Possibilistic Fuzzy C-means Topic Modelling for Twitter Sentiment Analysis

Mandhula Trupthi[1]*   **Suresh Pabboju[2]**   **Gugulotu Narsimha[3]**

[1]Jawaharlal Nehru Technological University Hyderabad, India
[2]Chaitanya Bharathi Institute of Technology, India
[3]Jawaharlal Nehru Technological University College of Engineering Sultanpur, India
* Corresponding author's Email: trupthijan@gmail.com

Abstract: Social media are generating an enormous amount of sentiment data in the form of companies getting their customers' opinions on their products, political sentiment analysis and movie reviews, etc. In this scenario, twitter sentiment analysis is undertaken for classifying and identifying sentiments or opinions expressed by people in their tweets. Usually, the raw tweets consist of more noises in terms of URLs, stop-words, positive emojis and negative emojis, which are essentially reduced. After pre-processing, an effective topic modelling methodology Latent Dirichlet Allocation (LDA) is implemented for extracting the keywords and identifying the concerned topics. The extracted key words are utilized for twitter sentiment analysis using Possibilistic fuzzy c-means (PFCM) approach. The proposed clustering method finds the optimal clustering heads from the sentimental contents of twitter-sanders-apple2 database. The acquired results are obtained in two forms such as positive and negative. Finally, the experimental outcome shows that the proposed approach improved accuracy in twitter sentiment analysis up to 3-3.5% compared to the existing methods: pattern based approach and ensemble method.

Keywords: Latent dirichlet allocation, Possibilistic fuzzy c-means, Topic modelling, Twitter sentiment analysis.

1. Introduction

Now-a-days, social websites like Twitter, Facebook, Tumbler, etc. plays an essential role in human's life [1]. Among all social media, twitter is the most famous platform to communicate and share information with friends. People tweets about various topics on movies, products, brands, politics and many others [2, 3]. Also, twitter makes the information easy to spread and read, because it allows user to publish only 140 characters in a single tweet. Twitter is a micro blogging platform, which provides a vast amount of data that are utilized for several applications such as reviews, elections, sentiment analysis, marketing etc., [4, 5]. Sentiment analysis is the procedure of extracting the information from vast amounts of data and then classifies the data into dissimilar classes named as sentiments [6]. Sentiment analysis often named as opinion mining, where it mines the important features from the people opinions [7]. Sentiment

analysis is done by employing several machine learning methods, statistical approaches for extracting the features of twitter data [8]. Recent methodologies extract the twitter text from online blogs, for classifying the text as positive or negative [9]. Challenges faced by the researchers in twitter sentiment analysis are: neutral tweets are more common than positive and negative ones, which is very difficult to classify.

Tweets are very short and often show limited sentiment cues. Numerous researchers are focused on the use of traditional classifiers, like naive bayes, maximum entropy, and support vector machines to solve these problems [10]. In order to improve the classification accuracy, a new machine learning topic modelling methodology is implemented with an appropriate clustering approach [11]. In this experimental research, twitter sentiment analysis is performed on the reputed dataset: Twitter-sanders-apple2 dataset. The proposed methodology consists of three phases such as, pre-processing, topic modelling and classification. In the first phase, the

twitter data is pre-processed by reducing the unwanted noise in terms of URLs, stop-words, positive emojis and negative emojis. The pre-processed twitter data are used for topic modelling using LDA method, which helps to extract the relevant keywords and also for identifying the relevant topic. LDA not only analyse the content in a single tweet, but also handle more complex cases where multiple events mix together. These keywords are stored in the dictionary with an individual weight value, then the testing data is matched with the dictionary. The extracted weight value of testing data is utilized for twitter sentiment analysis using PFCM clustering approach. The PFCM methodology combines the benefit of both fuzzy c means and possibilistic c-means methods. The PFCM is a significant clustering algorithm to perform classification tests, because it possesses capabilities to give more importance to membership values. The proposed clustering methodology identifies the optimal clustering heads from the sentimental contents of twitter data. The acquired results are obtained in two forms such as positive and negative.

