



A Study on Various Algorithms to Enhance Classification Accuracy in M-Learning

Dr. Nisha Jebaseeli

Ph.D, Department of Computer Science,
Bharathidasan University Constituent College,
Lalgudi, Trichy, India

Abstract— *This research paper aims to study about the accuracy of categorization by enhancing 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. This research paper focuses on extraction and categorization of opinions from the web-documents. As the opinions have many potential applications, mining on the reviews of the M-Learning users are contemplated. Artificial Neural Networks (ANNs) are one of the promises 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. Generating the solutions for optimization, the techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover are implemented. KN Preprocessing is utilized in the present study. The datasets are acquired from the reviews on the M-Learning system and they are preprocessed with the existing opinion mining preprocessors 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 main aim in this research is to develop a model for extracting opinions efficiently that represents evaluation of consumer products, in a structured form and classify the scattered opinions. Subsequently, the classification accuracy is enhanced using novel methods to complement the analysis.*

Keywords— *categorization of opinions, Artificial Neural Networks, optimization, KN Preprocessing, M-Learning system.*

I. INTRODUCTION

Mining different kinds of knowledge from opinion-driven sources is interestingly a huge need and growing demand^[1]. For instance, organizations wish to know relevant and qualitative opinions towards products or services rendered, individuals wish to retrieve relevant information for personal use, and the need for understanding human behaviours through trends at real-time is also becoming more apparent and necessary. The ability to classify the large number of reviews from the Web sources is valuable to both academic research and practical applications^[2]. In this section, the outcomes of the present research study are expounded. The Opinion Mining is an area of interest that is applied in varied domains to extract knowledge from subjective and human generated opinions^[3]. This section discusses some of the areas identified for the application of the proposed algorithms and opinion analysis model. In the areas of education and training, the use of M-learning is becoming inevitable and efficient in reaching its objectives. The use of various M-Learning materials and products available on-line could be instructed based on the views of the users^[4]. Thus, to enhance the use of such resources effectively the proposed algorithms and model could be implemented. Opinion mining is an emerging area of research which summarizes the customer reviews of a product, service and Learning system. It expresses whether the opinions are positive or negative^[5]. The main objective of Opinion mining is to determine the polarity of comments (positive, negative or neutral) by extracting attributes and components of the object that have been commented on in each document.

The businesses and organizations in general spend a lump sum of the investments in finding the consumer sentiments and opinions^[6]. Hence, the mining on the opinions and sentiments are made for decision making process. Especially, the explosive growth of the social media has intuited the business and organizations to use the content for extracting and summarizing the information in them. When purchasing a product or using a service, every individual could scrutinize based on the opinions of those who have bought^[7]. The opinions of a population on Politics and political issues could be mined for valuable information. In the field of analysing the subjectivity of the sentiments/opinions, advanced preprocessing techniques involving the removal of superfluous and malapropos data in the corpus will enhance the information retrieval^[8]. Feature-based opinion summary will complement the mining of opinions as a whole. Dictionary-based approach and corpus-based approach shall improve the mining of the reviews^[9]. Self-Organizing Feature map that

considers training and mapping which automatically classifies the new input vector could be designed for better classification with neural networks^[10]. Factors other than learning rate and momentum could be genetically optimized for speedy convergence.

II. OBJECTIVES OF THE OVERALL RESEARCH

- To study the preprocessing algorithm with word importance, term document frequency and filtering to enhance the classification accuracy.
- To compare the accuracy rates between existing preprocessing techniques and the proposed technique with various classification algorithms.
- To study KINN algorithm to enhance the classification accuracy.
- To analogize the proposed Classifier with the existing classification algorithms.
- To study Genetic Optimized GKINN algorithm to improve the classification accuracy.
- To optimize Learning rate and Momentum values with genetic approach.
- To study the Opinion Analysis Model to enhance polarization of the documents.

