IJCTA, 10(20), 2017, pp. 125-138
ISSN: 0974-5572

A Hybrid Kernel Based Extreme Learning Machine for Effective Sentiment Analysis

T.M. Saravanan* and A. Tamilarasi*

ABSTRACT

Sentiment analysis or Opinion mining is a popular and efficient scheme for web data analysis and information retrieval. The exponential growth of the user produced content has opened new horizons for research in the field of sentiment analysis. In this paper, a new model proposed for sentiment analysis of product reviews based on opinion mining and classification methods. In this work, Hybrid Kernel based Extreme Learning Machine (HKELM) algorithm proposed to categorize the sentiments of customer's in web applications. The initial process of proposed work is prepared the data from web documents of customer's review dataset. The data preparation is done by using pre-processing with Expectation Maximization (EM) clustering algorithm. In pre-processing, the unwanted data and outlier data is removed using clustering algorithm. Next, Multivariate Autoregressive Model (MAR) is used for potential feature extraction. After that, the extracted features are ranked and opinion scores are provided for every noun using Modified Cat Swarm Optimization (MCSO) algorithm. Finally, to extract the maximum used opinion words using HKELM algorithm, which classify the review into three kinds positive (good), negative (bad) and neutral based on the opinions. In this proposed Hybrid process, the Particle Swarm Optimization (PSO) scheme used for optimizing the kernel functions parameters in KELM. The experimental results show that proposed model is better than ID3 and Bagging classifier models in terms of the performance of precision, accuracy, recall, F-Measure and processing time.

Keywords: opinion mining, sentiment analysis, hybrid kernel based extreme learning machine, expectation maximization, cat swarm optimization.

1. INTRODUCTION

An Opinion is a decision about a specific thing formed by a majority of people, and is not essentially based on knowledge or fact. Generally, opinion is mentions to what a person considers about something or a subjective belief opinion, and the consequence of facts or emotion analysis [1]. Opinion Mining (OM) is also known as Sentiment Analysis (SA), is a Natural Language Processing (NLP) type to find public thought about a topic or product. In web a new product or item is given means, the OM or SA tool progression a collection of search results and then providing item attributes like features, cost, quality, etc., and gathering their opinion. This research work is focused the problem of OM accurate classification from others opinions on a specific item. The process entails gathering and examining opinions about item in social websites like twitter etc., blog posts, statements and reviews.

In marketing area, SA or OM is very helpful for economic and advertising strategizing where victory of a new item launch can be evaluated, make a decision about which item is famous and also identifies demographics like specific features. This sentiment analysis has challenges to face that opinion sentence, because in some situation, it defined positive opinion and in another situation, it defined negative opinion. So the opinions are not stated evenly by various persons. But, most of the sentence only mentioned in positive or negative and are examined sentence by sentence. However, in more social media like twitter,

* Department of Computer Applications, Kongu Engineering College, Perundurai, Erode, Tamil Nadu 638052, India.

people given their opinion which contain more various combined opinions in same sentence. It couldn't identify which may (positive) or may not (negative) and this analysing is very difficult for using an algorithm [2].

With the growth of social media like Twitter, micro-blogs on the Web and their reviews, forum discussions, comments, blogs and postings are using opinion in such media to make decisions by individuals and organizations. In general, overall background polarity or writer sentiment about some feature is established using SA. In SA, classification found major challenge and is sentiment may be judgment, mood or evaluation of a product or item which can be a sentence or document or feature that is categorized positive or negative [3]. However, due to the broadcast of various sites in web, they had difficult to monitoring and finding opinion sites and extracting information. In long blogs and forum locations, each and every site has some opinion sentence and it can easily decode. In websites, determine specific item related sites and their opinion extraction and analysing was critical through using an average human reader. In that situation, the automated SA or OM systems are needed [4]. Many researches [5-19] are present on SA of user opinion review and it mainly decides the polarities of user reviews. In this analysis are, the nature NLP is used in specific document sentiment analysis. Nowadays, sentiment analysis is the directive at the intersection of NLP and recovery of information and also shares the numeral of features along with other processes like computational linguistics, information extraction, predicative analysis, text-mining and psychology.

In this paper, HKELM algorithm proposed to categorize the sentiments of customer's in web applications. Initially, the data is pre-processed using EM clustering algorithm. Secondly, MAR is used for feature extraction. Then, MCSO is used to rank the noun. To extract the maximum used opinion words using HKELM algorithm. The experimental results show that proposed model is better than earlier sentiment analysis methods.

2. RELATED WORK

In this section, the existing research of sentiment analysis has been discussed. Farhadloo and Rolland [5] proposed a new scheme to utilize bag of nouns rather than bog of words to improve the clustering results for aspect identification and score representation, a new feature set, which leads to more accurate sentiment identification. This scheme is based upon the three scores such as positive ness, neutral ness and negative ness that are learned from the data for each term. This scheme improves the performance of 3-class sentiment analysis on sentences by 20% in terms of f1-measure compared than existing methods.

Njolstd et al., [6] proposed five different machine learning methods and evaluated the four different feature categories, compiled of 26 article features for sentiment analysis. These schemes classify Norwegian financial internet news articles, and attain classification precisions up to ~71%.

Valakunde et al.., [7] presented a hierarchical SA based on feature level based sentiment. Specifically, the document level sentiment was analysed via sentiment score of feature level sentiment and respectively given weightage of the entities. Here, multiclass SA was present great attention detail of sentiments, rather than positive and negative information. Based on the performance evaluation, feature based document level SA has high accuracy when compared to direct document level SA and also it reduced misclassification of the documents. The obtained performance result was better than existing schemes of Support Vector Machine (SVM) and Naive Bayesian (NB). But the processing time of this scheme is less.

