Opinion Analysis Model for Classification of Opinions for M Learning

A.Nisha Jebaseeli, Ph.D
Assistant Professor in Computer Science,
Bharathidasan University Constituent College, Lalgudi
Tiruchirappalli

ABSTRACT

The main objective of this research is to propose an Opinion Analysis Model for classification of opinions with the evolutionary approach and to improve the classification accuracy. The research presented three proposals to heed to various real-world problems. Some processes were posed, described, discussed and compared with existing classification algorithms. An improvement of classification accuracy is noticed when the data is preprocessed by KN preprocessing algorithm. Thus, the formed corpuses were applied in the existing classification algorithms which showed an increase in the accuracy rates. The enrichment of the classification algorithms with the neural network and genetic optimization brought a significant intervention in the classification accuracies. The research, for analyzing, took into account Mobile Learning (M-Learning) dataset from Android Market website. In order to assess the effectiveness of the proposed algorithms, the benchmark experiments are conducted with Internet Movie Database (IMDb) dataset comprising of movie reviews.

Keywords

Opinion Analysis Model (OAM), KN preprocessing algorithm, Classification Algorithms, Neural Network, Genetic Optimization, M-Learning, Internet Movie Database (IMDb) dataset.

1. INTRODUCTION

With the massive growth in the Web, one could post reviews of products at merchant sites and express views on almost anything in Internet forums, discussion groups, and blogs, which are collectively called the user generated content. A formidable task to find and monitor opinion sources on the Web, however, has become a point of departure. Hence, opinion mining models evolve with varied significance and analysis. Many researchers are at work in mining the opinions and identifying the semantic orientations. The objective of these tasks is to classify each review document as expressing a positive or a negative sentiment about an object (e.g., a movie, a camera, or a car or a learning system) [1].

The exponential growth in the field of telecommunication technology offers the features like network coverage, speed and technological innovation. The availability of diverse products in mobile devices such as smart phones, tablets etc. helps the students to access the course content via these devices. Mobile Learning is also helpful to perform learning activities utilizing learner's spare time at any time from any place with wireless networks [2].Thus, mobile learning (m-learning) is defined as: "the process of learning and teaching occurs with the use of mobile devices anywhere (context aspects) and anytime (device aspects)" [3][4]. The rapid growth of mobile communication and hand-held devices put forward a new and innovative pedagogical approach in the learning system. Anytime, anywhere learning scenario offered in M-Learning system enhances the learners learning ability. So it is necessary to improve M-Learning

system, and to know the users' opinions and evaluation about them. It involves applying the automatic text analysis to extract the opinions and adopting automatic sentiment analysis to identify the sentiment of opinions from the reviews on which users are discussing or describing their personal opinions and evaluation of the services. The recent researches on opinion mining suffer a limitation like low classification accuracy [5]. This research paper addresses the above issue and the main goal is to propose Opinion Mining algorithms to enhance classification accuracy in M-Learning.

As the opinions have many potential applications, mining on the reviews of the M-Learning users were contemplated. This research enhances the accuracy of classification by refining the preprocessing methods and competent use of evolutionary algorithms for classification. The real world datasets in general contain inconsistent, noisy and incomplete attributes which pose a challenge of preprocessing the datasets. Artificial Neural Network (ANN) is one of the promising areas in the computing domain. ANNs offer an ability to perform tasks outside the scope of traditional processors. ANNs recognize patterns within vast data sets and then generalize those patterns into recommended courses of action. The primordial distinction against the traditional algorithms is that the neural networks learn and are not programmed. Eventually the art of employing ANN involves the selection of learning rules, transfer functions, summation functions, and how to connect the neurons within the network.

Genetic approaches are finding prevalent applications in solving problems requiring efficient and effective search in synthesis of neural network architectures, scheduling, and numerical optimization. In Genetic training algorithms, the learning of an ANN is formulated as a weight optimization problems such as learning rate, momentum rate, tolerance level are optimized using genetic approach [6]. The use of genetic algorithm in the design of ANN improves the classification Accuracy [7]. Generating the solutions for optimization, the techniques inspired by natural evolution, such as mutation, selection, and crossover are implemented.

