IJCTA, 10(20), 2017, pp. 139-152 ISSN: 0974-5572

An Efficient Hierarchical Improved Relevance Vector Machine for Effective Sentiment Analysis

T.M. Saravanan* and A. Tamilarasi*

ABSTRACT

Sentiment analysis is the procedure by which information is extracted from the opinions, reviews and emotions of people in regards to events, entities and their attributes. In decision making, the opinions of others have a significant effect on customers ease in making choices regards to online shopping, choosing events, products, entities. In this work, an efficient Hierarchical Improved Relevance Vector Machine (HIRVM) algorithm proposed to categorize the sentiments of customer's in web applications. Initially, the data is prepared by pre-processing using Independent Component Analysis (ICA) to remove unwanted data. Then Windowed Multivariate Autoregressive Model (WMAR) is proposed for potential feature extraction. After that, the extracted features are ranked and opinion scores are provided for every noun using Improved Bat Algorithm (IBA). Finally, to extract the maximum used opinion words using HIRVM algorithm, which classify the review data based on the opinions. In HIRVM, Analytical Hierarchy Process (AHP) method is used to select the input weights and hidden biases. It is used to analytically determine the output weights and the Iterative Learning Mechanism (ILM) algorithm is employed in order to learn the review through IRVM. It is established by developing a probabilistic Bayesian learning structure which is capable enough to derive accurate prediction models. The experimental results show that proposed model is better than existing sentiment analysis classifier models based on evaluation metrics.

Keywords: sentiment analysis, opinion mining, hierarchical improved relevance vector machine, bat algorithm, iterative learning mechanism.

1. INTRODUCTION

The sudden increase and success of Web 2.0 applications has invented various social media and user produced contents. These contents are magnificent benefits for supporting censorious business intellect applications. The user information gathered from social media and is used to develop the services with implementation novel ideas. It helps to identify the users need and to offer them some new products; it will increase economic of business. Online user reviews are the censorious one along with various user or consumer generated contents. Consistent with a review of 2,400 U.S. adults [1], 81% of Internet users use web applications to do investigate about particular product or item to purchase the product based on users reviews. Another review of more than 2,000 U.S. internet users [2] performed by the Kelsey Group and comScore in 2007, 24% of users check with users reviews with most prior one considered as a paying service and 79% of users review information that consumer reviews have a major influence on their shopping judgment. In addition, various types of service have been considered in business such as excellent, good, normal, and bad. Those all are decided from the consumer reviews.

Nowadays various websites have been launched for gathering consumer reviews, and is very helpful for customers and product manufactures. In product manufacture fields, reviews are useful to analyse the customer's response at the same time improves the marketing promotion. Rapidly, the numeral of customer reviews has been increased; it becomes issue for users to attain complete view of consumer opinions

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

relating to the item of interest by manual analysis. Therefore, to develop an effectual and successful SA, the review is significant and efficient scheme competent of concluding the sentiments of consumer reviews automatically. This analysis has two main functions: 1. Feature extraction: it extracts all feature of item in the consumer review, 2. Identification of opinion orientation: it analysis sentiments on the item feature in the consumer reviews [3, 4 and 5].

Though, various researches [6-9] have been presented on the study of sentiment analysis, but they have some difficulties and that have not been solved effectively. In [6], the product feature extraction was not identifying the exact features and categorization or grouping of product feature synonyms are necessary, and it's not addressed in this analysis. The issue of absence of opinion words [7] and context or domain variations [5] are not effectively addressed in the opinion orientation recognition subtask. The task was using opinion words to decide sentiment types of extracted product features. The above mentioned issues of [7, 5] has been focused in few studies [8, 9] in sentiment analysis; however it only addressed in single point and it produced a combined framework for effectual SA.

Accordingly, the intention of this research work is to propose a sentiment analysis scheme competent of compacting with most of the abovementioned issues. In particular, this work proposed HIRVM algorithm for efficient categorization of customer's sentiments in Web 2.0 applications. To improve the prediction accuracy the data are pre-processed. For effective sentiment analysis, the potential features are extracted by WMAR. To reduce the processing time, the extracted features of opinions are ranked by IBAT. Finally, the HIRVM used to classify the customer's sentiments with effectual results compared with existing sentiment analysis methods.

2. LITERATURE SURVEY

In this section, sentiment analysis of opinion mining literature survey has been discussed. Yang et al., [10] presented two supervised learning schemes such as class association rules and naïve Bayes classifier. These two schemes are used to categorize opinion sentences into suitable product feature classes. Though, these schemes attained sensible efficiency, and preparation task for training examples was more time consuming. In addition, the efficiency of these approaches was based on the excellence of the training examples.

Jeevanandam et al., [11] extracted movie review features from Internet Movie Database (IMDb) by using Inverse Document Frequency (IDF) and the significance of the word found. Then, the features are selected by using Principal Component Analysis (PCA) and Classification and Regression Tree (CART) based on the significance of the work along with deference to the whole document. Finally, the sentiment classifies using by Learning Vector Quantization (LVQ). The obtained classification accuracy was 75% and positive opinions based precision was somewhat low.

Chetashri Bhadane et al., [12] proposed an integrated machine learning of Support Vector Machine (SVM) and domain specific lexicons to identify the polarity and classification of product review. In SVM process, the reviews are trained and classified the polarities. This scheme obtained result show the accuracy of 78%. But the system reliability was not good.

