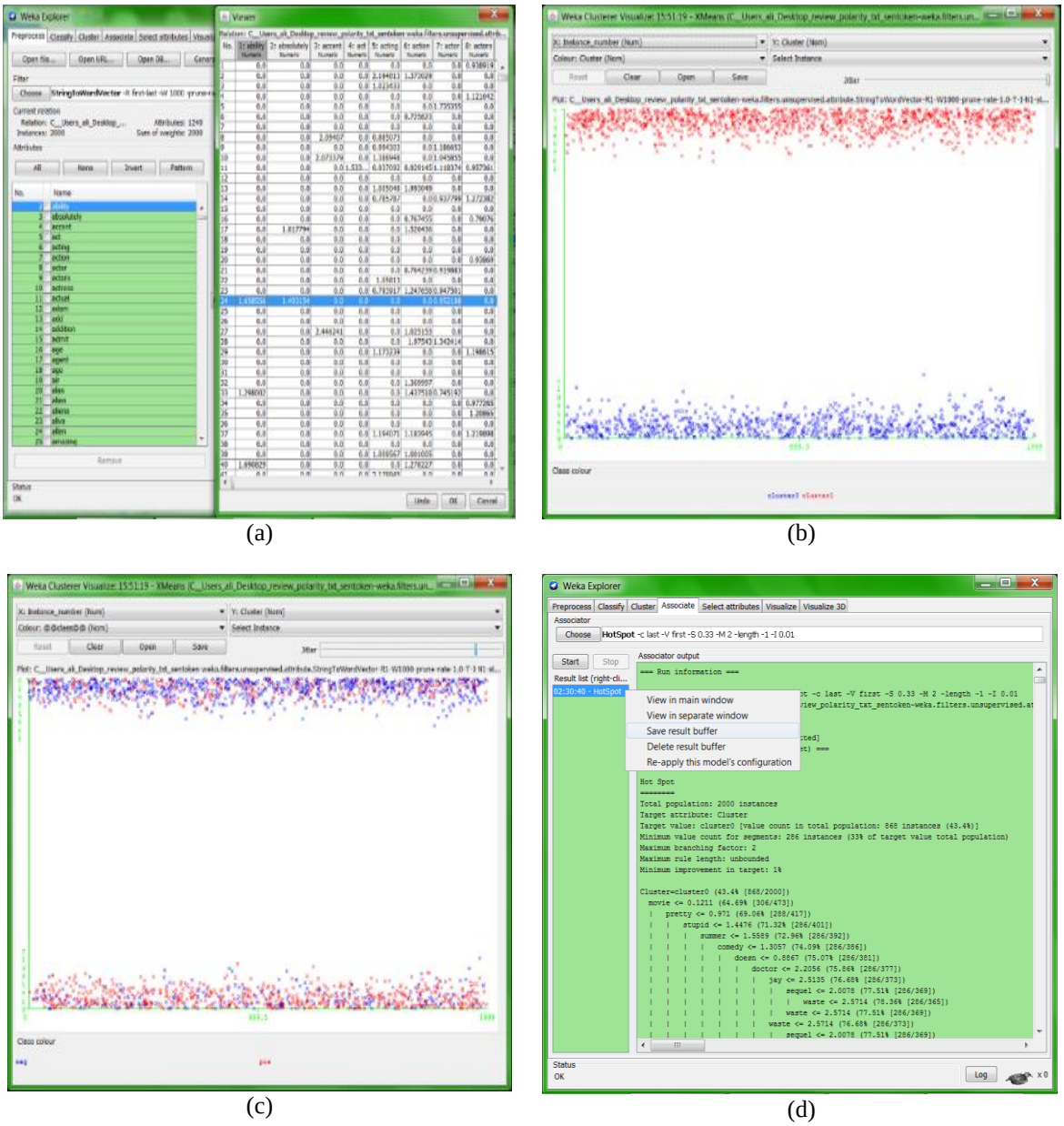

Figure 4. Implementation of framework functionalities, (a) matrix representation of terms and their respective frequency, (b) number of clusters captured by x-means algorithm application, (c) key information sequences within each cluster, (d) discovery result of maximum target values of interest

6. RESULTS AND DISCUSSION OF CLASSIFICATION ACCURACIES

The classification result accuracies are measured by way of classifying sentimental multi-view textual data as positive or negative information document categories. The classification results show the rate of correctly classifying the polarity data when both of simple-term based classification method and the proposed method are separately used for classifying sentiment multi-view textual data of polarity reviews dataset into their respective categories. The calculated correctly classifying rates are provided in Table 1.

The correctly classification rates have been measured alongside with every single classifier for predicting the correct category of each polarity information document. Consequently, the highest rate of correctly classification has the capacity with more accuracy assured in the process of decision making. Different settings of parameters have been tested by using Weka 3.7.12 software, in which different test modes like cross-validation with 10 folds and using training dataset are tested. The classifiers in the proposed method using rules for maximum target value of interest are all result improved accuracy than the simple term-based classification method, which is using (term frequency matrix) only.

Table 1. Results of correctly classification rates by test mode

Classification Method	Correct Classification Rate: using cross-validation with 10 folds		Correct Classification Rate: using training dataset	
	The proposed classification framework	Simple term-based classification model	The proposed classification framework	Simple term-based classification model
Decision Trees (J48)	70.3 %	66.75 %	98.5 %	96.85 %
Naive Bayes	79.85 %	76.7 %	85.55 %	83.2 %

7. CONCLUSION

This paper has been discussed the integration of different data mining algorithms for sentiment classification of multi-view polarity textual data. A data mining framework is proposed for generating Single-key Information sequences of knowledge by X-means clustering and maximum target value of interest by discovering Multi-Key Frequent Information Rules in each cluster by using HotSpot association rules will perform classification with improved accuracies of the classifiers. The analysis results have shown improved accuracies of classifying the sentimental multi-view textual data into positive or negative information document categories.

