International Journal of Soft Computing 5 (6): 211-213, 2010
ISSN: 1816-9503

BSclassifier for Balance Scale Weight and Distance Database

Ghanshyam Singh Thakur and Rekha Singh Thakur
Department of Computer Applications,
Maulana Azad National Institute of Technology, 462051 Bhopal (M.P.), India

Abstract: In this study researchers have proposed a new Balance Scale Classifier (BSclassifier) for Balance Scale Weight and Distance Database. The correctness of the classifier has been analyzed using real datasets. The proposed classification algorithm is classified the test data set as having the balance scale tip to the right, tip to the left or be balanced. The proposed model has been trained using large number of training data sets and performance evaluation done using real and synthetic datasets.

Key words: Data mining, classification, Bsclassifier, synthetic datasets, performance, India

INTRODUCTION

Today's due to globalization of the world the size of data set is increasing, it is necessary to discover the knowledge. Knowledge mining (Fayyad *et al*., 1996; Pujari, 2000; Han and Kamber, 2001) is non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data that means data mining extract meaningful knowledge from huge chunk of raw data. The discovery of knowledge can be typically in the form of association rules, classification rules, clustering, discovery of frequent episodes and deviation detection. Other mining problems are sequence mining, web mining, text mining, spatial and temporal data mining. An important task of data mining is to assign objects to predefined categories or classes a process called Classification (Pujari, 2000; Han and Kamber, 2001). The input to the classification system consists of a set of example records called a training set, over several fields or attributes. Attributes are continuous, coming from an ordered domain or categorical, coming from an unordered domain. One of the attributes called the classifying attribute in dicates the class or label to which each example belongs. The goal of classification is to induce a model from the training set that can be used to predict the class of a new record. Classification has applications in diverse fields such as retail target marketing, customer retention, fraud detection and medical diagnosis.

THE RESEARCH ALREADY DONE IN THE FIELD

Many models or algorithms have been adopted for classification. Classification has been studied intensively because of its wide applicability in areas such as data mining in formation retrieval etc. The majority of this information is in various forms like emails, news, web pages, reports, etc. Organizing and classifying them into a logical structure is a challenging task. The classification techniques (Fayyad *et al*., 1996; Quinlan, 1986; Yang *et al*., 2001; Carvalho and Freitas, 2000; Pujari, 2000; Han and Kamber, 2001) such as Support Vector Machine (SVM), perception with margin, k Nearest Neighbor (k-NN), decision Ada Boost, Logistic regression, neural network and Bayesian network, k-means, Naive Bayes and Decision Tree aremost popular. All these approaches are suffered from lack of high performance and high accuracy.

In addition, many existing classification algorithms require the user to specify the number of category as an input parameter. Incorrect estimation of the value always leads to poor classification. This variation tremendously reduces the resulting classification accuracy for some of the state-of-the art algorithms. But there are still problems to be tackled such as efficiency and accuracy. Owing to wide significant applicability of Balance Scale Weight and Distance Database classification and its challenges motivated to do research in this field. The poor classification accuracy and the weaknesses of the standard classification methods formulate the goal of this research. In this study an efficient classification algorithm is proposed.

PROPOSED CLASSIFIER

In this study the new classification algorithm have proposed for classifying Balances into various classes. The proposed BSclassifier is designed which based on a mathematical formulae where it classify balances into three classes i.e., Right, Left and Balance. The proposed BSclassifier is shown in Fig. 1.

Corresponding Author: Ghanshyam Singh Thakur, Department of Computer Applications,
Maulana Azad National Institute of Technology, 462051 Bhopal (M.P.), India

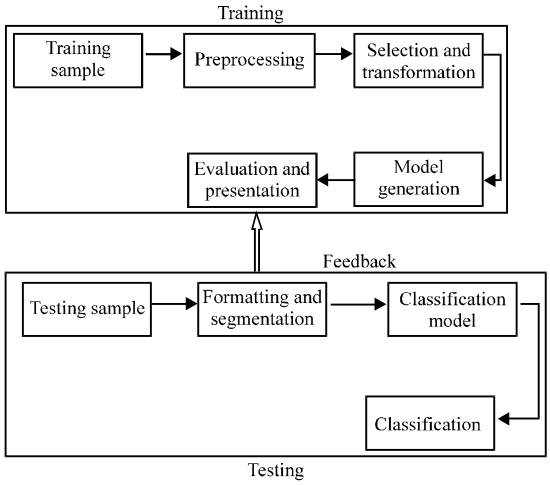

Fig. 1: Architecture of classification process

The proposed BSclassifier framework consists of four steps during the first step the data will be collected from various sources. The data will contain information about balance i.e., left weight, left distance, right weight and right distance. This balance information can be collected by measuring balances on the spot or ask from each shopkeeper submit their balances information in required format. Once data collection is over the system will move data preprocessing towards. The data gathered from various sources may contain noise and incorrect information. So during preprocessing noise will be removed from data and only essential information will be abstracted from raw data. After completion selection of process all essential information transferred at one source to make a form data warehouse or data cube, under unified schema. This Balance Scale data cube contains data into two dimension one is balance identification and another is balance attributes.