Weiyuan Li and Hua Xu [13], presented a text based emotion classification by the extraction of emotion cause. The emotion cause extracted by using web based data and also features are selected. Then, the selected features are merged in training process. In training, the unknown web post are given in to SVM and Support Vector Regression (SVR) classification schemes and it gives high classification accuracy. It has some challenges and high error.

Kai Gao et al., [14] introduced a rule-based scheme to emotion cause module recognition for Chinese micro-blogs. This scheme extracted emotion cause components in fine-grained emotions. From the quantity of cause component, the emotional lexicon can be created by automatically and manually. For now, in the influence of the multi-language features based on Bayesian probability used for calculated the proportions of cause components. The obtained results demonstrate the achievability of the scheme. The accuracy of prediction was less.

Jeyapriya et al., [15] presented the Bayesian algorithm based classification in product customer reviews and also extract features. From each review word, the nouns and noun phrases are extracted. To find all frequent features for the given review sentences, the Minimum support threshold was used. Then, to classify whether review word was positive or negative opinion by using Naïve Bayesian algorithm also recognizes the how many number of reviews. The prediction accuracy has achieved high but time consumption also high.

Ion Smeureanu et al., [16] proposed a naive Bayes scheme for sentiment analysis for user review of movie. The two classes are considered such as positive and negative. In training process, the predefined positive and negative based collected movie review data are used. The insignificant words are removed in this scheme and it increased classification accuracy. But the system stability was less and not focused those words which do not state opinions.

Richa Sharma et al., [17] presented a polarity calculation for mobile phone reviews data from amazon website and Wordnet used as a dictionary. The majority of opinion words are used to calculate the polarity. The obtained consequences of AIRC Sentiment analyzer scheme were evaluated and this scheme produced high accuracy. The sentence such as not, but, only, ne neither-nor, eitheror etc or not addressed in this scheme, these are decrease performance efficiency of system and take high computation time.

Richa Sharma et al., [18] suggested an Aspect based Sentiment Orientation System (ASOS) to extract and classify the opinion sentence. In this scheme, the features are extracted from review sentences and classified as positive review, negative review or neutral for extracted features and also it handled the negation. Unsupervised scheme was adopted with dictionary based schemes to establish the semantic orientation of the words (i.e. sentences). In this process, the opinion sentences and their synonyms and antonyms established by wordnet and it has been used as a dictionary. The orientation of word has been identified for each feature based on extracted features. The majority of opinion word only used to calculate polarity in this scheme. The obtained results demonstrate it attained good accuracy and system reliability was not good.

Haddi, Lui and Shi [19] explained the sentiment analysis for movie feedback of online data. In this text, the unwanted noise is decreased by using various pre-processing methods. In this system, chi-squared scheme was used to pre-process the data and it has benefit that does not collision on its orientation. It increased the accuracy of prediction category. The Sentiment analysis of accuracy was predicted by using SVM. The training and testing speed and size was less.

Valiati and Neto [20] presented two approaches such as SVM and Artificial Neural Network (ANN) is evaluated based on the condition to attain better accuracy for classification. In particular terms, all schemes are evaluated as a task in bag-of-words. This sentiment learning essential findings is in two pints. First, classification accuracy for benchmark dataset of movies reviews. Second, complete evaluation in the context of balanced data. The obtained result shows the ANN outperform than SVM. ANN does not give data about the relative significance of the various reviews.

3. PROPOSED METHODOLOGY

In this section, the customer's sentiment is analysed based on the proposed HIRVM has been discussed. Step by step process of sentiment analysis has been described in given below subsections.

3.1. Data pre-processing using ICA

Data pre-processing set is often beneficial to remove the stop words and stemming words and reduced the data dimensionality of the review data using ICA. It can be revealing the hidden data in user's review data and it can be improved the accuracy of analysis. It is not mainly focused reduction of dimensionality and also moderated the dimensionalities of data. The basic procedure of this process, initially, the data is linearly diverse with a collection of particular independent signal sources. After that, theses signal sources dissimilar consistent among their numerical independency computed by mutual data. The efficiency of this scheme is validated by using Gaussian source. In this ICA process, the irrelevant data are mentioned as the high dimensionality and it will remove from the data set. Considered, N-dimensional data vectors defined as $\left[V_X^{(1)}, V_X^{(2)}, \ldots, V_X^{(N)}\right]$, the vectors are independent components, which the data of projections of the data vectors are independent of each other. Here, T unknown matrix and it defines the transformation from the present reference data to the independent component reference data and the random variables defined as

$$V_x = Tp\left(e\right) \tag{1}$$

Such that $p(e) = \pi_{pa}(et)$. Where $p(e) \rightarrow$ the joint distribution over the n-dimensional vector e, $pa(et) \rightarrow$ the marginal distribution. Generally, the linear mixture of observed variable defined as

$$V_y = WV_X \tag{2}$$

Where W is the demixing matrix, each component of V_Y is defines the each y_i becomes independent of each other. If the entity marginal distributions are non-Gaussian, then the obtained marginal densities become a scaled variation of the original density functions and is defined as W and is derived from the natural gradient descent of Kullback-Leibler divergence [21] among joint density and the product of marginal densities and is defined as

$$\Delta W = \eta\left(I - \phi\left(V_y\right)V_y^T\right)W \tag{3}$$

Where $\eta > 0$ defined as leaner rating, I is mutual information that measure of random variables have other random variables, and $\phi(V_y)$ is a nonlinear function of the output vector V_y.