III. RESEARCH METHODOLOGY

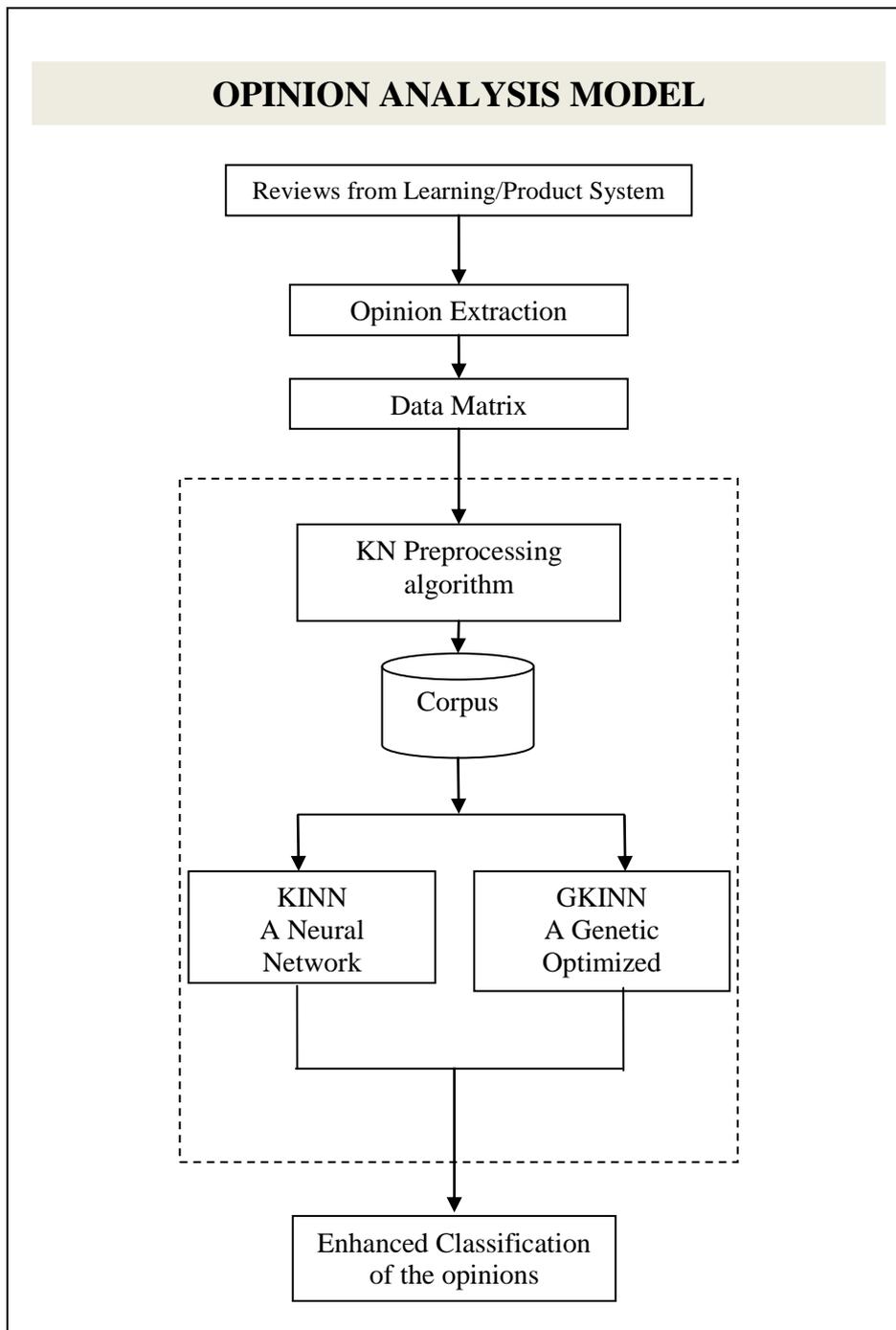


Figure 1 Opinion Analysis Model

III. KN PREPROCESSING

An In the formulation of corpus for the classification and further mining analysis, KN Preprocessing is utilized in the present study. The datasets were acquired from the reviews on the M-Learning system and they were preprocessed with the existing opinion mining preprocessors namely Stopword Removal, Stemming, Inverse Document Frequency, and Singular Value Decomposition^[11]. 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^[12]. The results proved that the classification accuracy enhanced with the use of the proposed preprocessing technique.

