

Effective and Efficient framework for Multimodal Summarization in neural network using optimized genetic algorithm

D. GOUSYA BEGUM, M.Tech (Ph.D)

Assistant Professor, Dept. of CSE
S.K .U College of Engineering & Technology
S.K. University, Anantapuramu, A.P

U. DHANUNJAYA, M.Tech (Ph.D)

Assistant Professor, Dept. of CSE
S.K .U College of Engineering & Technology
S.K. University, Anantapuramu, A.P

ABSTRACT:

A Data mining has become one of the widely researched fields off late especially with the increase in technological advancements which have caused the volume of data to be stored and processed for query based applications or decision based processes to increase in great multiplicative measures. Data mining refers to the process of extraction of useful information from a pool of data. Various algorithms have been proposed in the past for the mining process out of which neural based mining algorithms are predominant. Genetic algorithm is an important algorithm of association rule mining. However, there is some issues that genetic algorithm easy to lead premastering convergence and into the plight of local optimum, or convergence too much time and consume a large amount of time to search

For resolving this issue, the paper improves the algorithm through adopting an adaptive mutation rate and improving the methods of individual choice, and the improved genetic algorithm that applies to the mining association rules. The simulating experiments show that the improved genetic algorithm reduces the cost of computing, and improve the efficiency of association rule mining. The proposed algorithm has been tested with iris data set and results obtain indicate a comparatively higher efficiency based on classification and reduction of computation time. The proposed results have been compared against conventional techniques like SVM classifier based mining and neural network in its standalone architecture.

Keywords: Data mining, neural network, genetic optimization, knowledge discovery database, Association Rule

I. INTRODUCTION

A Data mining is the non-trivial process that automatically collects those useful hidden information from the data muster, and is taken on as forms of rule, concept, rule and pattern and so on. It is useful for decision-makers owing to the virtues of analyzing historical and current data, discovering hidden relation and pattern, predicting possible behaviors which may occur in the future. The process of data mining is also called as process of knowledge discover, which is a new interdisciplinary referring to such wide scope of subjects as databases, artificial intelligence, statistics, visualization, parallel computing and other fields. Data mining is one of research areas of the international database and decision-making information, of which the main purpose is to develop the relevant methods, theory and tools and to extract useful and interesting knowledge from the large amounts of data. Association rules mining is an important research area in data mining, which indicates relations among item sets in database. With the accumulation of the data, association rule mining in large data set attracts more and more attention. Apriori algorithm is a classical algorithm of association rule mining.

Lots of algorithms for mining association rules and their mutations are proposed on basis of Apriori algorithm, but apriori algorithm need to traverse the database many times, I/O overhead, and computational complexity, cannot meet the requirements of large-scale database mining. Genetic algorithm is an algorithm which based on the biological theory of evolution and molecular genetics of the global random search, the algorithm has a strong randomness, robust and implicit parallelism can be quickly and effectively search for global optimization, is an effective way to deal with large-scale data sets. At present, genetic algorithm-based data mining methods have yielded some progress, and based on genetic algorithms classification system has also yielded some results. Data mining (DM) often referred as knowledge discovery in databases (KDD), is a process of nontrivial extraction of implicit, previously unknown and potentially useful information from a large volume of data.

II. MOTIVATION

The extracted information is also referred as knowledge of the form rules, constraints and regularities. Rule mining is one of critical tasks as they provide a concise statement of potentially important information and most research contributions in the past till data have utilized neural network based techniques for deriving the mining rules. Researchers have been using many techniques such as statistical, AI, decision tree, database, cognitive etc. for rule mining. Several major kinds of data mining methods, including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization, and meta-rule guided mining.

A review on the types of mining techniques have been presented in and classified into association technique based, clustering technique based, prediction based, classification based, decision tree and pattern based techniques. The prediction is one of a data mining techniques that determine relationship between dependent and independent variables. The prediction analysis technique can be used in sale to predict profit, sale is an independent variable; profit could be a dependent variable. Then based on the past sale and profit data, a regression curve that is used for profit prediction. From the literature, it has been observed that no approaches or tools can guarantee to generate the accurate prediction in the organization. In this paper, they have analyzed the differential algorithm and prediction technique.

In spite the fact that the least median squares regression is known to produce better results than the classifier linear regression techniques from the given set of attributes. As comparison they found that Linear Regression technique which takes the lesser time as compared to Least Median Square Regression. Sequential patterns analysis is one of data mining techniques that seeks to discover or identify related patterns, regular events or trends in transaction data over a business period. Rule mining using neural networks (NNs) is a challenging job as there is no straight way to translate NN weights to rules. However, NNs have potential to be used in rule mining since they have been found to be a powerful tool to efficiently model data and modeling data is also an essential part of rule mining. As one of branches of DM methods, rule mining aims to apply algorithms of DM to stored data in databases.