1.1 KN Preprocessing

In the formulation of corpus for the classification and further mining analysis, KN Preprocessing is utilized in the present study. The quality of discovered knowledge requires data quality. Unfortunately real data contains noise, uncertainty, errors, and redundancies or even irrelevant information. Preprocessing the data set required great effort when dealing with real world applications [8]. Preprocessing is a critical task, mainly owing to two reasons:

➢ Real data sets tend to be imperfect, contains errors, outliers, missing data, extra noise and tools either for detecting or correcting it are required.

➢ Application of a certain data mining technique may require specific conditions for the data set (only binary

variables, centered data, normality, only qualitative variables, etc) [9].

The datasets were acquired from the reviews on the M-Learning system and they were preprocessed with the existing preprocessing methods namely Stopword Removal, Stemming, Inverse Document Frequency, and Singular Value Decomposition. The proposed KN preprocessing is applied in order to enhance the classification of the opinions. The significant intervention by KN preprocessing involved in introducing confidence interval, which improves the formulation of the corpus. After the preprocessing, the acquired corpus is used with different classification algorithms. The results proved that the classification accuracy enhanced with the use of the proposed preprocessing technique.

1.2 KINN

As ANN, configured for pattern recognition or data classification, is widely perceived to be accurate to the level of 80%-90% [10]. In a move to increase classification accuracy there by deducing uncluttered knowledge, neural network based algorithm is designed. In the proposed KINN algorithm for increasing the classification accuracy, the preprocessed dataset forms the inputs $x_1, x_2, …, x_n$ for the KINN architecture. The input weights presented to the input nodes are offered to the Hidden layer with the product of weights and the data in the first epoch. In the Hidden layer, summation and activation function occurs. For performing the activation function, sigmoidal operation is adopted. As the Bias value is set to 0 in the process, the summation of the calculated value with Bias do not make any significant change.

The output from the Hidden layer is handed to the output layer. In the Output layer, the calculated values from the Hidden layer with the weights from the Input layer are dispensed. In this scheme, the additional weights are feed forward from input layer to the output layer. The additional weights, top five positive and top five negative are selected with the help of Information Gain. The pinnacle and nadir values, calculated through IG, have brought the significant improvement in the classification accuracy.

1.3 GKINN

The relevance in optimization technique will lead to the genetic approach which has a little chance to get stuck at local minima. Genetic approaches are finding prevalent applications in solving problems requiring efficient and effective search in synthesis of neural network architectures, scheduling, numerical optimization [11]. etc., and results in solutions that are globally optimal or nearly so. Topology optimization, genetic training algorithms and control parameter optimization are the ways in which genetic algorithms are applied in Artificial Neural Network (ANN). ANN is the best tool for classification and regression. The parameters of the ANN are optimized with genetic approach to enhance the classification accuracy [12]. When the ANN is combined with genetic approach, it is known as hybrid model which is performed better than the conventional ANN [13].

The Genetic Approach often works with a form of binary coding [14]. If the problems are coded as chromosomes, the populations are initialized. Once the population size is chosen, the initial population is randomly generated. After the initialization step, each chromosome is evaluated by a fitness function. According to the value of the fitness function, the chromosome associated with fittest individuals will be reproduced more often than those associated with unfit individuals. As to outwit the performance of the neural network based algorithms, genetic optimization is regarded in mining the opinions [15]. By genetic optimizing the

moment and learning rate the convergence becomes more profound. Hence, the genetic optimized techniques are used to enhance the classification accuracy [16].