This paper is composed as follows. Section 2 presents a broad survey of several recent papers on twitter sentiment analysis. In section 3, a topic modelling approach named as LDA with a clustering methodology (PFCM) is presented. In Section 4, comparative experimental result of a proposed twitter sentiment analysis strategy for twitter-sanders-apple2 database is presented. The conclusion is made in Section 5.

2. Literature review

Several techniques are suggested by researchers in the twitter sentiment analysis. In this scenario, a brief evaluation of some important contributions to the existing literatures is presented.

R. Dehkharghani, H. Mercan, A. Javeed, and Y. Saygin, [12] proposed a new model named as sentimental causal rules for extracting the textual data sources from twitter data. The proposed methodology was the combination of both sentimental evaluation and causal rule discovery. The sentimental evaluation was the process of extracting the public sentiment from textual data. In addition, the causal rules refer to the relationship between dissimilar concepts in a context. The experimental research was performed on a publicly available dataset (i.e. Twitter dataset) to validate its classification accuracy in terms of polarity percentage. Computational time was a bit high in causal rules compared to other existing approaches,

while clustering the sentimental contents in large dataset.

A.C. Pandey, D.S. Rajpoot, and M. Saraswat, [13] proposed a new metaheuristic methodology, which was based on K-means and cuckoo search. The proposed methodology was used to identify an effective clustered heads from the twitter sentimental contents. Also, the proposed methodology generalized the practical implications for designing a system, which provide conclusive reviews on any social issues. Extensive experiments were conducted and the efficiency of the proposed approach was verified using Test-data.manual.2009.06.14, Twitter sanders apple 2 and 3 dataset. In a few cases, the proposed methodology is not efficient for large scale database and also not robust in content interpretation from multimedia documents.

G. Yang, D. Wen, N.S. Chen, and E. Sutinen, [14] evaluated a new methodology named as hierarchical Bayesian topic methods that incorporates the concepts of n-grams and hierarchical latent topics. The quantitative and qualitative evaluation results showed that the proposed model has outperformed the existing models in document modelling. In addition, the experimental results demonstrated that the proposed summarization system significantly improved the summarization performance. The proposed N-gram features have limited sets and unable to apply directly in different languages.

N. El-Fishawy, A. Hamouda, G.M. Attiya, and M. Atef, [15] illustrated a new machine learning based methodology, which was the combination of model tree approach and correlation feature selection. The main objective of this work was to reduce the summary of Arabic tweets associated with a particular topic. The proposed methodology was assessed and the results were compared with multi-document summarization systems. The experimental outcome demonstrated that the proposed approach has outperformed the existing methodologies in multi-document summarization. The major drawback of proposed approach was language dependent, for a particular language only the proposed approach will summarize the data, not for all the languages.

R. Abbasi-ghalehtaki, H. Khotanlou, and M. Esmaeilpour, [16] proposed a new methodology for topic modelling, which was the combination of evolutionary approaches, fuzzy logic systems and Cellular Learning Automata (CLA). At first, the most important features of the text, length and position of the sentence and similarity measures
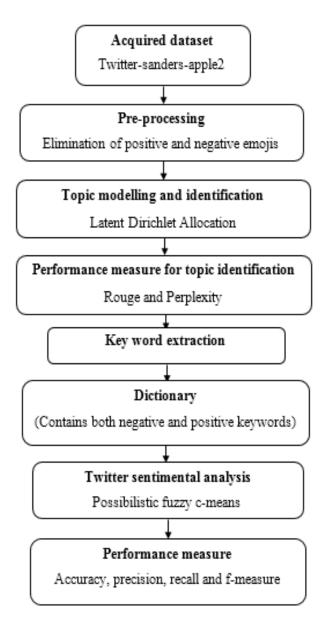
Figure.1 Work procedure of proposed approach

Acquired dataset
Twitter-sanders-apple2

Pre-processing
Elimination of positive and negative emojis

Topic modelling and identification
Latent Dirichlet Allocation

Performance measure for topic identification
Rouge and Perplexity

Key word extraction

Dictionary
(Contains both negative and positive keywords)

Twitter sentimental analysis
Possibilistic fuzzy c-means

Performance measure
Accuracy, precision, recall and f-measure

3. Proposed methodology

The proposed twitter sentiment analysis consists of three steps such as, data acquisition and pre-processing, topic modelling and classification. A general block diagram of twitter sentiment analysis is represented in the Fig. 1. The brief description about the proposed methodology is presented below.