Liu et al., [8] researched A Fuzzy Domain Sentiment Ontology Tree (FDOT) can be automatically constructed to facilitate opinion mining, comprising the extraction of product features, sentiment words and extraction of feature-relation. It also predicts the polarities of sentiments. The results shows it can develop efficient business strategies related to marketing, support to the customer, and product design functions in a timely fashion.

Sumathi et al., [9] used an optimal feature selection for reducing feature subset size and computational complexity thus increasing the classification accuracy. For solving combinatorial optimization problems, the Artificial Bee Colony (ABC) algorithm is used and it is a powerful optimization technique. Hence, this method is incorporated for optimizing the feature subset selection. In this system, using opinion mining the movie reviews is classified. This system evaluated feature selection techniques based on Inverse Document Frequency (IDF) and ABC scheme. The results show that the classification accuracy of the proposed ABC has attained high in tune of 1.63% to 3.81% for feature selection.

Basari et al., [10] has shown that PSO increases the accuracy of SVM after the hybridization of SVM-PSO. In this study, after data cleansing Support Vector Machine (SVM) with Particle Swarm Optimization (SVM-PSO) achieved the best accuracy of 77%. Also, the accuracy level of SVM-PSO can be improved by using enhancements of SVM that might be using another combination of SVM with other optimization method.

Manjul et al., [11] proposed Probabilistic aspect algorithm using K-Nearest Neighbour (K-NN) based classifier for sentiment analysis. In this proposed scheme, semantic-oriented subjective information extraction and calculated raking based on the aspect frequency.

Vigneshkumar et al., [12] learned so many classification techniques such as K-NN, SVM and Principal Component Analysis (PCA), Artificial Neural Network (ANN) and Fuzzy logic. In this system, to improve the prediction accuracy, to enhance the quality factor considers Autoregressive Sentiment and Quality Aware (ARSQA), and is to build the quality for predicting sales performance.

Basiri et al., [13] proposed a new score aggregation method by using the Dempster-Shafer theory of evidence. Initially, the polarity of reviews is detected using a machine learning approach. Then, consider sentence scores as evidence for the overall review rating. The two public social web datasets evaluation results show the higher performance of that method in comparison with existing score aggregation methods and machine learning approaches.

Tripathi et al., [14] presented a survey on sentiment analysis and the related methods. It also discusses the application areas and challenges for sentiment analysis with insight into the past researches. Joshi et al, [15] provides a study of the most of the researchers reported that SVM has high accuracy than other algorithms. The NB classifier produces the best results compared to SVM. But the accuracy and processing time of those classifiers is less.

3. PROPOSED METHODOLOGY

In this section, the proposed HKELM based automatic sentiment analysis has been described. Each and every process has been discussed in given below subsections.

3.1. Pre-processing using EM algorithm

In this pre-process, stop words and stemming words are removed from data set. And the outlier is detected based on EM clustering. This clustering process is grouped dissimilar and similar data. Finally, the dissimilar data are removed from dataset and it increases the accuracy of proposed performance.

3.1.1. Stop Words Removal

After processing of natural language data, the Stop words are removed. These stop words (the, is, at, which, on etc) can cause problem while searching for phrases that include them.

3.1.2. Stemming

Customer online reviews are usually mentioned with informal language and they include internet slang and contemporary spellings such as use of apostrophes, ing form of words to name a few. So such words must be re-visited and stemmed for correct data retrieval. Porter Stemming algorithm is used remove the stemming words.

3.1.3. Outlier detection

It is a task that finds objects that are dissimilar with respect to the remaining data. The data prepared for using EM clustering algorithm. It is used in maximum likelihood estimation where the problem involves two data sets of random variables of which observable one X, and hidden one Z. In simpler words algorithm works in following two steps:

E-step: the expectation of the missing value (unlabeled class) information is estimated. It is used to perform classify each unlabeled document. Using current parameter the probability distribution is calculated.

M-step: the likelihood of the model parameter maximizes using the previously computed expectation of the missing values as if were the true ones.

Algorithm 1: Pre-processing of customers review data

Inputs: Collections of customer review data set, no of attributes (words), missing values or unobserved hidden data z and vector of unknown parameter α

Output: predict the cleaned data

maximum probability calculation method:

Initial naive Bayes classifier, α, from the input datasets, maximum a posteriori parameter estimation used to find α' argmax $P(X \mid \alpha)\, P(\alpha)$

- while loop parameters of classifier improved, as measured by the change in L_c (α / D, z) c // α, parameter, D-labelled data, c-class
- **(E-step)** the current classifier a' used to estimate component membership of each unlabeled dataset, *i.e.*, the probability that each mixture component (and class-c and document d) generated each dataset, $P(c_j/d_i;\ \alpha')$.

(M-step) the classifier ⊖ˆ Re-estimated to given the estimated component membership of each dataset. Maximum a posteriori parameter estimation Used to find $\alpha' = \text{argmax } \alpha P(X|)P(\alpha)$

- **Preprocessing steps:**

1. Periods, commas, punctuation, stop words are removed. Data's that have occurrence frequency more than once in the data set is collected.
2. The frequent words viewed as data sets by matching the words which are in the vocabulary as well as training datasets.
3. Search for matching word s-sets or its subsets i.e. containing items more than one in the list of data sets collected from training data with that of s-subset i.e. containing items more than one of frequent data set of new dataset.
4. The corresponding probability values of matched data set(s) for each target class are collected.
5. The probability is calculated
6. Score algorithm applied to calculate the range in which the attributes must be lying.
7. The probability class calculated by applying expectation maximization algorithm.
8. The dataset Categorized in the class having maximum probability as cleaned dataset.