2. ARCHITECTURE OF OAM

The Opinion Analysis Model proposed in this research paper analyses the opinions from the given corpus using methods from machine learning such as support vector machines, stop word removal, stemming, and inverse document frequency. The proposed KN preprocessing is applied on the dataset. Through the medium of KN preprocessing, the uncluttered corpus is formed for the further processes in data mining.
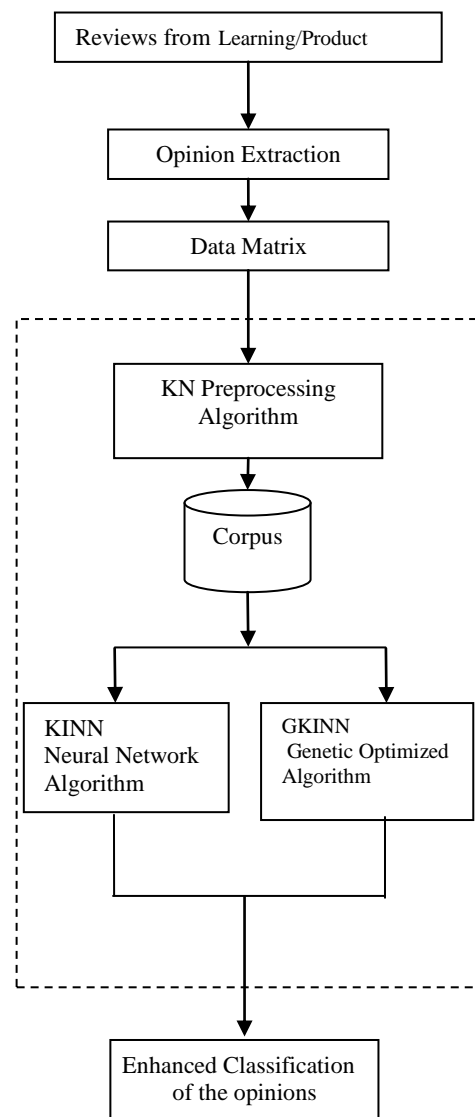
OPINION ANALYSIS MODEL

Figure 1: Opinion Analysis Model

It is essential that the data is preprocessed such that the accuracy of classification and abstraction of knowledge are augmented. In this model, the corpus-formed is classified using the proposed KINN algorithm. The neural networks based algorithm is applied to classify the opinions based on the trained data as they are the better alternative for conventional methods of classification [17]. Moreover, the feed forward neural network algorithm is also enhanced through Input-Output-Positive-Negative values for

opinion classification.

In addition to improve the classification accuracy, genetic optimization is undertaken. The operators that are used for converging to an optimal solution precisely are learning rate and momentum. In the model, the corpus is rendered to the proposed GKINN algorithm for obtaining

It is essential that the data is preprocessed such that the accuracy of classification and abstraction of knowledge are augmented. In this model, the corpus-formed is classified using the proposed KINN algorithm. The neural networks based algorithm is applied to classify the opinions based on the trained data as they are the better alternative for conventional methods of classification [17]. Moreover, the feed forward neural network algorithm is also enhanced through Input-Output-Positive-Negative values for opinion classification.

In addition to improve the classification accuracy, genetic optimization is undertaken. The operators that are used for converging to an optimal solution precisely are learning rate and momentum. In the model, the corpus is rendered to the proposed GKINN algorithm for obtaining better classification accuracy. Thus, the Opinion Analysis Model classifies the opinions efficiently [18]. The proposed researches were evaluated with the bench mark dataset namely IMDb dataset comprising of movie reviews. The obtained results proved that the classification of the reviews by the proposed methods hold an upper hand over the existing methods.

The proposed Opinion Analysis Model involving KN preprocessing algorithm, KINN and GKINN algorithms to enhance the classification accuracy is elaborated in achieving the objectives.

3. RESULTS AND DISCUSSIONS

The proposed Opinion Analysis Model is operated with M-Learning system reviews as dataset. Learner's opinions about free M-Learning system available are considered. The benchmark experiments are also conducted with Internet Movie Database (IMDb) dataset comprising of movie reviews. The results show that the Opinion Analysis model gives enhanced classification accuracy for both M-learning dataset and IMDb dataset.

Table 1. Comparisons of Results with preprocessing

Classifiers	Classification Accuracy without Proposed Preprocessing Algorithm (%)	Classification Accuracy With Proposed Preprocessing Algorithm (%)
Naïve Bayes	53.67	55
KNN	49.33	52.67
Random Forest	57.66	60
LVQ	54.33	56.67
Elman Neural Network	72.33	76.67
FFNN	77	78.67