3.2. Feature Extraction using WMAR

For individual products, The MAR model extracts features from user's reviews, so, feature vector values are signified as a linear summation of earlier activities. Assumed, d to be the time series review data with n attributes and where m be the order of MAR(m) model for potential features extraction, y_n is linear summation is defined as

$$u_n = \sum_{i=1}^{m} u_{n-i} Q\left(i\right) + e\left(n\right) \tag{4}$$

Where $u_n = \left[u_n\left(1\right), u_n\left(2\right), \ldots, u_n\left(d\right)\right] \rightarrow$ the n^{th} $n$sample of a d-dimensional user's review related extracted features results at various stages and $Q(i) \rightarrow a\ d$ – by – d weight matrix values for each review data of each subjects. And $e_n = [e_n(1), e_n(2), \ldots, e_n(d)] \rightarrow$ preservative Gaussian noise for each data with consideration of zero mean value and covariance R. Here the Gaussian additive noise assessment is used to approximation of user's review features results from MAR. General illustration of MAR model for feature extraction is represented as follows,

$$u_n = x_n W + e_n \tag{5}$$

where $x_n = [u_{n-1}, u, \ldots \ldots, u_{n-m}]$ represents the m previous feature vector results from multivariate model for each user's review data, W is a $(m \times d)$ – by -d matrix of MAR coefficients. If the n^{th} rows of user's review data U with X and E are respectively u_n, x_n and e_n, where $n = 1 \ldots N$ user's review training samples can be written as,

$$U = XW + E \tag{6}$$

where U is an $(N-m)$ – by -d matrix, X is an $(N-m)$ – by - $(M \times d)$ matrix and E is an $(N-m)$ –by –d matrix. In equation (6) MAR model is not effectively extract all features in user's review data's. For the efficient feature extraction, MAR model of reformulation operation is implemented via the Maximum Likelihood (ML) result for each user's review data along with MAR coefficients,

$$\widehat{W} = \left(X^T X\right)^{-1} X^T Y \tag{7}$$

The ML feature extraction results for user's review data (UR_{ML}) is defined as below

$$UR_{ML} = \frac{1}{N-k}\left(U - X\widehat{W}\right)^T \left(U - X\widehat{W}\right) \tag{8}$$

where $k = m \times d \times d$. Weight values based feature (user's review) vector coefficient values defined as $\widehat{w} = vec\left(\widehat{W}\right)$ where *vec* denotes the total number of column weight values feature vector for each user's review data. The re-estimation process of $\widehat{W}$ is carried out by restore the value of $\widehat{w}$ using eqn (7). Then, ML solutions of vector weighted coefficient matrix $\widehat{w}$ is represented as

$$\hat{\Sigma} = UR_{ML} \otimes \left(X^T X\right)^{-1} \tag{9}$$

Where ⊗ mentioned the Kronecker product. The *m* of best value can be predicted by using Minimum Description Length (MDL) [22]. It is also improves the performance results of Bayesian structure [23]. The ML feature extraction from review data along with MAR coefficients is assumed as initialization to Bayesian approach. The Bayesian structure accurately analysis the results of feature extraction results for review data $N(0, P^{-1})$ with mean bc and precision b^2c. Also, $Ga(b, c)$ is the represented as gamma distribution along with the parameters b and c. The Bayesian structure makes use of the probabilities to estimate feature extraction results for each user's review data,

$$p\left(W \mid m\right) = N\left(0, \alpha^{-1} I\right)$$

$$v \qquad {}_{i}^{t} \quad {}_{j}$$

$$p\left(\Lambda \mid m\right) = \left|\Lambda\right|^{-(d+1)/2} \tag{10}$$

where m is the order value of MAR model, α is the precision value of weights feature vector which drawn randomly and Λ is the noise precision matrix posterior likelihood value for each review datas distributions are specified by,

$$p\left(W \mid Y, m\right) = N\left(\widehat{W}_{B,} \widehat{\Sigma}_B\right)$$

$$p\left(\alpha \mid Y, m\right) = Ga\left(\widehat{b}, \hat{c}\right)$$

$$p\left(\Lambda \mid Y, m\right) = Wi\left(s, B\right) \tag{11}$$

3.3. Improved Bat Algorithm (IBA)

In IBAT, bats 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. Traditional bat algorithm works efficiently, but it has some insufficiency [24] due to the easy trap of local minimum function in multimodal test functions. To improve the bat process, the velocity equation is updated by inertia weight modifications. The step by step process of proposed scheme is described in given below.

3.3.1. Bat Population initialization process

First, randomly initialized the population generated from real valued vectors with d-dimension and an n-number of bats along with lower and upper boundaries are considered as given below

$$p_{ij} = p_{min\,j} + \varphi\left(p_{max\,j} - p_{min\,j}\right), i = 1,\ 2,\ \dots\ n, j = 1,\ 2,\ \dots\ d \tag{12}$$

Where $p_{min\,j}$ defined as lower boundary and $p_{max\,j}$ are defined as upper boundary for dimension j, φ mentioned randomly produced value from 0 to 1.

To improve the local search ability, the given below structure has been used

$$p_{new} = p_{old} + \sigma A^t \tag{13}$$

Where $p_{old} \rightarrow$ high quality solution of existing, $A^t \rightarrow t^{th}$ time step of average loudness value for all bats and $\sigma \rightarrow$ randomly generated value from -1 to 1. The loudness and pulse emission are represented as A and r. In case, the r increased means the A is decreased.