3.2. Feature Extraction using MAR

The MAR model extracts features from user's reviews for individual products, therefore feature vector values are represented as a linear summation of earlier activities. Let us consider, to be the time series review data with attributes and where m be the order of MARmodel for potential feature extraction, y_n is linear summation is represented as,

$$y_n = \sum_{i=1}^{m} y_{n-i} A(i) + e(n) \tag{1}$$

Where $y_n = [y_n(1), y_n(2), ..., y_n(d)] \rightarrow$ the n^{th} sample of a d-dimensional user's review related extracted features results at various stages and A(i) $\rightarrow$ $a\ d-\text{by}-d$ weight matrix values for each review data of each subjects. And $e_n = [e_n(1), e_n(2), ..., e_n(d)] \rightarrow$ preservative Gaussian noise for each data.

3.3. Cat Swarm Optimization (CSO)

CSO is a new optimization algorithm in the field of swarm intelligence [16]. The CSO algorithm models the behaviour of cats into two modes such as seeking mode and Tracing mode. To search in the solution space, initially, swarm is made of initial population composed of particles. In CSO, cats are used as particles for solving the rank problems in sentiment analysis. This ranking process is calculated based on the fitness value. The high fitness value of features is marked as high rank as well as calculated opinion score.

In CSO, every cat position is mentioned as D dimensions, each dimension has own velocities, and a fitness value represents the accommodation of the cat to the fitness function, and to identify whether the cat is in seeking mode or tracing mode mentioned as a flag. The best position of one of the cats is predicted as final solution. The CSO keeps the best solution until it reaches the end of the iterations [17].

3.3.1. Seeking Mode

The behaviour of cats in resting time and being-alert is modelled as seeking mode. It is used for thinking and deciding about next move in a time. In this mode, Seeking Range of the selected Dimension (SRD), Seeking Memory Pool (SMP), Counts of Dimension to Change (CDC) and Self-Position Consideration (SPC) [16] parameters are used. The seeking mode step by step process is given below.

Step 1: construct p = SMP replicas of the current position of C_k and C defines the cat. In case, the value of SPC is correct, let $p = (SMP - 1)$, and then one of the cat's current position is maintained.

Step 2: For each replica, SRD produced arbitrarily plus or minus rates and replace the old values along with CDC.

Step 3: the Fitness Values (FV) of all candidate points are calculated.

Step 4: If all FV are not exactly equal, then for each candidate point the selecting probability computed by using equ (2), or else set value of 1 for all the selecting probability of each candidate point.

Step 5: picked the point randomly from the candidate points to move, and change the position of C_k.

$$P_o = \frac{|FV_o - FV_x|}{FV_{\max} - FV_{\min}} \tag{2}$$

If the goal of the FV is to be find the minimum solution, then $FV_x = FV_{max}$, or else $FV_x = FV_{min}$.

3.3.2. Tracing Mode

The second mode of algorithm is tracing mode. In this mode, cats are desired to trace targets and foods. Tracing mode process is given below.

Step 1: according to (3) for every dimension the velocities are updated.

Step 2: the velocity value is checked, it is in the maximum velocity range or not. In case, the new velocity is exceed the range means it is set equal to the limit.

$$V_{k,d} = V_{k,d} + r_1 c_1 \left(X_{bset,d} - X_{k,d} \right) \tag{3}$$

Where $X_{best,d} \rightarrow$ the best fitness value of cat position, $X_{k,d} \rightarrow$ the position of C_k, $C_1 \rightarrow$ an acceleration coefficient (here, it equal to 2.05) and is used to extend the velocity of the cat, $r_1 \rightarrow$ a random value uniformly generated in the range of [0,1].

Step 3: according to (4) the position of C_k is updated

$$X_{k,d} = X_{k,d} + V_{k,d} \tag{4}$$

3.3.3. Core Description of CSO

The integration of seeking and tracing modes defined as a Mixture Ratio (MR). In this process, MR parameter is used to decide how many cats will be moved into seeking mode process. For instance, considered 100 population size and 0.7 values for MR parameter, then it calculated the cat by ($100 \times 0.7 = 70$). In 100 cats, 70 cats move to seeking mode and remaining 30 cats move to tracing mode in this iteration [18]. The CSO algorithm is described given below steps

1. Initially, N cats are created and the positions of cat, velocities and flags for cats are initialized.
2. Using objective function, the fitness value of the each cat is evaluated and the best cat is keep into memory and is mentioned as X_{best}.
3. Then, along with cat's flag, cat is applied to the seeking or tracing mode.
4. According to MR parameter, the number of cats is re-picked and set them into seeking mode or tracing mode
5. Finally, checked the termination condition, if condition satisfied, then program terminated, or else goes to fitness function [18].

3.4. Modified Cat Swarm Optimization

In CSO, the tracing mode has two equations like velocity update and position update equations. To achieve efficient adaptive CSO, in velocity equation the parameters are changed to compute the current position of cats based on the existing and next dimensions information's. A new dynamic position update equation is achieved via using a special factor.