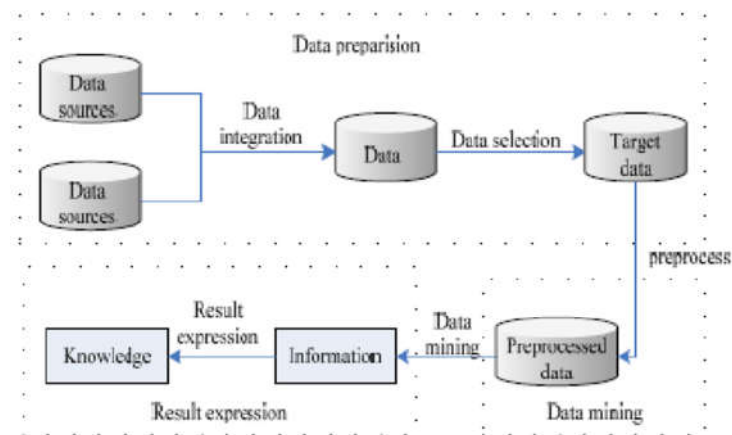


Figure 1: A general data mining system

III. PROPOSED ALGORITHM

The program of premature is a phenomenon that can not be ignored in genetic algorithm, and A concrete manifestation of it as follows:

- 1) All of the individual in groups stopped the evolution at the same extreme value;
- 2) The individual of closing to the optimal solution always be eliminated, not convergent in the process of evolution.

In order to solve these two problems: i) A method of adaptiving mutation rate to avoid excessive variation causing non-convergence, or into a local optimal solution;
ii) A sort of individual-based selection method, it will be applied to the evolution of the latter in genetic algorithm, in order to prevent the high-fitness individuals converging early by the rapid growth of the number of individual and differences is too small.

Adaptive Mutation Rate :

In the early stages of genetic algorithm evolution, such as the use of a fixed mutation probability P_m , when P_m values is small, mutation operator does not have an impact on the group and conducive to introduce a new genes; when P_m value is great, it could undermine the group's excellent genes, the algorithm does not even slow down the convergence. Here, a method of adaptive mutation rate, in the early stages of evolution and mutation rate are used as follows:

$$P_m^{(n+1)} = \lambda \cdot P_m^0 \sqrt{\frac{\sum_{i=1}^m (f_{\max}^{(n+1)} - f_i^{(n)})^2}{\sum_{i=1}^m (f_{\max}^{(n)} - f_i^{(n)})^2}}$$

Improved Genetic Algorithm for Mining Association Rules:

Genetic Algorithm Encoding is the most basic question in the genetic algorithm. A good encoding method will make crossover and mutation genetic manipulation easier to achieve. When we found an association rule such as $A_1 \ A_2 \ \dots \ A_n \Rightarrow B_1 \ B_2 \ \dots \ B_n$. In fact, is found such a rule in the database : $A_1 \ A_2 \ \dots \ A_n$ and $B_1 \ B_2 \ \dots \ B_n$ are true at the same time when the percentage of the entire affairs is greater than the support given by users .and $B_1 \ B_2 \ \dots \ B_n$ is true in the conditions $A_1 \ A_2 \ \dots \ A_n$ is true, and the probability of confidence greater than the confidence given by user. ($A_1, A_2, \dots, A_n; B_1, B_2, \dots, B_n$) is a database field in affairs.

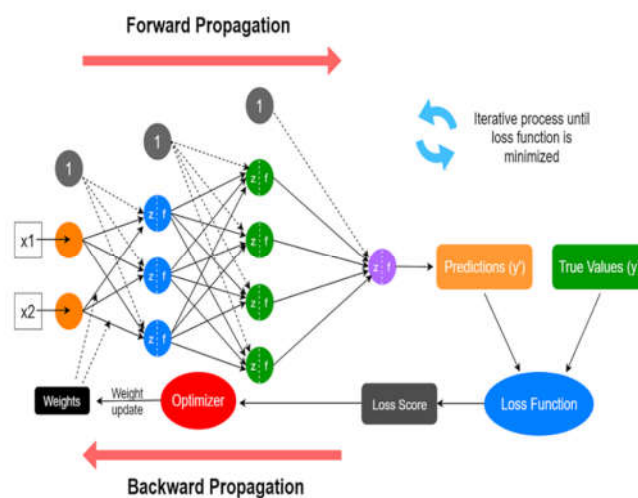


Figure 2: A multi layered artificial neural network

Every NN model must be trained with representative data before using. There are basically two types of training, supervised and unsupervised. The basic idea behind training is to pick up set of weights (often randomly), apply the inputs to the NN and check the output with the assigned weights. The computed result is compared to the actual value. The difference is used to update the weights of each layer using the generalized delta rule [6, 10]. This training algorithm is known as ‘back propagation’. After several training epochs, when the error between the actual output and the computed output is less than a previously specified value, the NN is considered trained. Once trained, the NN can be used to process new data, classifying them according to its required knowledge. When using supervised training it is important to address the following practical issues.