3.3.2. Frequency, velocity and new solution generated process

Generally, in bat process, randomly selected the frequency value and is allocated a solution. But, this frequency value has attained same result for all dimensions and it reduces the local search ability. To improve the local search, in IBA, a frequency fromf_{min} to f_{max} is assigned for each dimension of a solution separately.

$$dis = max\left(\sqrt{\left(p_{ij} - p_j^*\right)^2}\right) - min\left(\sqrt{\left(p_{ij} - p_j^*\right)^2}\right) \tag{14}$$

$$f_i = f_{min} + \frac{\sqrt{\left(dis\right)^2}}{dis} * \left(f_{max} - f_{min}\right) \tag{15}$$

Where f_{min} and f_{max} are defined as minimum and maximum frequency values correspondingly. Forever depends on frequency only the step size of a solution has been defined. The basic velocity formulation has some premature convergence problem. To overcome the problem, the given below equation is used as the velocity update

$$v_{ij}^t = \omega\left(v_{ij}^{t-1}\right) + \left(p_{ij}^t - p_j^*\right) f_i \tag{16}$$

Where v_{ij}^t is velocity update and ω is defined as inertia weight value. It is used to balance the local and global search intensity of the i and j^{th} solution via controlling the value of old velocity.

3.4. IBAT for ranking the features based on opinion

In this section, the features are ranked and their opinion scores are calculated using IBA. In this process, the position of bats encodes the subset of features and the detailed explanation is given in algorithm 1. First, the population of bats are initialized. Then, randomly the bats' position is initialized with chosen binary values. According to whether a feature will be has high or not to compose the opinion score of data. In addition, if a new solution has been accepted, then loudness and pulse emission are updated. In order to, the min and values are observed. At last, the output vector with the ranked features is produced and revisited by the algorithm.

Algorithm 1: ranking the feature and opinion score calculation

Input: m-Population size, n-number of features, T-number of iterations, A-loudness, and pulse emission rate r.

Output: Subset of features F

Objective function: $f(p)$, $p = (p_1, \ldots p_d)t$

bat population p_i and velocity v_i are Initialized defined pulse frequency f_i and initialized r, A while ($t <$ *maximum number of iterations*) Generated New solutions and velocity updated by using ω

F (rand>r)

one best solution is selected among the solutions around the selected best solutions a local solution is generated

end if

if (rand<A)

stored new solution in memory when r increased, A reduced

end if

the bats are ranked and current best p^* mean found while the opinion score is represented

end while

selected features are displayed

3.5. HIRVM based classification

In this section, the reviews are classified into types of sentiments (i.e Positive, Negative and Neutral) based on the users opinions that are incorporated in the reviews using HIRVM classification algorithms.

3.6. Improved Relevance Vector Machine

The given training inputs $\{p_i, t_i\}_{i=1}^{n}$, where $p_i \in R^n$, $t_i \in \{0, 1\}$ and n is the number of samples. The RVM makes predictions for new inputs $\hat{p}$ based on the Support Vector Machine (SVM)-like function; this scheme obtains the form of a linear mixture of basic functions transformed through a logistic sigmoid function

$$q(\hat{p}, w) = \sigma\left(\sum_{i=1}^{n} \omega_i k(p_i, \hat{p})\right) = \sigma\left(w^T K\right) \tag{16}$$

Where $k(\hat{p}) = \left[k(p_1, \hat{p}) \ldots k(p_n, \hat{p})\right]^T$ is the kernel function vector, $w = (\omega_1 \ldots w_n)^T$ is the weight vector, and □σ(.) is the logistic sigmoid function defined by:

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \tag{17}$$

The logistic sigmoid function assures the given below symmetry property

$$\sigma(-a) = 1 - \sigma(a) \tag{18}$$

Consequently, RVM scheme can be used as the posterior probability. For the input $\hat{p}$, the class c_1 posterior probability can be defined as

$$Pr\left(t=1 \mid \hat{p}\right)=q\left(\hat{p},w\right) \tag{19}$$

Correspondingly, the class c_2 posterior probability can be defined as

$$Pr\left(t=0 \mid \hat{p}\right)=1-p\left(\hat{p},w\right) \tag{20}$$

RVM model can be treated as the posterior probability, because it adopts a Bayesian probabilistic framework to train the system. The key feature of RVM is using the Automatic Relevance Determination (ARD) prior over the weight vector w, in which there is a separate hyper parameter α_i for each of the weight parameters ω_i. Through the inference procedure, a lot of hyper parameters are determined to high values, therefore that corresponding weights are efficiently forced to zero. Thus the corresponding kernel functions can be pruned out, resulting in a sparse model. The remaining nonzero weights of the inputs p_i are known as relevance vectors.

For an input vector $\hat{p}$, the RVM decision model, as pre-defined by Equation (16), can be rewritten only based on the w_{MP} and RVM as given below

$$q\left(\hat{p},w_{MP}\right)=\sigma\left(\sum_{p_i\in RVM}\omega_i k\left(p_i,\hat{p}\right)+\omega_0\right) \tag{21}$$

As can be seen in Equations (16) and (21), kernel function plays an important role in the RVM decision model. There are several common kernel functions for selection, such as linear, polynomial, sigmoid, Gaussian Radial Basis Function (RBF) and so on. In this improved RVM the Elliptical Radial Basis Function (ERBF) used for kernel function.

$$\left(p,z\right)=\exp\left(-\sum_{i=1}^{D}\left(p_i-z_i\right)^2/\left(\sigma_i^2.r^2\right)\right) \tag{22}$$

Where p and z are D-dimension feature vectors (i.e. = $(p_1, \ldots p_D)^T$, $z = (z_1, \ldots z_D)^T$, r is scale factor, σ_i^2 variance.