3.4.1. Adaptive Parameters

In this proposed approach, to produce constancy among global and local search ability the adaptive parameters are used. In each dimension, an adaptive inertia weight combined to the velocity equation and is updated. The inertia weight with high value assists a global search and a small value assists a local search. First, the high values are processed and it will slowly decreased to the small value via using given below equation

$$w(o) = w_s + \left(\frac{o_{\max} - o}{2 \times o_{\max}} \right) \tag{5}$$

Where $w(o) \rightarrow$ defines the inertia weight update $w_s \rightarrow$ the initial weight, $o_{max} \rightarrow$ the maximum dimension of benchmark and $o \rightarrow$ the current dimension. In every dimension, the first dimension of the each iteration, the maximum inertia weight happens and it will be updated decreasingly. Generally c_1 is mentioned as 2.05 value, but here, an adaptive formula updated by using given formula (6) and this formula shows the gradual increment of adaptive acceleration coefficient in every dimension also the high value achieved in last dimension

$$c(o) = c_s - \frac{(o_{\max} - o)}{2 \times o_{\max}} \tag{6}$$

In this proposed scheme, c_s is equal to 2.05 and for each cat, the velocity update equation changed based on a new appearance defined in (7)

$$V_{k,d} = w(d) + V_{k,d} + r_1 \times c(d) \times (x_{best,d} - X_{k,d}) \tag{7}$$

3.5. HKELM based classification

The ELM learning algorithm has rapid learning speed and better generalization performance [19, 20]. In ELM, there is no need to tune the hidden layer initial parameters but nearly all nonlinear piecewise continuous tasks are mentioned as hidden neurons. So, N arbitrary distinct samples $\{(p_i, q_i)|p_i \in R^n, q_i \in R^m, i = 1, ..., N\}$ in ELM of the output function with L hidden neurons is defined as

$$f_L(p) = \sum_{i=1}^{L} \omega_i hl_i(p) = hl(p)\omega \tag{8}$$

Where $\omega = [\omega_1, \omega_1, ..., \omega_L] \rightarrow$ output weights vector among the L neurons and the output neuron of hidden layer, $hl(x) = [h_1(p), hl_2(p), ..., hl_L(p)] \rightarrow$ the output vector of data as of input space to the ELM feature space [20].

In this neural network, to decrease the training error and increase the generalization performance, the training error and the output weights are reduced at similar time and it defined as

$$minimize: \|H\omega - T\|, \|\omega\| \tag{9}$$

The least squares result of (9) based on Karush-Kuhn-Tucker(KKT) conditions can be defined as

$$\omega = H^T \left(\frac{1}{C} + HH^T \right)^{-1} T \tag{10}$$

If the feature mapping $hl(p)$ is unidentified, then based on Mercer's conditions the kernel matrix of ELM can be defined as

$$M = HH^T : m_{ij} = hl(p_i) hl(p_j) = k(p_i, p_j) \tag{11}$$

Thus, the output function $f(p)$ of the KELM can be described as

$$f(p) = [k(p, p_1), ..., k(p, p_N)] \left(\frac{1}{C} + M \right)^{-1} T \tag{12}$$

Where $M = HH^T$ and (p, q) is single hidden layer feed-forward neural networks of hidden neurons kernel function. In this proposed scheme, Wavelet kernel function used for simulation and performance analysis and the wavelet kernel functions defined as

$$fk(p,q) = \cos\left(d \frac{\|p-q\|}{e} \right) \exp\left(\frac{\|p-q\|^2}{f} \right) \tag{13}$$

To increase the performance of neural networks, the adjustable parameters such as d, e, and f (From eqn 13) are used and these are tuned under the solved difficulty.

Evaluated with the ELM learning algorithm, the hidden layer feature need not be chosen in the KELM. In addition, the KELM achieved better generalization performance and it has efficient stability evaluated to traditional ELM and it is faster than SVM [20].

3.6. PSO for KELM

In KELM, the C - regulation coefficient and kernel parameters are chosen for better generalization performance of neural networks. In [20], the optimization process of parameters is only focused in a wide range and it takes more time consumption. And in [21], proposed a hybrid kernel function for getting better generalization performance of KELM. But, the issue of how to select the optimal value for kernel function parameters has not been determined. In this paper, PSO optimization scheme is proposed to the KELM for choosing the optimal parameters of kernel function. Compared with other optimization schemes, the PSO is a biologically inspired computational stochastic optimization scheme presented by Eberhart and Kennedy [22]. It is becoming popular due to its little memory requiring, easy implementation, and capability to converge to get a sensibly optimal result quickly [23].

Considered to other swarm optimization schemes, the PSO algorithm also works by randomly initialling the population of individuals in the search space. In the search space, each particle of PSO can fly around to find the best solution, also the particles all look at the best particle (i.e. best solution) in their path.

Let the dimension of search space of PSO is D and the population size is $\wedge\hat{N}$. Then, x_i^d mentioned the current position of i^{th} particle and V_i^K mentioned the current velocity of i^{th} particle at iteration t. Then, the new position and velocity of the particles for the next iteration t are defined as

$$v_i^k(t+1) = w.v_i^k(t) + c_1.rand_0\left(p_i^k(t) - x_i^k(t)\right) + c_2.rand_0\left(g_i^k(t) - x_i^k(t)\right), \tag{14}$$

$$x_i^k(t+1) = x_i^k(t) + v_i^k(t+1), 1 \le i \le \hat{N}, 1 \le k \le D \tag{15}$$

where $p_i^k \rightarrow$ the best position of the ith particle, $g_i^k \rightarrow$ the global best position, which constitutes the best position found via the entire swarm, $w \rightarrow$ the inertia weight, c_1 and $c_2 \rightarrow$ the acceleration constants, and *rand*O $\rightarrow$ a random number between 0 and 1.