The proposed BSclassifier will get input from balance Scale data cube and classifier classify input data into three class i.e., class Right, class Left and class Balance. This classification works based on one mathematical formula which is shown in Fig. 2. The last phase is the result analysis and presentation. In This phase the system will present knowledge according to end user requirement.

Algorithm

Input: Left weight, left distance, right weight and right distance:

Output: Class L or B or R.

- Lc = left-distance * left-weight
- Rc = right-distance * right-weight
- If (Lc>Rc) then Class L
- If (Lc<Rc) then Class R
- If (Lc = Rc) then Class B

Fig. 2: Class distribution

Table 1: Balance scale weight and distance database

ID	Left-weight	Left-distance	Right-weight	Right-distance
B1	4	1	5	3
B2	4	1	5	4
B3	4	1	5	5
B4	4	1	4	1
B5	4	2	1	2
B6	4	2	1	3
B7	4	2	1	4
B8	4	2	4	2
B9	4	2	2	1

Table 2: Classification result of Table 1

Balance ID	Class
B1	R
B2	R
B3	R
B4	B
B5	L
B6	L
B7	L
B8	B
B9	L

Each record is classified as having the balance scale tip to the right, tip to the left or be balanced. The attributes are the left weight, the left distance, the right weight and the right distance. If the multiplication of left-distance and left-weight is greater than the multiplication of right-distance and right-weight then balance class will be Left otherwise, class will be Right. If they are equal, class will be Balance. For understanding the working of the BSclassifier suppose there is 9 balances with ID (B1, B2, …… … … B 9) and each ID having its own left weight, left distance, right weight and right distance which is shown in Table 1. After applying BSclassifier algorithm the each balance will classify with their respective classes which is shown in Table 2.

EXPERIMENTAL RESULT ANALYSIS AND DATA SETS

The Balance Scale Weight and Distance Database has been used to check correctness of the proposed BSclassifier. This is a real dataset and is available at UCI

Machine Learning Repository (Hettich and Bay, 1999). The Balance Scale Weight and Distance Database has 625 numbers of Instances. The datasets is divided into two parts training datasets and test datasets 80 and 20%, respectively. In this experiment 600 instances have used 480 instances for training datasets and 120 instances for test data sets. The total number of attributes are 5 (4 numeric+1 class name) in the Balance Scale Weight and Distance Database. The attributes and Classification result details are shown in Fig. 2.

Attribute information:

- Class Name: 3 (L, B, R)
- Left-Weight: 5 (1-5)
- Left-Distance: 5 (1-5)
- Right-Weight: 5 (1-5)
- Right-Distance: 5 (1-5)

EVALUATION MEASUREMENT

To evaluate the classification accuracy, two basic measures are used Precision and Recall (Han and Kamber, 2001). F-measure compares the results to the pre-classified classes. Where F-measure is derived from the definition of precision and recall in information retrieval. The precision and recall are defined as follows:

Precision: This is the percentage of retrieved records that are in fact relevant to the query. It is calculated as:

$$\text{Precision} = \frac{(\text{Relevant})\text{U}(\text{Retrieved})}{(\text{Retrieved})}$$

Recall: This is the percentage of records that are relevant to the query and were in fact, retrieved. It is calculated as:

$$\text{Recall} = \frac{(\text{Relevant})\text{U}(\text{Retrieved})}{(\text{Retrieved})}$$

The F-measure is defined as the harmonic mean of recall and precision:

$$\text{F-measure} = \frac{2\text{X recall X precision}}{(\text{Recall+Precision})}$$

The F-measure for one class is considered to be the best solution. The better the classification method is the higher overall F-measure will be obtained.

CONCLUSION

In this study, we proposed BSclassifier a new approach for Balance classification. The experimental results show that the proposed new approach for Balance classification outperforms. This new method works efficiently and gives efficient and accurate results. The research presented here is concentrating on Balance classification algorithm. The Balance Scale Weight and Distance Database has used to check correctness of proposed BSclassifier. This BSclassifier gave the better results for real dataset.

REFERENCES

Carvalho, D.R. and A.A. Freitas, 2000. A hybrid decision tree/genetic algorithm for coping with the problem of small disjuncts in data mining. Proceedings of the Genetic and 312 Evolutionary Computation Conference, July 2000, San Francisco, CA. USA., pp: 1061-1068.

Fayyad, U.M., G. Piatetsky-Shapiro and P. Smyth, 1996. From Data Mining to Knowledge Discovery: An Overview. In: Advances in Knowledge Discovery and Data Mining, Fayyad, U.M., G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (Eds.). AAAI/MIT Press, Menlo Park, CA., pp: 1-34.

Han J. and M. Kamber, 2001. Data Mining Concept and Techniques. Academic Press, London.

Hettich, S. and S.D. Bay, 1999. The UCI KDD archive [http://kdd.ics.uci.edu]. University of California, Department of Information and Computer Science, Irvine, CA.

Pujari, A.K., 2000. Data Mining Techniques. University Press, Hyderabad.

Quinlan, J.R., 1986. Induction of decision trees. Machine Learn., 1: 81-106.

Yang, L., D.H. Widyantoro, T. Ioerger and J. Yen, 2001. An entropy-based adaptive genetic algorithm for learning classification rules. Proc. Cong. Evolut. Comput., 2: 790-796.