Artificial Neural Network

In essence, an artificial neuron is a biological neuron made through engineering. It is a gadget with a single output and several inputs. An artificial neural network (ANN) is made up of numerous simple processing components that are coupled to one another. Neural networks are used in the human body to accomplish tasks. A neural network is merely a network of millions and millions of interconnected neurons. The human body is the best example of parallel processing since it uses these interconnected neurons to perform all of its parallel processing. A neuron is a unique type of biological cell that uses electrical and chemical changes to transfer information from one neuron to another.

An artificial neuron is essentially a biological neuron that has been engineered. This device has many inputs and a single output.

The many basic processing elements that make up an artificial neural network (ANN) are connected to each other. The human body uses neural networks to do out tasks. Millions and millions of linked neurons make up a brain network. Since all of the parallel processing in the human body is carried out by these interconnected neurons, the body is the best illustration of parallel processing. A neuron is a special kind of biological cell that transmits information from one neuron to another through electrical and chemical changes. A multi-layered artificial neural network is a prevalent architecture for supervised learning, specifically the feed-forward neural network. These networks are structured with layers of nodes, each connected to nodes in the adjacent layers through weighted connections. Training of feed-forward networks typically employs a back propagation learning method, which adjusts the weight values starting from the output layer and then progresses backward through the hidden layers of the network.

Input: Data set Target: Classified set

Generate $m \times n$ map with a seed neurons Initialize initial weight $W(0)$

Select an instance n Find the winning neuron Determine the error

Adapt the weight vectors of the k neuron using genetic optimization after encoding

Repeat steps until convergence

Add or delete connections between neurons according to the measure distances

GA Optimization:

The GA is basically based on biological principle of natural selection. The architecture of systems that implement GAs is able to adapt to a wide range of problems. A GA functions by generating a large set of possible solutions to a given problem. It then evaluates each of those solutions, and decides on a fitness level for each solution set. These solutions then breed new solutions. The parent solutions that are more fit are more likely to reproduce, while those that are less fit are more unlikely to do so. The elementary operation of genetic algorithm consists of three operands: selection, crossover and mutation. Select is also called copy or reproduction. By calculating the fitness f_i of individuals, we select high quality individuals with high fitness, copy them to the new population and eliminate the individual with low fitness to generate the new population

IV. RESULTS

The set contains three classes of fifty instances each, where every class refers to a type of Iris plant. 4 attributes are used to predict the iris class, *i.e.*, sepal length (T_1), sepal width (T_2), petal length (T_3), and petal width (T_4), all in centimeters. Among the 3 classes, class one is linearly separable from the other two classes, and classes two and three are not linearly separable from one another. To ease data extraction, we reformulate the data with three outputs, where class 1 is represented by $\{1, 0, 0\}$, class 2 by $\{0, 1, 0\}$, and class 3 by $\{0, 0, 1\}$. One class is linearly separable from the other 2; however, the latter are not linearly separable from each other. This data set is prepared by Ronald Fisher. In all four features of 150 irises samples of three classes/types are recorded. Class 1 is Setosa; Class 2 is Verginica; and class 3 is Versicolor.

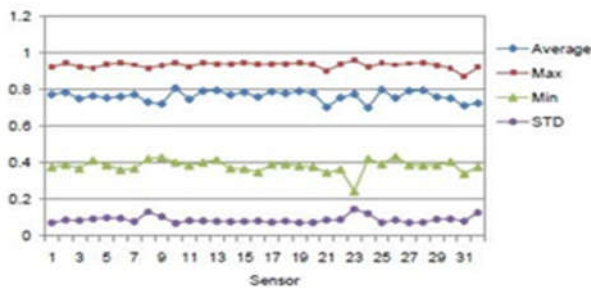


Figure 3: Extracted features for the dataset

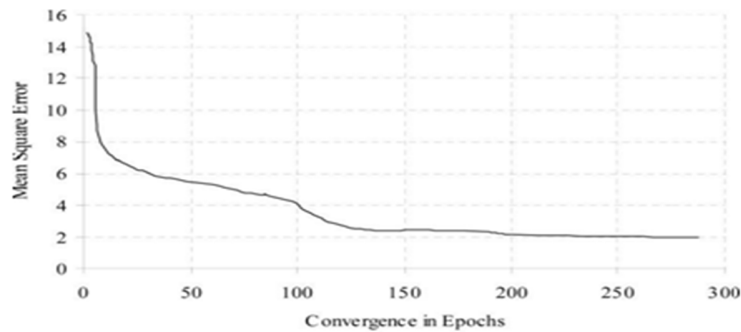


Figure 4: convergence of the mean square error

The convergence of the mean square error against the number of iterations or epochs is depicted in figure 4. The mean square error for the given data set converges to a minimum value in 108 iterations and maintains a near or less saturation.