In PSO, the inertia weight w preserves the development ability of exploring new areas in the search space. So, to ensure higher exploring ability in the early iteration and quick convergence speed in the last part iteration, w parameter can reduce gradually at the generation increases and is calculated as [24]

$$w(t) = w_{\max} - iter \times \frac{\left(w_{max} - w_{min}\right)}{max\ iter} \tag{16}$$

Where w_{max} and $w_{min} \rightarrow$ the initial inertia weight and the final inertia weight and *max iter* $\rightarrow$ the maximum searching iteration number and *iter* $\rightarrow$ the iteration number at the present.

Additionally, in the early part iteration, to enhance the global search, to support the particles to converge to the optimal solution of global search, and to increase the convergence speed in the final iteration period [23], the acceleration parameters c_1 and c_2 are described as

$$c_1 = \left(c_{1_{min}} - c_{1_{max}}\right)\frac{iter}{max\ iter} + c_{1_{max}} \tag{17}$$

$$c_2 = \left(c_{1_{max}} - c_{2_{min}}\right)\frac{iter}{max\ iter} + c_{2_{min}} \tag{18}$$

Where $c_{1_{max}}$ and $c_{1_{min}}$ are the initial and final acceleration constant of c_1, and $c_{2_{min}}$ and $c_{2_{max}}$ are the initial and the final acceleration constant of c_2. Consequently, the acceleration coefficients changing with time and reduced the cognitive component and the social component is increased in (14), respectively.

Based on PSO with self adaptive parameters w and c, in KELM, the parameters of kernel functions are optimized for improving the generalization performance. While the number of parameters of kernel functions is changed. In this paper, the dimension of the particle of the proposed algorithm also changes with the different kernel functions. So, the particle or individual θ of search space in the proposed algorithm can be defined as

θ □∈ $[a, b, c]$ for wavelet kernel

Hence, the kernel parameter optimization strategy of KELM based on the PSO also called as Adaptive Kernel Based Extreme Learning Machine Learning (AKELM) is described given below.

Considered the kernel function type, the training set $\{(p_i, q_i)|p_i \in R^n, q_i \in R^m, i = 1, ..., N\}$ and the initial value of regulation coefficient C, process of HKELM is given below.

Step 1: The population or particles are initiated the position and velocity of each particle based on the kernel function.

Step 2: The fitness function of each particle is evaluated based on the following conditions: the root means standard error for regression problems and classification problems of classification accuracy.

Step 3: Based on the formulas (14)–(18), modified the velocity and position of the particle.

Step 4: until the maximal iteration time is satisfied, the Step 2, 3 are iterated repetitively.

Step 5: determined the optimal parameters of kernel function. After that, the hidden layer kernel matrix is computed based on the optimized parameters.

Step 6: the final output weights ω determined in terms of the following equation

$$\omega = H^T\left(\left(\frac{1}{C}\right) + HH^T\right)^{-1} T \tag{19}$$

Thus the reviews are classified into Positive, Negative and Neutral based on the opinions of the users that are included in the reviews.

4. RESULTS AND DISCUSSION

In this section, the proposed HKELM sentiment analysis scheme performance is evaluated. And the evaluation of performance results of proposed scheme is compared with existing sentiment analysis schemes such as Interactive Dichotomizer version 3 (ID3) [25] and J48 [26].

4.1. Dataset description

The polarity dataset used for performance evaluation and it downloads from the URL http://www.cs.cornell.edu/people/pabo/movie-review-data. The data set folder has "txt_sentoken" are the 2000 processed down-cased text files used in Pang/Lee ACL 2004 and it two subdirectories such as "pos" and "neg", specify the true classification (sentiment) of the component files according to proposed automatic rating classifier and the performance evaluation is given below.

4.2. Evaluation Criteria

In this study, the sensitivity, and specificity [27] are used to quantify the accuracy of the sentiment analysis models. TP -True Positives, TN -True Negatives, FP - False Positives and FN -False Negatives are mentioned in the confusion matrix, and the sensitivity is calculated by using this formula (TP/(TP + FN). Sensitivity mentioned the proportion of positive cases (good or bad or neutral) which are predicted to be positive or correct.

The specificity is calculated by using this formula (TN/((TN + FP)). The specificity mentioned the proportion of negative cases which are predicted to be negative [27]. ROC analysis used to calculate the accuracy of a classifier. The horizontal axis and the vertical axis of an ROC curve are defined by Equations 20 and 21 respectively [27].

$$x = 1 - \text{specificity } (t) \tag{20}$$

$$y = \text{sensitivity } (t) \tag{21}$$

To measure the accuracy of a model, the AUC can be measured [27, 28].

Fig 1 shows the performance of the separation, with the cutoff point for best sensitivity and specificity. This HKELM model is then applied back to the training data and generates customer review data. The sentiment analysis classification rate of proposed HKELM schema is 96%, for specificity value of 90%.