3.7. Analytical Hierarchical Process (AHP)

AHP is a multi-criteria decision-making scheme has initiated for competent mathematical properties of the scheme and the required input data can be easily established from this scheme. A set of pair wise comparisons are used to derive the relevant data. To obtain the weights of significance of the decision criteria, these comparisons are used and the comparative performance computes of the options in terms of each individual decision criterion. If the comparisons are not completely reliable, then it offers a system for improving reliability.

3.8. Levenberg-Marquradt Learning

The extension of Error Back Propagation (EBP) algorithm is called as Levenberg-Marquardt (LM) learning scheme. It gives an important swap among the Newton algorithm based speed and steepest descent scheme based strength. The LM algorithm contains the above two fundamental theorems. In the EBP, the performance index $PI(w)$ to be decreased and is described as the sum of squared errors among the outputs of target and the outputs of network's simulation,

$$PI(w) = e^T e \tag{23}$$

Where $w = [w_1, w_2 ..., w_n]$ includes all weights of the network, $e \rightarrow$ specifies the error vector comprising the error for all the training samples.

The augmentation of weights Δw, when training with the LM scheme is computed as

$$\Delta w = \left[Jm^T Jm + \mu I \right] - 1 Jm^T e \tag{24}$$

Where $Jm \rightarrow$ indicates the jacobian matrix, $\mu \rightarrow$ denotes the learning rate which is to be updated using the β based on the resultant. Particularly, μ is multiplied via decay rate $\beta(0 < \beta < 1)$ whenever $PI(w)$ reduces, while μ is divided by β whenever $PI(w)$ increases in a new step.

This is attained through adding a small scalar λ to the diagonal elements in the Hessian H as defined as

$$(H + \lambda I)\Delta\sigma = -\nabla E(\sigma_0) \tag{25}$$

In this scheme, initially, a first order scheme is processed and automatically progresses towards the second-order scheme are given below. Based on the equation (25), the second order process computed by numerical perturbation as defined below

$$\left.\frac{\partial E}{\partial \sigma}\right|_{\sigma=\sigma_0} \approx \left.\frac{\Delta E}{\varepsilon}\right|_{\sigma=\sigma_0} = \frac{E(\sigma_i + \varepsilon)}{\varepsilon}\Big|_{\sigma=\sigma_0} \tag{26}$$

Where $\varepsilon \rightarrow$ indicates a small perturbation value acting on the i^{th} component in σ. $E(\sigma_i) \rightarrow$ is the performance metric Q^2 achieved from the change in the i^{th} component of σ before solving the above equations, A second approximation will be introduced. For the reason that the elements of the Hessian are costly to assess, so, a speedy and competent approximation for the Hessian matrix is presented. After that, $\Delta\sigma$ is solved numerically from equation (25) with a quick conjugate gradient based equation solver to keep away from computing the inverse of the Hessian matrix (H). In outcomes of Hessian the fairly accurate evaluation, a heuristic coefficient á will be introduced in the iterative updating procedure for the elements. LM has high convergence speed.

3.9. Hierarchical Improved Relevance Vector Machine (HIRVM) classifier

In this section, a HIRVM scheme has been discussed. The proposed scheme is defined as the integration process IRVM, AHP and LM algorithms. Bayesian counterpart introduced the IRVM. The proposed hierarchical model is prepared using AHP and it is very useful for decision support of sentiment analysis. Then, Relevance Vector framework gives solutions to solve sentiment classification issues. The incessant hyper parameters are used by this sentiment analysis scheme to manage model difficulty and so reduce the in-tractable issue of searching over a differentially large discrete space of model systems. So as to decrease the dimensionality of the hyper parameter space, a prior structure specified which replicates the likelihood of correlation among the hyper parameters of the coefficients distribution and thus it is potential to separate a unique solution.

In this proposed method, IRVM has been used for classification. It is an exceptional case of a sparse linear form in which the basic processes are formed by a kernel function. Then, AHP used for decision support. In IRVM, at each location, it automatically enables selecting good kernel by pruning all irrelevant kernels; therefore it is potential that two different kernels stay on the same location. First, followed by hidden biases the input weights are generated. In analytical hierarchy process method, this method used to choose the input weights and hidden biases. After that, the corresponding output weights are analytically calculated by using the IRVM algorithm once and randomly produce the output hidden biases. Finally, these parameters (all weights and biases) are updated by using LM algorithm. Using this procedure user's reviews are analysed which kind of sentiment mentioned in this system.

4. RESULTS AND DISCUSSION

In this section, the performance of proposed HIRVM sentiment analysis scheme is evaluated with existing sentiment analysis schemes such as Hybrid Kernel based Extreme Learning Machine (HKELM), Interactive Dichotomizer version 3 (ID3) [25] and J48 [26].

4.1. Dataset description

The performance is evaluated by using polarity dataset download from the URL http://www.cs.cornell.edu/people/pabo/movie-review-data. The dataset detailed description is referred from the above link.

4.2. Evaluation Criteria

In this work, the sensitivity and specificity are used to measure the accuracy of the sentiment analysis schemes. ROC analysis used to calculate the accuracy of a classifier. The sensitivity and specificity calculation are referred from [25].