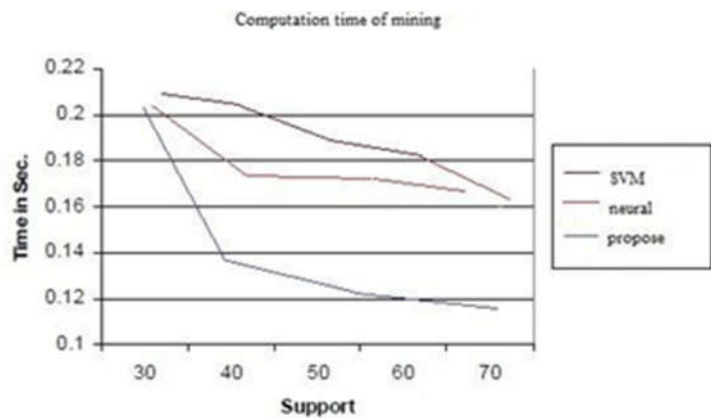


Figure 5: Comparison of computation times

IV. CONCLUSION

A research on data mining based on neural network optimized through genetic algorithm is presented in this paper. Data mining is known for its high robustness, self organizing adaptive, parallel processing capabilities, and distributed storage with a high degree of fault tolerance. The combination of data mining and neural network can greatly improve the efficiency of data mining, and it has been widely used & we have presented neural network based data mining scheme to mining classification rules from given databases. This work is an attempt to apply the approach to data mining by extracting symbolic rules. An important feature of the rule extraction algorithm is its recursive nature. A set of experiments was conducted to test the approach using a well define set of data mining problems. The results indicate that, using the approach, high quality rules can be discovered from the given datasets. The extracted rules are concise, comprehensible, order insensitive, and do not involve any weight values. The accuracy of the rules from the pruned network is as high as the accuracy of the fully connected networks. Experiments showed that this method helped a lot to reduce the number of rules significantly without sacrificing classification accuracy.

REFERENCES

- [1]. Meenakshi Sharma, "Data Mining: A Literature Survey", International Journal of emerging research in management and technology, Vo. 3, Issue. 2, 2014.
- [2]. Ranno Agarwal, " Genetic algorithms in data mining", Internaitonal Journal of advanced research in computer science and software engineering, Vol. 5, Issue. 9, 2015.
- [3]. Suguna and Nandhini, "Literature review on data mining techniques", International journal of computer technology and applications, Vol. 6, No. 4, pp. 583 – 585, 2015.
- [4]. T. Karthikeyan and N. Ravikumar, A Survey on Association Rule Mining International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 1, 2014.
- [5]. Maruthaveni.R, Mrs. Renuka Devi.S.V, "Efficient Data Mining For Mining Classification Using Neural Network", International Journal Of Engineering And Computer Science, Volume. 3, Issue. 2, 2014
- [6]. Diti Gupta, Abhishek Singh Chauhan, "Mining Association Rules from Infrequent Item sets: A Survey", International Journal of Innovative Research in Science", Engineering and Technology, Vol.2, Issue 10, 2013.
- [7]. J. Chang, C. Gao, Y. Zheng, Y. Hui, Y. Niu, Y. Song, D. Jin, and Y. Li, "Sequential recommendation with graph neural networks," in SIGIR, 2021, pp. 378–387.
- [8]. S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," Proc. AAAI Conf. Artif. Intell., vol. 33, no. 01, pp. 346–353, 2019.
- [9]. Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan et al, "Training a helpful and harmless assistant with reinforcement learning from human feedback," arXiv preprint arXiv:2204.05862, 2022..
- [10]. D. Ghosh, J. Rahme, A. Kumar, A. Zhang, R. P. Adams, and S. Levine, "Why generalization in rl is difficult: Epistemic pomdps and implicit partial observability," *Advances in Neural Information Processing Systems*, vol. 34, pp. 25502–25515, 2021..
- [11]. Sapna B Kulkarni, Yuvaraju BN, "The Top-N rule selection approach algorithm to split the multimedia traffic stream into multiple sub-streams prior to transmission in MANETS", IPASJ International Journal of Computer Science (IJCS), Volume 3, Issue 1, January 2015