3.1 Pre-processing and data acquisition

In the initial stage of the twitter sentiment analysis, the twitter data are taken from the standard benchmark dataset: twitter-sanders-apple 2 database. It consists of 479 tweets in a CSV format that is directly imported in the Light-SIDE. Twitter-sanders-apple 2 database has two categories such as positive (163 tweets that represents the positive sentiment) and negative (316 tweets that represents the negative sentiment) tweets. After the acquisition of twitter data, an important step in the twitter sentiment analysis is pre-processing of acquired data. The raw tweets consist of more noise in terms of URLs, stop-words, positive emojis, and negative emojis that are reduced before topic modelling.

3.2 Topic modelling and topic identification

The respective pre-processed twitter data are used for topic modelling and topic identification. Topic modelling is the automatic process of retrieving the concerned topic described by an original text based upon the distance. In this scenario, a new machine learning model (LDA) is utilized for automatic topic modelling and topic identification.

3.2.1. Latent Dirichlet allocation

The LDA is a general probabilistic topic model, where every document is represented as a random mixture of latent topics. Every latent topic is described as a distribution over fixed set of words in LDA and used to identify the underlying latent topic structure based on the observed data. For each document, the words are generated in a two phase process. In the first stage, a distribution over topics is randomly chosen for each word of document. In LDA, a word is a distinct data from a vocabulary index$\{1, .., V\}$, a series of N words are represented as $w = (w_1, w_2 \ldots \ldots w_n)$ and a collection of M documents are denoted as $D = (w_1, w_2 \ldots \ldots w_M)$.

The LDA approach are described as a three level Bayesian graphical model, where nodes are presented as random variables and the edges are stands for possible dependencies between the

were extracted. The proposed methodology identifies the most important and least important text features and then assigns fair weight to the text. The experiment was carried on a publicly available database (i.e., DUC2002 dataset) to validate its robustness and speed in terms of rouge. While implementing more approaches, summarization of the word was so complex. Also, the proposed method needs more repeated words for summarizing the particular word.

To overcome the above-mentioned drawbacks, a topic modelling method (LDA) with an appropriate clustering methodology (PFCM) is implemented that enhances the performance of twitter sentiment analysis.

variables. The LDA is observed from the three layered representation, π and μ parameters are examined during the generation of corpus. For every document, the document-level topic variables are examined and for each word of the document in LDA, the word level variables are examined.

A joint distribution over random variable is represented as the generative process of LDA. The Eq. (1) calculates the probability density function of $k-dimensional$ dirichlet random variable. An Eq. (2) compute the joint distribution of a topic mixture and the probability of a corpus is determined by the Eq. (3).

$$p(\aleph|\pi) = \frac{\Gamma(\sum_{i=1}^{k}\pi_i)}{\prod_{i=1}^{k}\Gamma(\pi_i)}\aleph_1^{\pi_1-1}\dots\dots\aleph_k^{\pi_k-1} \tag{1}$$

$$p(\aleph,x,y|\pi,\mu) = p(\aleph|\pi)\prod_{n=1}^{N}p(x_n|\aleph)\,p(y_n|x_n,\beta) \tag{2}$$

$$p(D|\pi,\mu) = \prod_{d=1}^{M}\int p(\aleph_d|\pi)\times \left(\prod_{n=1}^{N_d}\sum_{x_{dn}}p(x_{dn}|\aleph_d)p(y_{dn}|x_{dn},\mu)\right)d\aleph_d \tag{3}$$

Where, π is represented as the dirichlet parameter, $\aleph$ refers to document-level topic variables, x refers to per-word topic assignment, y refers to the observed word, μ refers to the topics, N is represented as number of words and M is denoted as the document.