IV. KINN - A NEURAL NETWORK BASED ALGORITHM

All Biologically inspired classification algorithms are employed for better categorization of the reviews on the M-Learning system^[13]. Data enters at the inputs and passes through the network, layer by layer, until it arrives at the outputs. In the neural network operation, that is when it acts as a classifier, there is feedback between layers. The Sigmoidal summation functions are used in the layers^[14]. With the help of Information Gain, top five positive and negative values are selected for the subsequent dispensation into the feedback process. The convergence of the output is authenticated when the threshold value or the number of epochs is reached in the process of classification^[15].

VI. GKINN - A GENETIC OPTIMEIZED ALGORITHM

Genetic Optimization is the evolutionary approach which aims to improve the solution to a problem by keeping the best combination of input variables. Selection, Mutation and Crossover operators were iteratively fed forward to the input layers. The process after the summation and activation is augmented by genetically optimizing the learning rate and momentum^[16]. Through the process of calculating fitness value, computing and selecting the offspring count, the crossover and mutation are applied in the process of generation of optimized values. Thus, with the amplified process the classification accuracy is found to be enhanced with M-Learning dataset^[17].

VII. OPIONION ANALYSIS MODEL

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^[18]. The proposed opinion Analysis Model (OAM), the proposed model comprises various techniques for classifying the user generated content. Opinion Mining is the computational study of people's opinions, appraisals, attitudes and emotions toward entities, individuals, events, topics and their attributes. It is a technically challenging and practically useful analysis^[19]. Discovering the polarity or the subjectivity of the opinions is the prime effort of the analysis. Among various sentiment analysis tasks, one of them is sentiment classification, i.e., identifying whether the semantic orientation of the given text is positive, negative, or neutral. Sentiment analysis or opinion mining aims to use automated tools to detect subjective information such as opinions, attitudes, and feelings expressed in text^[20].

VIII. LIMITATIONS OF THE RESEARCH

Although the research has reached its aims, there were some unavoidable limitations. A feature-based opinion summary of multiple reviews is not aforesought; instead sentence and document level classifications are made if they are opinionated. There exist various approaches to compile sentiment words, namely dictionary-based approach, corpus-based approach which is less considered in the research. Feed Forward neural network approach alone is thought-about, though other approaches like Self-Organizing Feature maps of ANN that exist to enhance the classification. The analyses of the opinions are based on supervised learning. In the research, Learning rate and momentum are the factors that alone are genetically optimized.

IX. CONCLUSION

This research work is concentrated on enhancing the classification accuracy. Hence, three algorithms to enhance classification accuracy are analyzed. The clean data set is achieved through the proposed preprocessing algorithm for enhancing the classification accuracy. The proposed KINN and genetic optimized GKINN algorithms classify the M-Learning system opinions as positive, negative or neutral with enhanced classification accuracy. Finally, the Opinion Analysis Model helps to enhance the classification accuracy. The experimental results reveal that the proposed algorithms enhance the classification accuracy. The three proposed algorithms are integrated to form a model. The outcomes of the model reveal that the classification accuracy is enhanced.

REFERENCES

- [1] Alexandra Balahur, Andrés Montoyo, OpAL: Applying opinion mining techniques for the disambiguation of sentiment ambiguous adjectives in SemEval-2 task 18", Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 444-447, 2010.
- [2] R. Alejo, V. Garcia, J.M. Sotoca, R.A. Mollineda, J.S. Sánchez, "Improving the Classification Accuracy of RBF and MLP Neural Networks Trained with Imbalanced Samples", Proceedings of the 7th international conference on Intelligent Data Engineering and Automated Learning, pp. 464-471, 2006.