REFERENCES

[1] Özerk Yavuz, Adem Karahoca, Dilek Karahoca, "A Data Mining Approach For Desire And Intention To Participate In Virtual Communities," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 3714-3719, Oct. 2019.

[2] T. Cruz, et al., "Exploring data analytics of data variety," in *World Conference on Inforation Systems and Technologies*, pp. 920-930, 2018.

[3] M. Sammour, et al., "DNS Tunneling: a Review on Features," *International Journal of Engineering and Technology*, vol. 7, no. 3.20, pp. 1-5, 2018.

[4] B. Mcwilliams and G. Montana, "Multi-view predictive partitioning in high dimensions," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 5, no. 4, pp. 1-31, 2012.

[5] T. Baltrusaitis, et al., "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2. pp. 423-443, 2019.

[6] G. Miner, et al., "Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications," *Academic Press*, 2012.

[7] J. L. Solka, "Text data mining: Theory and methods," *Statistics Surveys*, vol. 2, pp. 94-112, 2008.

[8] A. Amado, Paulo Cortez, Paulo R., Sergio M., "Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis," *European Research on Management and Business Economics*, vol. 24, no. 1, pp. 1-7, 2018.

[9] H. Hashimi, et al., "Selection criteria for text mining approaches," *Computers in Human Behavior*, vol. 51, pp. 729-733, 2015.

[10] M. Jakubik, "Becoming to know. Shifting the knowledge creation paradigm," *Journal of Knowledge Management*, vol. 15, no. 3, pp. 374-402, 2011.

[11] S. Sangam and S. Shinde, "Sentiment Classification Of Social Media Reviews Using An Ensemble Classifier," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 16, no. 1, pp. 355-363, Oct. 2019.

[12] C. G. Reddick, et al., "A Social Media Text Analytics Framework for Double-Loop Learning for Citizen-Centric Public Services: A Case Study of a Local Government Facebook Use," *Government Information Quarterly*, vol. 34, no. 1, pp. 110-125, 2017.

[13] O. Müller, et al., "Using Text Analytics To Derive Customer Service Management Benefits From Unstructured Data," *MIS Quarterly Executive*, vol. 15, no. 4, pp. 243-258, 2016.

[14] J. S. Sowmiya and S. Chandrakala, "Joint Sentiment/Topic Extraction From Text," *2014 IEEE International Conference on Advanced Communications Control and Computing Technologies*, pp. 611-615, 2014.

[15] A. S. Halibas, et al., "Application of text classification and clustering of Twitter data for business analytics," in *Proceedings of 2018 Majan International Conference (MIC 2018)*, pp. 1-7, 2018.

[16] R. Ahuja and W. Anand, "Sentiment classification of movie reviews using dual training and dual predition," in *2017 4th International Conference on Image Information Processing (ICIIP 2017)*, pp. 594-597, 2017.

[17] P. Zola, et al., "Social Media Cross-Source and Cross-Domain Sentiment Classification," in *International Journal of Information Technology and Decision Making*, vol. 18, no. 5, pp. 1469-1499, 2019.

[18] R. Wang, et al., "Convolutional Recurrent Neural Networks for Text Classification," in *Proceedings of the 2019 International Joint Conference on Neural Networks*, pp. 1-6, 2019.

[19] A. W. Haryanto, et al., "Influence of Word Normalization and Chi-Squared Feature Selection on Support Vector Machine (SVM) Text Classification," in *Proceedings of 2018 International Seminar on Application for Technology of Information and Communication*, pp. 229-233, 2018.

[20] M. Fikri and R. Sarno, "A Comparative Study of Sentiment Analysis using SVM and SentiWordNet," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 13, no. 3, pp. 902-909, 2019.

[21] B. Tang, et al., "Toward Optimal Feature Selection in Naive Bayes for Text Categorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2508-2521, Sep. 2016.
[22] Z. Tang, W. Xiao, Bin Lu, and Y. Zuo, "A Parallel Algorithm for Bayesian Text Classification Based on Noise Elimination and Dimension Reduction in Spark Computing Environment," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11513 LNCS, pp. 222-239, 2019.
[23] F. Zablith and I. H. Osman, "Review Modus: Text classification and sentiment prediction of unstructured reviews using a hybrid combination of machine learning and evaluation models," *Applied Mathematical Modelling*, vol. 71, pp. 569-583, Jul. 2019.
[24] D. Pelleg and A. W. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 727-734, 2000.
[25] Y. Cai, et al., "Discipline Hotspots Mining Based on Hierarchical Dirichlet Topic Clustering and Co-word Network," *Journal of Software*, vol. 11, no. 11, pp. 1089-1101, 2016.
[26] M. Hall, "Class HotSpot," *weka.sourceforge*, 2015. [Online], Available: http://weka.sourceforge.net/doc.packages/hotSpot/weka/associations/HotSpot.html.
[27] J. H. Friedman and N. I. Fisher, "Bump hunting in high-dimensional data," *Statistics and Computing*, vol. 9, pp. 123-143, 1999.
[28] M. A. Burhanuddin, R. Ismail, N. Izzaimah, Ali Abdul-Jabbar M., and N. Zainol., "Analysis of Mobile Service Providers Performance Using Naive Bayes Data Mining Technique," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 5153-5161, 2018.
[29] B. Pang and L. Lee, "A sentimental education," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 271-es, 2004.