Table 1 shows the classification accuracy of proposed preprocessing algorithm with various classification algorithms. The result shows that the proposed preprocessing algorithm gives better enhancement in the classification accuracy.
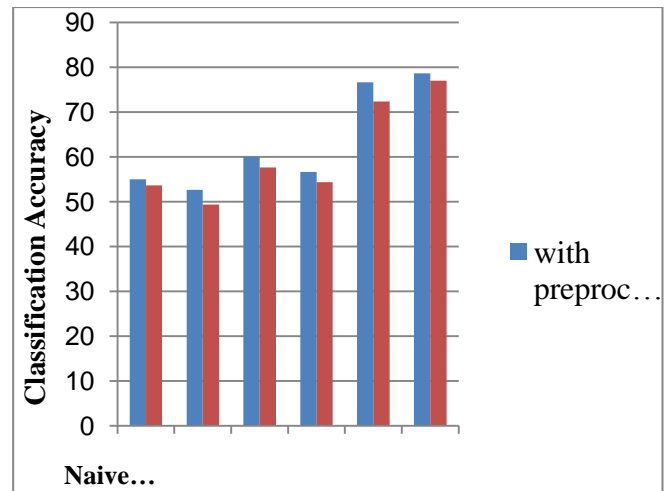

Figure 2: Comparisons of Results

Figure 2 shows the graphical representation of the classification accuracy of M- Learning dataset with proposed preprocessing and without proposed preprocessing algorithm. From this result the classification accuracy is enhanced with proposed preprocessing algorithm with existing classifiers.

Table 2. Comparisons with IMDb and M-Learning Datasets

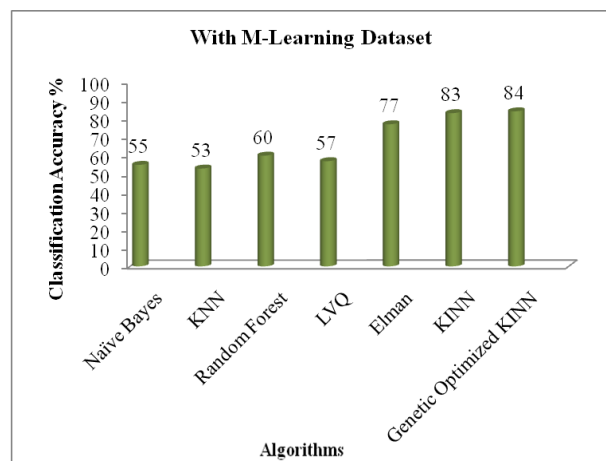| Algorithms	Classification Accuracy % M-Learning	Classification Accuracy % IMDb
Naive Bayes	55	55
kNN	53	53
Random Forest	60	68
LVQ	57	72
Elman	77	88
Proposed KINN	83	91
Proposed GKINN	84	93

Figure 3: Results with M-Learning Dataset

From the results, it can be shown that the research explorations with the benchmark dataset have an improvement over M-Learning dataset. Figure 3 shows the graphical representation of the classification accuracy with M- Learning dataset.

The comparison of results with various existing and proposed classification algorithms for IMDb dataset is presented in Figure 4.
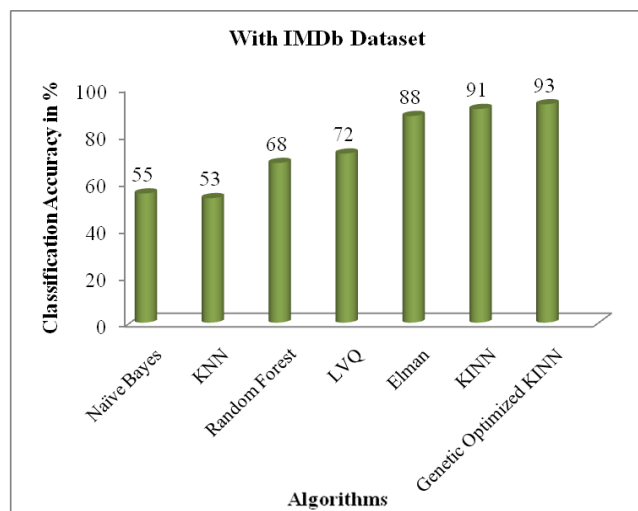

Figure 4: Results with IMDb Dataset

From the obtained results, it is concluded that the proposed algorithms perform better in classification of the opinions based on their subjectivity.

4. CONCLUSION

In the process of identifying subjective information such as opinions, and sentiments, from the human generated text, mining has become a tedious and difficult effort. Opinion extraction and the classification algorithm play important role to obtain the classification accuracy. The amplified process such as preprocessing is used to enrich the data sets. Thus, enhances the classification accuracy with M-Learning dataset. Neural networks have high acceptance ability for noisy data, high accuracy and are preferable in data mining. In Knowledge Discovery in Databases (KDD), Neural Networks are employed in classification process. The proposed KINN based algorithm enhances the classification accuracy. Evolutionary algorithm namely GKINN is incorporated to enhance the classification accuracy of the M-Learning.