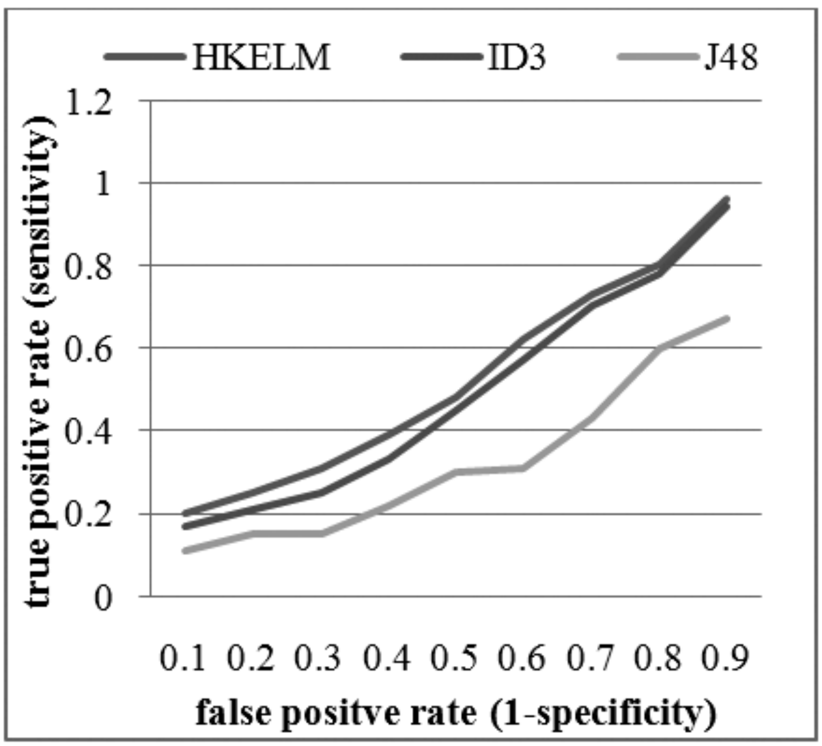

Figure 1: ROC curves of the separation

4.3. Processing Time Comparison

The HKELM sentiment analysis model classifies customers review data and much greater efficiency than the existing sentiment analysis model of ID3 and J48 shown in Fig 2 .The proposed HKELM prediction model takes less computation time to analyze the customer review when compared to existing systems. The HKELM is reportedly working efficiently and in many cases, it’s much faster than ID3 and J48.

4.4. Accuracy, sensitivity, specificity comparison

The HKELM sentiment analysis model classifies customers review data and achieved greater accuracy results than existing sentiment analysis models shown in Fig 3. When the number of products is increases the accuracy of the result is increases. Fig 3 shows the performance evaluation of accuracy, sensitivity and

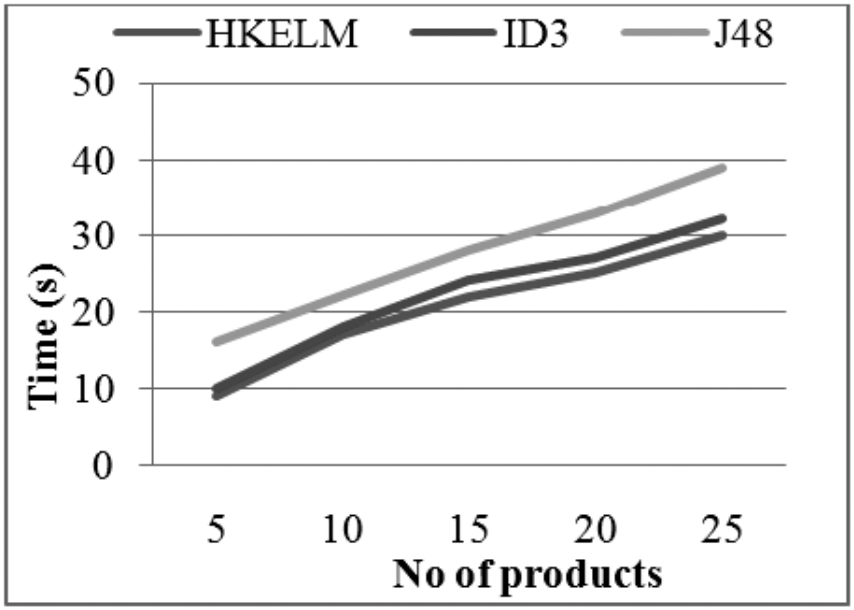

Figure 2: Processing time comparison

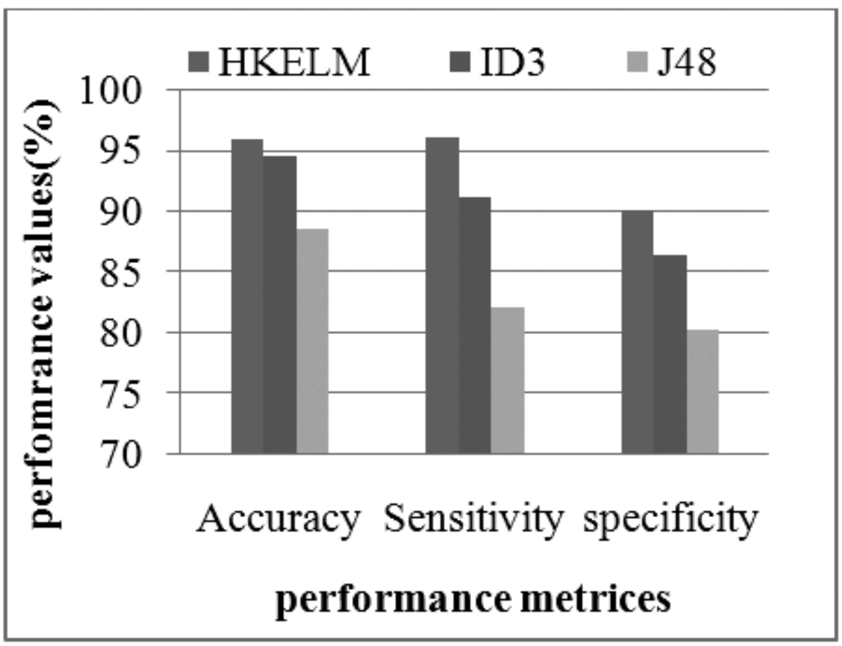

Figure 3: Accuracy, sensitivity, specificity comparison

specificity. The proposed HKELM produces high accuracy rate of 95.55% and sensitivity rate of 96.12% and specificity rate of 90.12% when compared to existing ID3 and J48.