- [3] Andrew McCallum, Kamal Nigam, "A Comparison of Event Models for Naive Bayes Text Classification", In AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41-48, 1998.
- [4] Bekim Fetaji, Majlinda Fetaji, "Software Engineering Mobile Learning Software Solution Using Task Based Learning Approach", World Academy of Science, Engineering and Technology, Volume 6, Issue 4, pp. 395-399, 2009.
- [5] Bing Liu, "Exploring User Opinions in Recommender Systems", Proceedings of the second KDD workshop on Large Scale Recommender Systems and the Netflix Prize Competition, USA, Article No.1, August 2008.
- [6] Debotosh Bhattacharjee, Ayan Seal, Suranjan Ganguly, Mita Nasipuri, Dipak Kumar Basu, "A Comparative Study of Human Thermal Face Recognition Based on Haar Wavelet Transform and Local Binary Pattern", Computational Intelligence and Neuroscience, Volume 12, pp. 1-12, 2012.
- [7] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, "Indexing by latent semantic analysis", Journal of the American Society for Information Sciences, Volume 41, pp. 391-407, 1990.
- [8] Doug Muchoney, James Williamson, "A Gaussian Adaptive Resonance Theory Neural Network Classification Algorithm Applied to Supervised Land Cover Mapping Using Multi temporal Vegetation Index Data", IEEE Transactions On Geoscience And Remote Sensing, Volume 39, Number 9, pp. 1969-1977, 2001
- [9] J.L. Elman, "Finding structure in time", Cognitive Science, Volume 14, Number 2, pp. 179-211, 1990.
- [10] Faisal Kamiran, Toon Calders, "Data preprocessing techniques for classification without discrimination", Knowledge Information System, Springer, Volume 33, pp.1-33, 2011.
- [11] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, "From Data Mining to Knowledge Discovery in Databases", Artificial Intelligence Magazine, Volume 17, Issue 3, pp. 37-54, 1996.
- [12] Y. Freund, R. Schapire, "Experiments with a New Boosting Algorithm", Machine Learning: Proceedings of the Thirteenth International Conference, pp. 148-156, 1996.
- [13] K. Gnana Sheela, Dr. S.N. Deepa, "Analysis of computing algorithm using momentum in neural networks", Journal of Computing, Volume 3, Issue 6, pp. 163-166, 2011.
- [14] R.M. Gray, "Vector quantization", IEEE Acoustic, Speech, and Signal Processing Magazine, volume 1, pp. 4-29, 1984.
- [15] Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen, "Opinion Word Expansion and Target Extraction through Double Propagation", Computational Linguistics, Volume 37, Number 1, pp. 9-27, 2011.
- [16] Guoqiang Peter Zhang, "Neural Networks for Classification: A Survey", IEEE Transactions On Systems, Man, And Cybernetics-Part C: Applications and Reviews, Volume 30, Number 4, pp. 451-462, 2000.
- [17] L. Gupta, M. McAvoy, J. Phegley, "Classification of Temporal Sequences via Prediction using the Simple Recurrent Neural Network", Pattern Recognition, Volume 33, Number 10, pp. 1759-1770, 2000.
- [18] M.H. Hassoun, "Fundamentals of Artificial Neural Networks", Massachusetts Institute of Technology Press, London, 1995.
- [19] Heikki Maaranen, KaisaMiettinen, Antti Penttinen, "On initial populations of a genetic algorithm for continuous optimization problems", Springer Science, Volume 37, pp. 405-436, 2006.
- [20] Inaki Inza, Borja Calvo, Ruben Armananzas, Endika Bengoetxea, Pedro Larranaga, Jose A. Lozano, "Machine Learning: An Indispensable Tool in Bioinformatics", Springer Science Business Media, Volume 593, pp. 25-48, 2010.