Performance comparison for ROC curves of the separation with the cutoff point for best sensitivity and specificity based graphical representation is showed in fig 1. It illustrates the proposed HIRVM model is attained good performance compared than existing schemes. Because of the data effectual pre-process and efficient feature extraction. The classification rate of proposed HIRVM sentiment analysis schema is 97.21%, and specificity rate of 92.12%.

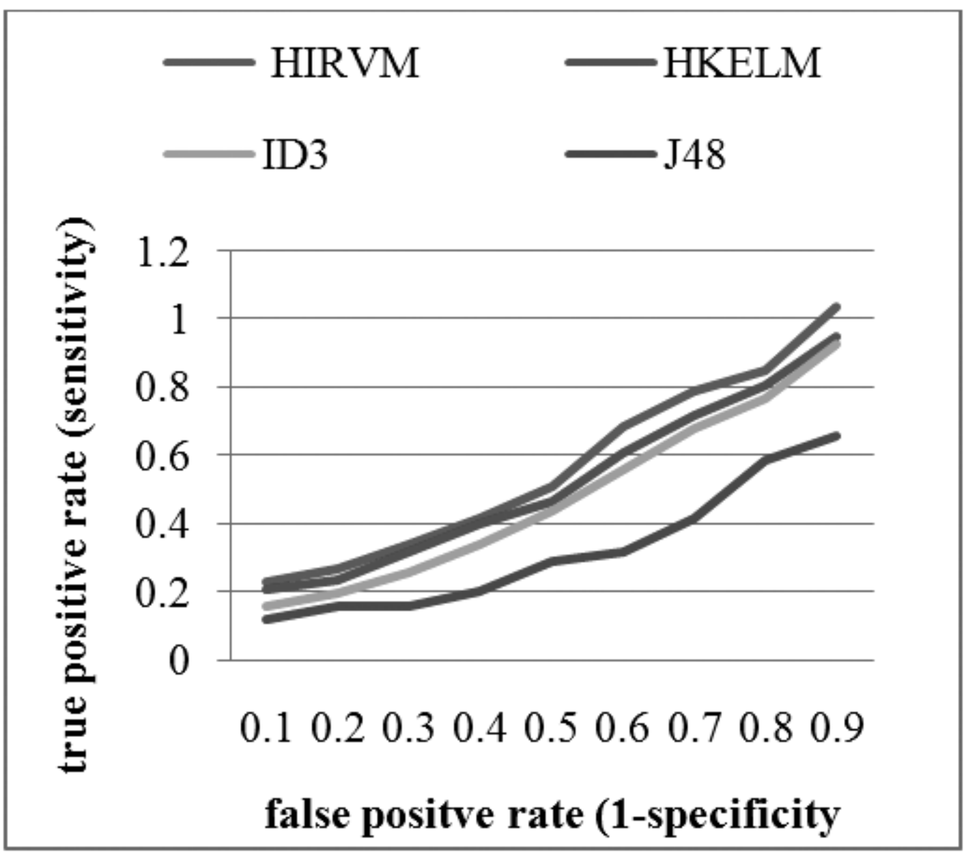

Figure 1: Performance comparison for ROC curves of the separation

4.3. Comparison of Processing Time

In Fig 2 shows the processing graphical representation of proposed HKELM and existing schemes of ID3 and J48. It shows the proposed HIRVM attain high efficiency than the existing schemes, because it takes low time for computation to predict the customer review. In many cases, the HIRVM is working efficiently and much faster than HKELM, ID3 and J48.

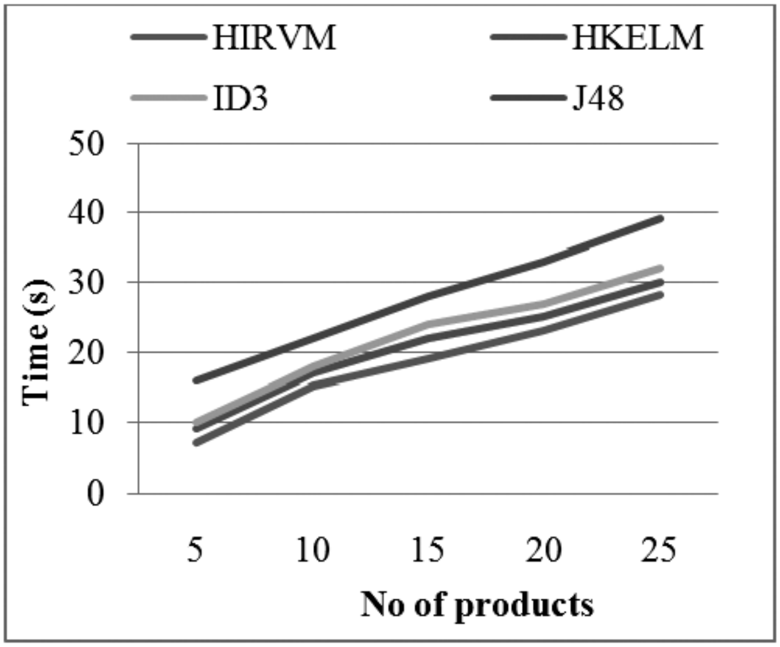

Figure 2: comparison of processing time vs. No of products

4.4. Comparison of Accuracy, sensitivity, specificity

In Fig 3 shows the graphical representation for comparison of Accuracy, sensitivity, specificity for various SA methods such as HIRVM, HKELM, ID3 and J48. If the product reviews or products increased, then the