4.5. Precision, recall and F-measure comparison

Fig 4 shows the performance evaluation of proposed and existing schemes. It shows the comparison of Precision, recall and F-measure performance for proposed and existing schemes. The proposed HKELM produces high precision, recall and F-measure rate when compared to existing systems. It defines the proposed scheme has high precision rate of 95.1%, recall rate of 88.14 % and F-measure rate of 87.45% when compared to existing ID3 and J48.

Fig 5 shows the performance cost of proposed scheme. It shows the number of iteration or products increases means the performance cost is decreased.

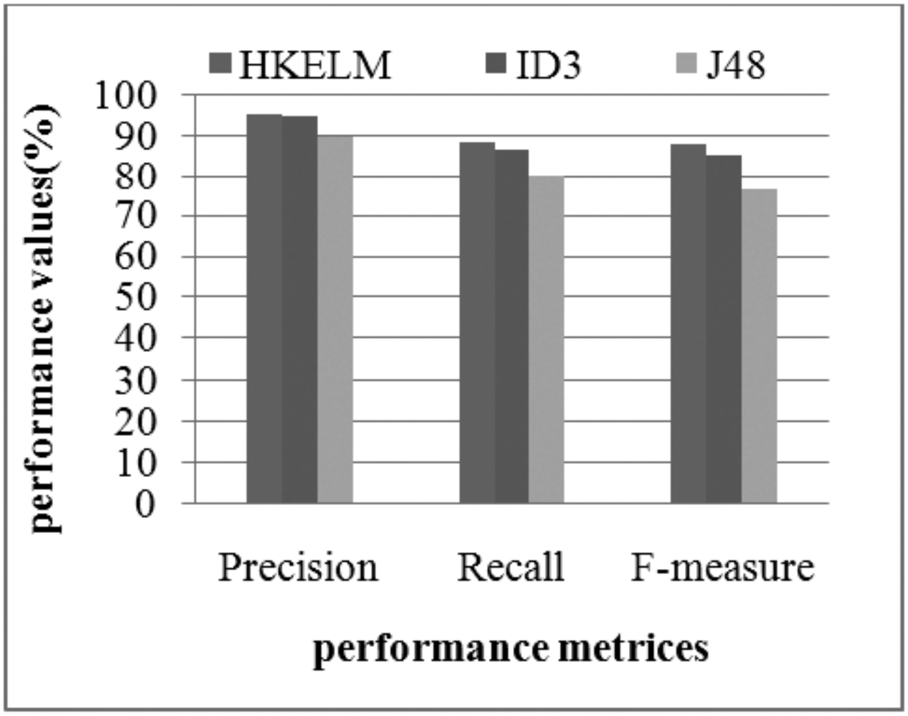

Figure 4: Precision, recall and F-measure comparison

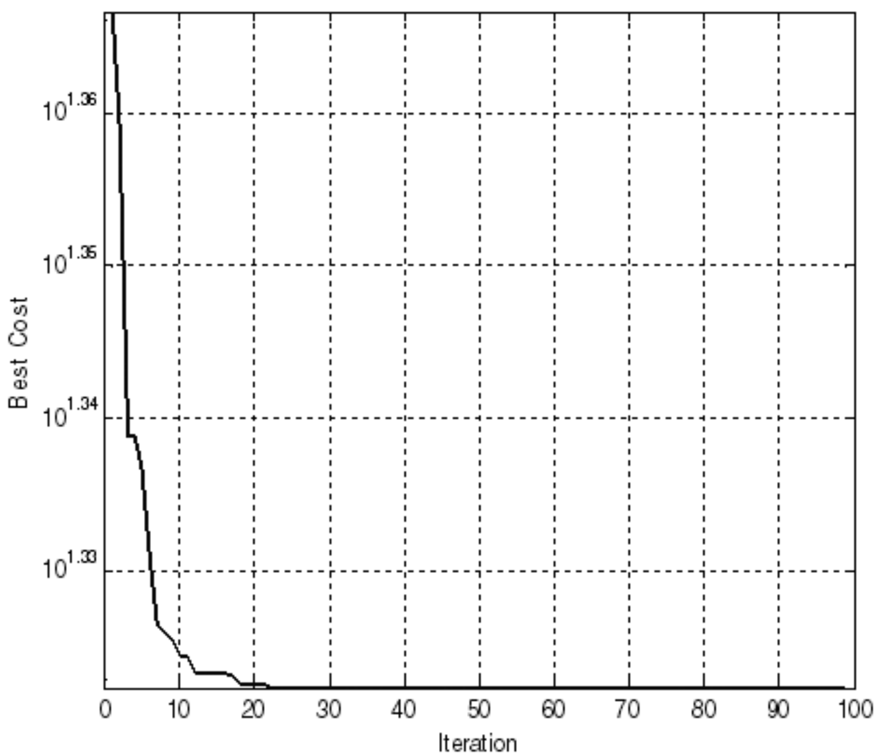

Figure 5: performance cost of proposed scheme

5. CONCLUSION

In this paper, a HKELM algorithm has been proposed based on the KELM and PSO with self-adaptive parameters for efficient sentiment analysis. Initially, the data is pre-processed using EM clustering algorithm to improve proposed scheme accuracy. Secondly, MAR is used for efficient feature extraction. Then, MCSO is used to rank the features and opinion score calculated for noun. MCSO is used to increases speed of processing time. Finally, the maximum used opinion words are extracted using HKELM algorithm. In HKELM, the kernel functions parameters are changed for searching the optimal values by using PSO algorithm. The obtained performance results demonstrate that the proposed HKELM scheme attained efficient performance in terms of accuracy, sensitivity, precision, specificity, recall and f-measure compared than existing sentiment analysis schemes. In future, enhance new methods for data preparation and rank and opinion score prediction. And other classification methods are used to classify the sentiment analysis.