The proposed Opinion Analysis Model involving KN preprocessing algorithm, KINN and GKINN algorithms enhances the classification accuracy. The results clearly shows that the proposed opinion analysis model enhances the classification accuracy of M-Learning dataset and also applicable to all the datasets.

5. REFERENCES

[1] Chenghua Lin, Yulan He, Richard Everson, Stefan Ruger, "Weakly Supervised Joint Sentiment-Topic Detection from Text", IEEE Transactions On Knowledge and Data Engineering, Volume 24, Number 6, pp. 1134-1144, 2012.

[2] Chih-Ming Chen, Shih-Hsun Hsu, "Personalized Intelligent Mobile Learning System for Supporting Effective English Learning", International Forum of Educational Technology & Society, Volume 11, Issue 3, pp. 153-180, 2008.

[3] Traxler, J. "Current State of Mobile Learning", Mobile Learning: Transforming the Delivery of Education and Training. Edmonton: Athabasca University Press, pp. 9-24, 2009.

[4] Kukulska-Hulme, A., "Practitioners as Innovators: Emergent Practice in Personal Mobile Teaching, Learning, Work,

and Leisure", in Ally, M. (ed.) Mobile Learning: Transforming the Delivery of Education and Training. Edmonton: Athabasca University Press, pp. 135-155, 2009.

[5] Bo Pang, Lillian Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, Volume 2, Number 1-2, pp. 1–135, 2008.

[6] S.N. Sivanandam, S.N. Deepa, "Principles of Soft computing", Wiley India(P) Ltd., 2007.

[7] Weiyang Zhou, "Verification of the Nonparametric Characteristics of Backpropagation Neural Networks for Image Classification", IEEE Transactions on Geoscience and Remote Sensing, Volume 37, Number 2, 1999.

[8] D.G. Rees, "Essential statistics", 4th edition, Chapman and Hall/CRC. ISBN 1-58488-007-4 (section 9.5), 2001

[9] Karina Gibert, Joaquín Izquierdo, Geoff Holmes, Ioannis Athanasiadis, Joaquim Comas, Miquel Sànchez-Marrè, "On the role of pre and post-processing in environmental data mining", International Congress on Environmental Modeling and Software Integrating Sciences and Information Technology for Environmental Assessment and Decision Making, pp. 1937-1958, 2008.

[10] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417-424, 2002.

[11] A. Rangel-Merino, J. L. Lopez-Bonilla and R. Linares Y. Miranda, Optimization method based on genetic algorithms", Apeiron, Volume 12, pp. 393-408, 2005,

[12] Marylyn D Ritchie, Bill C White, Joel S Parker, Lance W Hahn and Jason H Moore, "Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases", BMC Bioinformatics, Volume 4, Number 28, pp.1-14, 2003

[13] Kyoung - Jae Kima, Ingoo Hanb, "Application of a hybrid genetic algorithm and neural network approach in activity-based costing", Expert Systems with Applications, Volume 24, pp. 73–77, 2003

[14] Gwo-Jen Hwang, Peng-Yeng Yin, Tzu-Ting Wang, Judy C.R. Tseng, Gwo-Haur Hwang "An enhanced genetic approach to optimizing auto-reply accuracy of an e-learning system" Computers and Education, Volume 51, pp. 337-353, 2008,

[15] Mohamed Ettaouil, Mohamed Lazaar, Youssef Ghanou, "Architecture Optimization Model For The Multilayer Perceptron And Clustering", Journal of Theoretical and Applied Information Technology, Volume 47, Number 1, pp. 64-72,2013,

[16] Fadl Mutaher Ba-Alwi, "Knowledge Acquisition Tool for Classification Rules using Genetic Algorithm Approach" International Journal of Computer Applications, Volume 60, Number 1, pp. 29-35, 2012,

[17] Guoqiang Peter Zhang, "Neural Networks for Classification: A Survey", IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications and Reviews, Volume 30 -No. 4, pp. 451-462, 2000.

[18] Jack G. Conrad, Frank Schilder, "Opinion Mining in Legal Blogs", Proceedings of the 11th International Conference on Artificial Intelligence and Law, ICAIL, pp.231-236, 2007