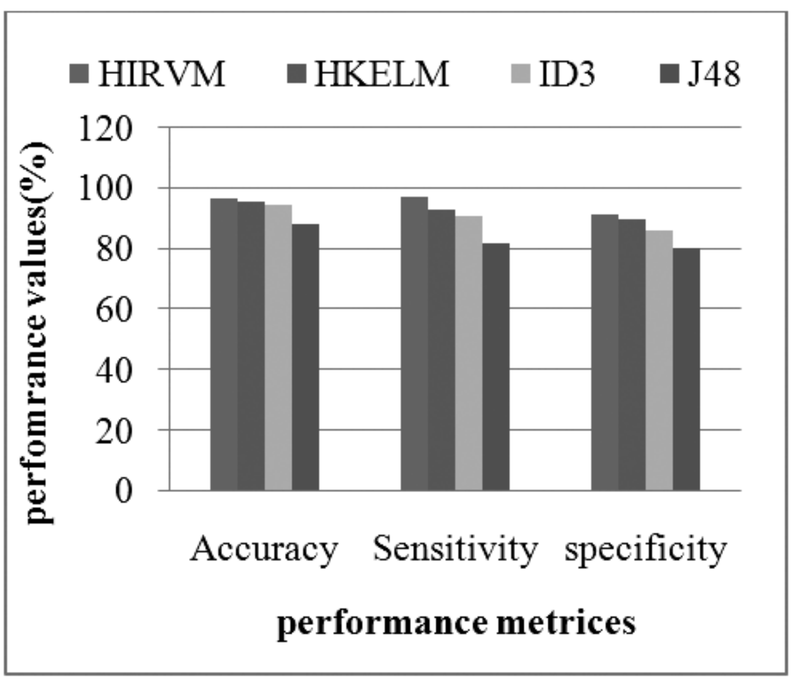

Figure 3: Comparison of Accuracy, sensitivity, specificity for various SA methods

performance also increases. The proposed HIRVM produces high accuracy rate of 96.87% and sensitivity rate of 97.14% and specificity rate of 91.24s% when compared to existing HKELM, ID3 and J48.

4.5. Precision, recall and F-measure comparison

In Fig 4 shows the graphical representation for comparison of Precision, recall and F-measure for various SA methods such as HIRVM, HKELM, ID3 and J48. The proposed HIRVM produces high precision, recall and F-measure rate when compared to previous approaches. It describes the proposed HIRVM scheme has high precision rate of 96.21%, recall rate of 90.24 % and F-measure rate of 89.78% when compared to existing HKELM,s ID3 and J48.

Fig 5 shows the performance cost of proposed HIRVM scheme. It shows the number of iteration (i.e. products) is increases means the performance cost is decreased.