For an individual document, the calculation of the posterior distribution of the hidden variable is an important inferential task in LDA. The exact inference of the posterior distribution of the hidden variable is an intractable problem. The combination of LDA with approximation algorithms such as, Laplace approximation, variational approximation, Gibbs sampling, and Markov chain are widely utilized. The negative and positive key words are extracted with an individual weight values and these key words are stored in dictionary. In order to attain negative and positive weight value, the testing twitter data are coordinated with the dictionary in the testing phase. The clustering process is carried-out by using PFCM after obtaining the negative and positive weight value.

3.3 Possibilistic fuzzy c-means

The PFCM is a data clustering technique, where each data point is a part of the cluster to a level indicated by the membership grade. In the clustering module, PFCM is dependent on the reduction of the objective function, which is illustrated in the following Eq. (4).

$$J_{PFCM}(U,T,V) = \sum_{i=1}^{C}\sum_{j=1}^{n}(u_{ij}^{m} + t^{n})d^{2}(x_j,v_i) \tag{4}$$

With the following constraints:

$$\sum_{i=1}^{c}\mu_{ij} = 1, \forall j \in \{1,\dots,n\} \tag{5}$$

$$\sum_{j=1}^{n}t_{ij} = 1, \forall i \in \{1,\dots,c\} \tag{6}$$

Where, J_{PFCM} is the objective function, U is represented as the partition matrix, T is represented as the typicality matrix, V is denoted as the vector of cluster centres. The outcome of the objective function is achieved using an iterative approach, where the degree of membership and the cluster centres are mathematically represented in the Eqs. (7), (8) and (9).

$$\mu_{ij} = [\sum_{k=1}^{c}(\frac{dx_j,v_i}{dx_j,v_k})^{\frac{2}{m-1}}]^{-1}, 1 \le i \le c, 1 \le j \le n \tag{7}$$

$$t_{ij} = [\sum_{k=1}^{n}(\frac{dx_j,v_i}{dx_j,v_k})^{\frac{2}{n-1}}]^{-1}, 1 \le i \le c, 1 \le j \le n \tag{8}$$

$$v_i = \frac{\sum_{k=1}^{n}(u_{ik}^{m}+t_{ik}^{n})x_k}{\sum_{k=1}^{n}(u_{ik}^{m}+t_{ik}^{n})}, 1 \le i \le c \tag{9}$$

Where, n is represented as the number of data points, c is represented as the number of cluster centers, which are described by the co-ordinates (x_j,v_i) and it is used to calculate the distance between cluster center and data sets.

PFCM constructs possibilities and memberships with normal prototypes and cluster centers for every cluster. Choosing the objective function is the important aspect for the performance of cluster methodology for accomplishing better clustering. Whereas, the clustering performance is based on the objective function, which is utilized for clustering. For developing an effective objective function, the following set of requirements are considered.

- Distance between the clusters should be reduced.
- Distance between the data points, which allocated in the clusters should be reduced.

The desirability between the clusters and data are modelled by the objective function. Further, the objective function of PFCM is improved by using driven prototype learning of parameter α. The

learning procedure α is a dependent exponential separation strength between the clusters and it is updated at every iteration. The parameter α is represented in the Eq. (10).

$$\alpha = \exp(-\min_{i \neq k} \frac{\|v_i - v_k\|^2}{\beta}) \qquad (10)$$

Where, β is represented as sample variance, mathematically, it is represented in the Eq. (11).

$$\beta = \frac{\sum_{j=1}^{n} \|x_j - \bar{x}\|^2}{n} \qquad (11)$$

Where, $\bar{x} = \frac{\sum_{j=n}^{n} x_j}{n}$

Then, a weight parameter is introduced to calculate the common value of α. Each point of the database consists of a weight in relationship with each cluster. So, the usage of weight function delivers a better classification outcome, especially in the case of noise data. The general equation of weight function is determined in the Eq. (12).

$$w_{ji} = \exp(-\frac{\|x_j - v_i\|^2}{(\sum_{j=1}^{n} \|x_j - \bar{v}\|^2) \times c/n}) \qquad (12)$$

Where, w_{ji} is denoted as the weight function of the point j with the class i. The step by step procedure of PFCM is represented in the Fig. 2, and also effectively explained below,
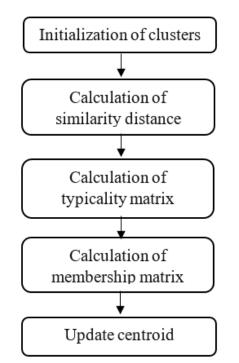