REFERENCES

[1] Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis–An Assessment of Peoples' Belief: A Survey. International Journal.

[2] Vinodhini, G., &Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. International Journal, 2(6).

[3] Buche, A., Chandak, D., &Zadgaonkar, A. (2013). Opinion Mining and Analysis: A survey. arXiv preprint arXiv:1307.3336.

[4] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[5] Farhadloo.M, Rolland.E, "Multi-Class Sentiment Analysis with Clustering and Score Representation" 13th International Conference on Data Mining Wokshop, IEEE, 2013.

[6] S.Njolstd.P, S.Hoysaeter.L, Wei.W and Atle Gull.J. "Evaluating Feature Sets and Classifiers for Semtiment Analysis of Financial News," IEEE/WIC/ACM International Joint Conferences on Web Intelligence(W1) and Intelligent Agent Technologies(IAT),IEEE,2014.

[7] Valakunde.N, Patwardhan.M, "Multi-Aspect and Mutli-Class Based Document Sentiment Analsis of Educational Data Caterin Accreditaion Process" International conference on Cloud & Ubiquitous Computing & Emerging Technologies,IEEE,2013.

[8] Liu.L, Nie. X, Wang. H, "Toward a Fuzzy Domain Sentiment Ontology Tree for Sentiment Analysis" 5th International Congress on Image and Signal Processing(CISP 2012),IEEE,2012.

[9] Sumathi.T, Karthik.S, Marikkannan.M, "Arificial Bee Colony Optimation For Feature In Opinion Mining", Journal of Theoretical and Applied Information Technoloy, Elsevier, 2012.

[10] Basari.A, Hussin.B, Ananta.I, Junta Zeninarja, "Opinion Mining of Movie Review using Hybrid Method of support Vector Machine and Particle Swarm Optimization" Malaysian Technical Universites Conference on Engineering & Techonlogy 2012, MUCET 2012 Part 4-Information And Communication Technology, Elsevier, 2013.

[11] Manju S.R, Kalaiman. E, R. Bhavani, "Product Aspect Ranking Using Sentimentic Oriented Sentiment Classifier" IJERA, 2014.

[12] Vigneshkumar K, Gnanavel S, "Mining Online Reviews for Predicting Sales Performance in Movie Domain" TIJCSA, 2013.

[13] Basiri. M.E, Naghsh-NilchiA.R, AND Ghasem-Aghaee.N, "Sentiment Prediction Based On Dempster-Shafer Theory Of Evidence" HINDAWI, 2014.

[14] Tripathi. G, S.N, "Opinion Mining: A Review" IJICT, 2014.

[15] Joshi.N, Itkat. S, "A Survey on Feature Level Sentiment" IJCSIT, 2014

[16] Chu, S.C, Tsai, P.W, and Pan, J.S. 2006, CatSwarm Optimization, LNAI 4099, 3 (1), BerlinHeidelberg: Springer-Verlag, pp. 854– 858.

[17] Santosa, B, and Ningrum, M. 2009, Cat Swarm Optimization for Clustering, International Conference of Soft Computing and Pattern Recognition, pp. 54-59.

[18] Orouskhani, M, Mansouri, M and Teshnehlab,M. 2011, Average-Inertia weighted Cat swarm optimization, LNCS, Berlin Heidelberg: SpringerVerlag, pp. 321–328.

[19] G.-B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," International Journal of Machine Learning and Cybernetics, vol. 2, no. 2, pp. 107–122, 2011.

[20] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," IEEE Transactions on Systems, Man, and Cybernetics B,vol. 42, no. 2, pp. 13–9, 2012.

[21] S.F. Ding, Y.A. Zhang, X.Z.Xu, and L.N. Bao, "A novel extreme learning machine based on hybrid kernel function," Journal of Computers,vol. 8,no. 8, pp. 2110–2117, 2013.

[22] R. C. Eberhart and J. Kennedy, “New optimizer using particle swarm theory,” in Proceedings of the 6th International Symposium on Micro Machine and Human Science, pp. 39–43, October 1995.

[23] A. Ratnaweera, S.K. Halgamuge, and H.C. Watson, “Self organizing hierarchical particle swarm optimizer with time varying acceleration coefficients,” IEEE Transactions on Evolutionary Computation, vol. 8, no. 3, pp. 240–255, 2004.

[24] F. Han, H. Yao, and Q. Ling, “An improved extreme learning machine based on particle swarm optimization,” Neurocomputing, vol. 116, pp. 87–93, 2013.

[25] Saravanan, T. M., & Tamilarasi, A. (2016). Effective Sentiment Analysis for Opinion Mining Using Artificial Bee Colony Optimization. Research Journal of Applied Sciences, Engineering and Technology, 12(8), 828-840.

[26] Kamal, A. and M. Abulaish, 2013. Statistical features identification for sentiment analysis using machine learning techniques. Proceeding of the International Symposium on Computational and Business Intelligence (ISCBI, 2013). New Delhi, pp: 178-181