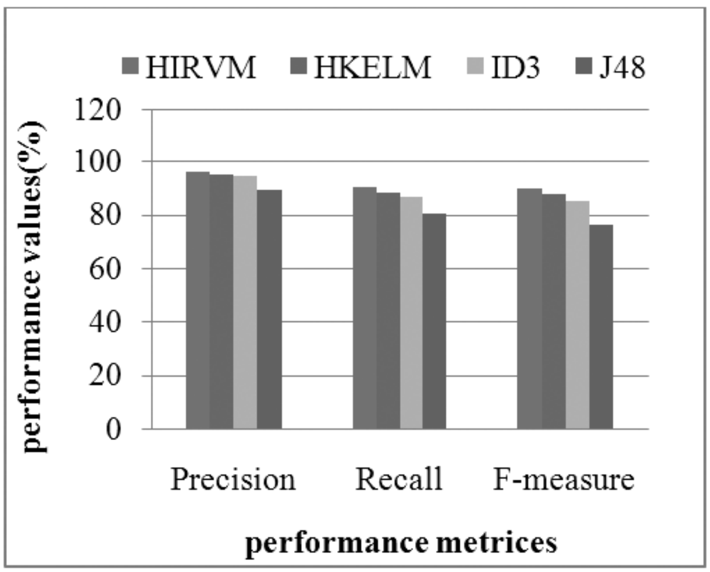

Figure 4: Comparison of Precision, recall and F-measure for various SA methods

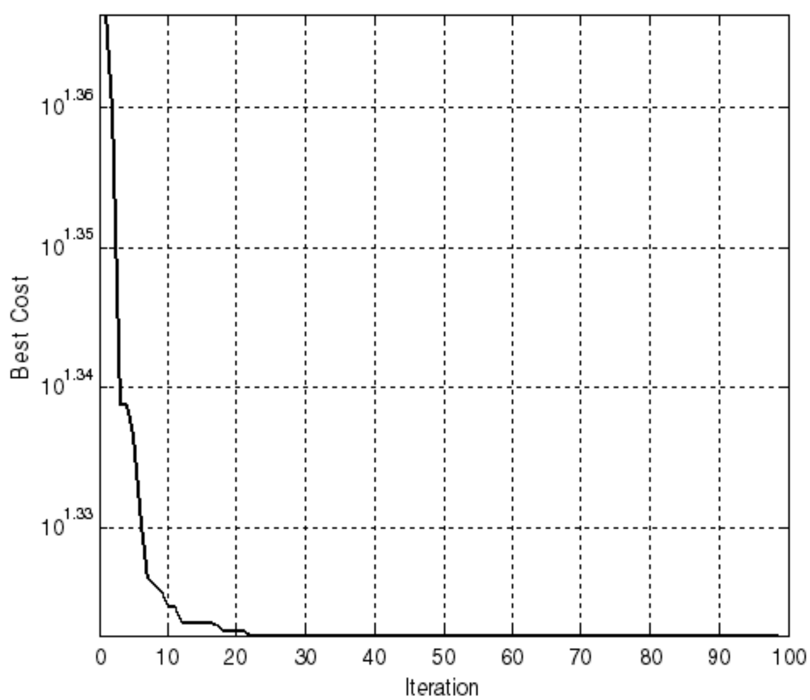

Figure 5: The proposed scheme performance cost

5. CONCLUSION

In this paper, machine learning based hierarchical algorithm is proposed for efficient sentiment analysis of user's review data in web applications. Initially, the data is pre-processed using ICA to improve proposed scheme accuracy and reduced the dimensionality. Secondly, WMAR is used for efficient feature extraction and it reduced the processing time. Then, IBAT is used to rank the features and opinion score calculated for noun and reduced processing cost also increased speed of processing time. Finally, the maximum used opinion words are extracted using HIRVM algorithm. In HIRVM, approach using combination of IRVM, LM and AHP algorithms and the IRVM was introduced by as a Bayesian counterpart. The experimental results show that the proposed HIRVM scheme attained efficient performance in terms of accuracy, sensitivity, precision, specificity, recall and f-measure compared than existing sentiment analysis schemes. In future, various opinion summarization algorithms should be applied to generate summary of all reviews provided by users.

REFERENCES

[1] Horrigan, J. B. (2008). Online shopping. Pew Internet & American Life Project.http://www.pewinternet.org/Reports/2008/Online-Shopping.aspx

[2] comScore and The Kelsey Group (2007). Online consumer-generated reviews have significant impact on offline purchase behaviour. http://www.comscore.com/press/release.asp?press=1928.

[3] Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'04),168-177.

[4] Popescu, A. and Etzioni, O. (2005). Extracting product features and opinions from reviews. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP 2005), 339-346.

[5] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2 (1-2), 1-135.

[6] Liu, B. (2010). Sentiment analysis and subjectivity. In Handbook of Natural Language Processing,(Indurkhya, N. and Damerau, F.J. Eds.), NY, Chapman and Hall/CRC.

[7] Yang, C.S., Chen, C.H., and Chang, P.C. (2011). Improving the effectiveness of opinion orientation identification in sentiment analysis of consumer reviews: A gloss-based approach. In Proceedings of the 5th China Summer Workshop on Information Management (CSWIM 2011).

[8] Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., and Swen, B. (2008). Hidden sentiment association in Chinese Web opinion mining. In Proceeding of the 17th International Conference on World Wide Web (WWW'08), 959-968.

[9] Guo, H., Zhu, H., Guo, Z., Zhang, X., and Su, Z. (2009). Product feature categorization with multilevel latent semantic association. In Proceeding of the 18th ACM Conference on Information and Knowledge Management (CIKM'09), 1087-1096.

[10] Yang, C.C., Tang, X., Wong, Y.C., and Wei, C. (2010). Understanding online consumer review opinions with sentiment analysis using machine learning. Pacific Asia Journal of the Association for Information Systems, 2 (3), 73-89.

[11] Jeevanandam Jotheeswaran, Dr. Y. S. Kumaraswamy, "Opinion Mining Using Decision Tree Based Feature Selection Through Manhattan Hierarchical Cluster Measure", Journal of Theoretical and Applied Information Technology, 2013

[12] Chetashri Bhadane,Hardi Dalal, Heenal Doshi, "Sentiment Analysis: Measuring Opinions", Science Direct, 2015

[13] Weiyuan Li, Hua Xu, "Text-based emotion classification using emotion cause extraction", ELSEVIER, 2013

[14] Kai Gao, Hua Xu, Jiushuo Wanga, "A Rule-Based Approach To Emotion Cause Detection For Chinese Micro-Blogs", ELSEVIER, 2015

[15] A.Jeyapriya, C.S.Kanimozhi Selvi, "Extracting Aspects And Mining Opinions In Product Reviews Using Supervised Learning Algorithm", IEEE, 2015

[16] Ion Smeureanu, Cristian Bucur, "Applying Supervised Opinion Mining Techniques On Online User Reviews", Informatica Economicã, 2012

[17] Richa Sharma, Shweta Nigam, Rekha Jain, "Polarity Detection at Sentence Level", International Journal of Computer Applications, Volume 86- No 11, 2014

[18] Richa Sharma,Shweta Nigam and Rekha Jain, "Mining Of Product Reviews At Aspect Level", International Journal in Foundations of Computer Science & Technology (IJFCST), Vol.4, No.3, May 2014.

[19] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Preprocessing in Sentiment Analysis," Procedia Comput. Sci., vol. 17, pp. 26–32, Jan. 2013.

[20] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," Expert Syst. Appl., vol. 40, no. 2, pp. 621–633, Feb. 2013.

[21] Zhaoyang Zhang, Honggang Wang, Athanasios V. Vasilakos, And Hua Fang ¯ECG-Cryptography And Authentication In Body Area Networks, IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 6, November 2012.

[22] A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. ACMTrans. Math. Softw., 27, 27-57, 2000.

[23] W.D. Penny and S.J. Roberts. Bayesian multivariate autoregressive models with structured priors. IEE Proc.-Vis Image Signal Process, 149(1):33–41, 2002.

[24] G. Wang and L. Guo, "A Novel Hybrid Bat Algorithm with Harmony Search for Global Numerical Optimization," Journal of Applied Mathematics, vol. 2013, pp. 21, 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.