Figure.2 Step by step procedure of PFCM

- **Initialization:** At first, the number of clusters is furnished by the user, which are identical in respect of every segment.
- **Calculation of similarity distance:** When the number of clusters is determined, the evaluation of the distance between the centroids and data points for each segment is carried out.
- **Calculation of typicality matrix:** After the estimation of distance matrix, the typicality matrix is evaluated, which is obtained from the PFCM.
- **Calculation of membership matrix:** The evaluation of the membership matrix M_{ik} is performed by means of assessing the membership value of data point, which is gathered from the PFCM.
- **Update centroid:** After the generation of clusters, the modernization of the centroids is updated.

The relative procedure is performed again and again till the modernized centroids of each and every cluster becomes identical in successive iterations. Finally, the acquired results are obtained in two forms such as positive and negative.

4. Experimental outcome

For experimental simulation, Net-Beans (version 6.2) was employed on PC with 3.2 GHz with i5 processor. In order to estimate the efficiency of the proposed algorithm, the performance of the proposed method was compared with the ensemble classifiers [17] and Pattern-based approach [6] on the reputed datasets: Twitter-sanders-apple 2 dataset. The performance of the proposed methodology was compared in terms of rouge, perplexity, accuracy, precision, recall and f-measure.

4.1 Performance measure

The relationship between the input and output variables of a system understand by employing the suitable performance metrics like precision and recall. The general formula for calculating the precision and recall of the twitter sentiment analysis is given in the Eqs. (13) and (14).

$$Precision = \frac{TP}{TP+FP} \times 100 \qquad (13)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \qquad (14)$$

Accuracy is the measure of statistical variability and a description of random errors. The general formula of accuracy for determining twitter sentiment analysis is given in the Eq. (15).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (15)$$

Where, TP is represented as true positive, FP is denoted as a false negative, TN is represented as true negative and FN is stated as a false negative.

F-measure is the measure of accuracy test and it considers the both precision P and recall R of the test in order to calculate the score. The general formula for F-measure is given in the Eq. (16).

$$F - measure = 2. \frac{Precision \; . \; Recall}{Precision+Recall} \qquad (16)$$

4.2 Result for topic modelling

In this scenario, the result of topic modelling (LDA) was measured in terms of rouge and perplexity. Rouge and perplexity is the performance measure utilized for assessing automatic summarization in natural language processing. The general formula for perplexity and rouge is represented in the Eqs. (17), (18) and (19).

4.2.1. Perplexity

$$l(w) = logp(w|\phi,\alpha) = \sum_d logp(w_d|\phi,\alpha) \qquad (17)$$

$$Perplexity(testsetw) = exp\{-\frac{l(w)}{countoftokens}\} \qquad (18)$$

Where, w_d is represented as the unseen document, ϕ is stated as the topics of unseen document, α is denoted as the topic distribution of documents and $l(w)$ is represented as the log likelihood function.

4.2.2. Rouge

$$Rouge - 1 = \frac{\sum_{s \in summ_{ref}} \sum_{1-gram \in s} count_{match}(1-gram)}{\sum_{s \in summ_{ref}} \sum_{1-gram \in s} cunt(1-gram)} \qquad (19)$$

Where, $summ_{ref}$ is represented as the number of words trained to dictionary and $count_{match}$ is the represented as the number of matched words. The performance evaluation of LDA is stated in the Figs. 3, 4 and Table 1.

Table 1. Performance evaluation of LDA approach

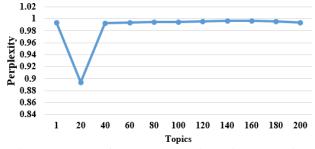| Topic	Perplexity	Rouge
1	0.994	0.111
20	0.894	0.334
40	0.993	0.033
60	0.994	0.076
80	0.9952	0.076
100	0.9952	0.105
120	0.996	0.067
140	0.997	0.067
160	0.997	0.45
180	0.996	0.87
200	0.994	0.071

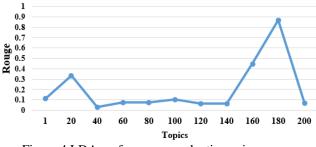Figure.3 LDA performance evaluation using perplexity

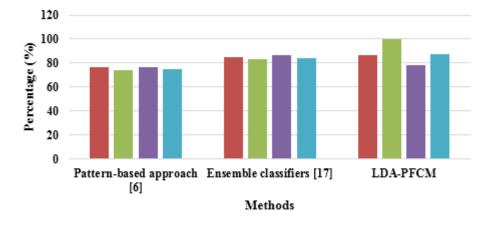Figure.4 LDA performance evaluation using rouge

4.3 Result for twitter-sanders-apple 2 dataset

In this experimental analysis, twitter-sanders-apple 2 dataset is assessed for comparing the performance evaluation of existing method and the proposed scheme. In Table 2, the accuracy, precision, recall and f-measure value of proposed and existing methodologies are compared for both the positive and negative classes. The accuracy, precision, recall and f-measure value of existing methods in the positive class delivers 76.90% and 84.89%, 73.70% and 82.90%, 76.90% and 86.30%, 75.30% and 84.20%. In addition, the proposed approach delivers 87%, 100%, 78.12%, and 87.71%.

Similarly, the accuracy, precision, recall and f-measure value of existing methods in negative class delivers 74.30% and 84.89%, 72.40% and 87.50%, 74.30% and 85.50%, 73.30% and 84.85%. Whereas, the proposed approach delivers 87%, 72.5%, 100% and 83.88%.

Table 2. Performance comparison of existing and proposed method

Performance measure	Positive class			Negative class		
	Pattern-based approach [6]	Ensemble classifiers [17]	LDA-PFCM	Pattern-based approach [6]	Ensemble classifiers [17]	LDA-PFCM
Accuracy (%)	76.90	84.89	87	74.30	84.89	87
Precision (%)	73.70	82.90	100	72.40	87.50	72.5
Recall (%)	76.90	86.30	78.12	74.30	85.50	100
F-measure (%)	75.30	84.20	87.71	73.30	84.85	83.88

Figure.5 Performance evaluation of existing and proposed method for positive class

Figure.6 Performance evaluation of existing and proposed method for negative class

Table 2 shows the performance evaluation of existing methodologies and the proposed method. In this scenario, topic modelling is performed to identify the relevant topic, later classification process is done to predict the best results, where in the existing methods the result is directly predicted without use of th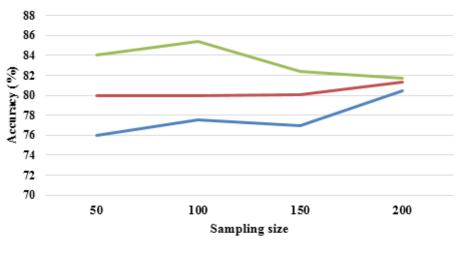e topic modelling. The topic-based mixture model for twitter sentiment helps to further improve the sentiment classification accuracy. Evaluation metrics confirms that the proposed scheme performs significantly in twitter sentiment analysis compared to existing methodology in terms of accuracy, precision, recall and f-measure.

Figure.7 Accuracy comparison of existing and proposed methodology

The graphical representation of performance evaluation of existing and proposed method (positive and negative class) is represented in the Figs. 5 and 6. In Figs. 5 and 6, the precision value in positive class and recall value in negative class attain low result, because of the highly dense spread of topics in the dataset.

Table 2 and Fig. 7 clearly shows that the proposed approach improves the classification accuracy in twitter sentiment analysis up to 2% compared to the existing methods in the twitter-sanders-apple 2 dataset. M. Bouazizi, and T. Ohtsuki, [6] proposed a pattern based approach for both binary and ternary classification in twitter sentiment analysis. Additionally, N.F. Da Silva, et al. [17] proposed an approach (ensemble classifiers) for automatically classifying the sentiment of tweets. In this literature, the tweets were classified as either positive or negative based on user query term. Compared to the existing scheme, the proposed method works effectively in terms of precision, recall, accuracy and f-measure.

5. Conclusion

Twitter sentiment analysis is one of the emerging fields for identifying and analysing the sentiments and viewpoints of users. In this experimental research, LDA and PFCM methods are under-taken for topic modelling and classification. LDA methodology extracts the key words and identifies the topic related to the key words. The extracted key-words are utilized for twitter sentiment analysis using PFCM clustering methodology. This clustering methodology identifies the optimal cluster heads from the sentiment contents. This experimental investigation is verified on a publicly available database named as

twitter-sanders-apple2 dataset, which shows the superiority of the proposed approach. The classification rate in the twitter-sanders-apple2 data is better in the proposed methodology than the previous methodologies. Associated to the other obtainable approaches for twitter sentiment analysis, the proposed scheme delivered an effective performance by means of accuracy, around 3-3.5% enhancement than the previous methods.

In the future work, for further improving of classification rate, a new multi-objective classification approach is developed with an adaptive topic modelling methodology.

References

[1] N. Öztürk and S. Ayvaz, "Sentiment Analysis on Twitter: A Text Mining Approach to the Syrian Refugee Crisis", *Telematics and Informatics*, Vol.35, No.1, pp.136-147, 2017.

[2] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter", *Information Processing & Management*, Vol.52, No.1, pp.5-19, 2016.

[3] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "CNN for situations understanding based on sentiment analysis of twitter data", *Procedia Computer Science*, Vol.111, pp.376-381, 2017.

[4] K. Philander and Y. Zhong, "Twitter sentiment analysis: Capturing sentiment from integrated resort tweets", *International Journal of Hospitality Management*, Vol.55, pp.16-24. 2016.

[5] T. Singh and M. Kumari, "Role of Text Pre-processing in Twitter Sentiment Analysis", *Procedia Computer Science*, Vol.89, pp.549-554, 2016.

[6] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", *IEEE Access*, Vol.5, pp.20617-20639, 2017.

[7] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis", *IEEE Access*, Vol.5, pp.2870-2879, 2017.

[8] P. Ray and A. Chakrabarti, "Twitter sentiment analysis for product review using lexicon method", In: *Proc. of International Conf. on Data Management, Analytics and Innovation (ICDMAI)*, pp.211-216, 2017.

[9] S. Anastasia and I. Budi, "Twitter sentiment analysis of online transportation service providers", In: *Proc. of International Conf. on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 359-365, 2016.

[10] J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, "Election result prediction using Twitter sentiment analysis", In: *Proc. of International Conf. on Inventive Computation Technologies (ICICT)*, Vol.1, pp.1-5, 2016.

[11] M. Trupthi, S. Pabboju, and G. Narasimha, "Improved feature extraction and classification Sentiment analysis", In: *Proc. of International Conf. on Advances in Human Machine Interaction (HMI)*, pp.1-6, 2016.

[12] R. Dehkharghani, H. Mercan, A. Javeed, and Y. Saygin, "Sentimental causal rule discovery from Twitter", *Expert Systems with Applications*, Vol.41, No.10, pp.4950-4958, 2014.

[13] A.C. Pandey, D.S. Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method", *Information Processing & Management*, Vol.53, No.4, pp.764-779, 2017.

[14] G. Yang, D. Wen, N.S. Chen, and E. Sutinen, "A novel contextual topic model for multi-document summarization", *Expert Systems with Applications*, Vol.42, No.3, pp.1340-1352, 2015.

[15] N. El-Fishawy, A. Hamouda, G.M. Attiya, and M. Atef, "Arabic summarization in twitter social network", *International Journal of Ain Shams Engineering*, Vol.5, No.2, pp.411-420, 2014.

[16] R.Abbasi-ghalehtaki, H. Khotanlou, and M. Esmaeilpour, "Fuzzy evolutionary cellular learning automata model for topic modelling", *Swarm and Evolutionary Computation*, Vol.30, pp.11-26, 2016.

[17] N.F. Da Silva, E.R. Hruschka, and E.R. Hruschka, "Tweet sentiment analysis with classifier ensembles", *Decision Support Systems*, Vol.66, pp.170-179, 2014.