253. An Extended Weighted Classification Technique using Emerging Patterns and Feature Ranking for Breast Cancer

Yamini C
Anna University, Coimbatore
yamini_c@yahoo.com

M. Punithavalli
SNS College of Engineering, Coimbatore
mpunitha_srcw@yahoo.co.in

Abstract. Decision Tree (DTs) classifier is most important and powerful solution of classification methods. One of the major problems in DTs is that they were built using crisp classes assigned to the training data. In the existing systems this drawback gets override with the concept of Emerging Pattern (EPs). Emerging pattern are those itemsets whose support in one class is significantly higher than their support in other classes. Hence DTs classifier are generalized along EPs so that they can take into account weighted classes assigned to the training data instances .The WDTs classifiers compared with other classifiers and proved that this methods have excellent noise tolerance and good performance. In the proposed system a new weighted decision trees classifiers is constructed using EPs and is compared with weighted Decision tree by applying Fuzzy feature ranking algorithm. Feature selection aims to reduce the dimensionality of patterns for classification by selecting the most informative instead of irrelevant and/or redundant features. In this paper, fuzzy feature clustering is proposed for grouping features based on their interdependence and selecting the best one from each cluster. Feature ranking is determined by means of different criterion functions. The accuracy and speed of both classifiers are evaluated, this comparative evaluation outsource which classifier has best performance.

Keywords: Classification, Decision Tree, Emerging Pattern, Feature Ranking Method

Introduction

Emerging pattern

Decision tree classifier is considered as effective classification technique despite of their simplicity. However, DTs assume that each training data instance is related only to one class (a crisp class). That is, the calculation assumes that each training data instance is related completely to one class only. This assumption conflicts with the fact that most real life datasets suffer from noise. That is, a training instance might not always be assigned to its real class. The notion of weighted classes is proposed in previous research [1]. Assume a dataset consisting of three classes: C1, C2, and C3. An instance i is said to have a crisp class if it is assigned completely to one of the three classes. However, instance i may still have some relations with the other two classes. The notion of weighted classes indicates that i is related to the three classes with different weights. Figure 1 shows examples of a crisp class and a weighted class. In the crisp class, 100% of the weight of instance i is assigned to one of the three classes (in this example, class C1). In the weighted class, the weight is distributed among the three classes. The weight assigned to each class is proportional to the strength of the relation between this class and instance i.

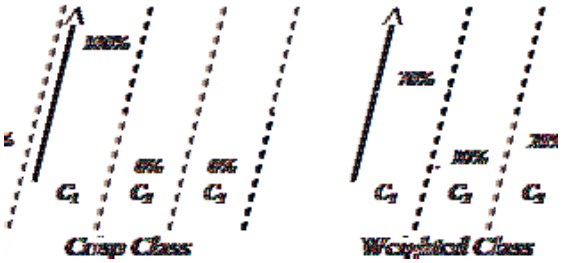


Figure1. Examples of a crisp class and a weighted class [1]

In this paper, the concept of weighted class is assigned to the DTs. Our weighting scheme, proposed in [1], is based on emerging patterns [EPs]. EPs pattern introduced in [3].They have been proved to have a great impact in many

applications [4] [5] [6] [7] [8] [9]. EPs can capture significant changes between datasets. They are defined as itemsets whose supports increase significantly from one class to another. The discriminating power of EPs can be measured by their growth rates. The growth rate of an EP is the ratio of its support in a certain class over that in another class. Usually the discriminating power of an EP is proportional to its growth rate.

Fuzzy feature ranking

By now, many applications have been introduced in which, feature selection is utilized as a preprocessing stage for classification. This process speeds up both the training and reasoning stages, reduces memory space, and improves classification accuracy. Reducing the cost of gathering data is another advantage of feature selection. Small number of samples narrows the acquirable knowledge. Hence it reduces the probability of correct reasoning whether a specified feature effects on the class label or not. Moreover, a classifier can generate the classification rules more easily with small number of features. But increasing the number of features may lead to ambiguity in training so that it would not even converge.

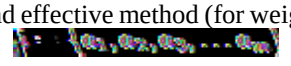
In addition, the more features, the more processing time and memory space is needed. But a few influential features are usually adequately used in classification of samples. Indeed:

The class label is usually independent of the most of features. Some features may be correlated and selecting only a few candidates seems to be sufficient for classification. Fuzzy feature clustering is proposed for grouping features based on their interdependence and selecting the best one from each cluster.

The next section describes previous work of Emerging Pattern and Feature Ranking method. In section 3, Experimental evolutions on dataset are demonstrated. Finally section 4 deals with conclusion.

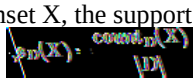
Related Work

Weighted classifiers

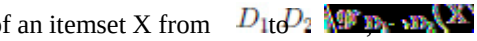
The shortcomings of normal classifiers were Unrealistic Weight, Sensitive to noise, Low Accuracy, Dependent on distance metric. These problems are rectified by most sophisticated and effective method (for weighting the training instances) Emerging Patterns. Initially EPs is defined as follows:  is a data object (instance) following the schema $\{A_1, A_2, A_3 \ldots A_m\}$ $A_1, A_2, A_3 \ldots A_m$ $a_1, a_2, a_3, \ldots a_m$. are attributes and , are values related to these attributes. Each pair (attribute, value) is denoted as an item.

Let Z denote the set of all items in an encoding data set D. Itemsets are subsets of Z. Consider an instance Y contains an itemset X, if $X \subseteq Y$.

Definition1. Given a data set D and an itemset X, the support of X in D, $s_D(X)$, is defined as



where $count_D(X)$ is the number of instances in D containing X.

Definition2. Given two different sets of data D_1 and D_2 , where instances belong to class C_i , let $s_i(X)$ denote the support of the itemset X in the data set D_i. The growth rate of an itemset X from D_1 to D_2  , is defined as

Definition3. Given a growth rate threshold $\rho > 1$, an item set X is said to be a ρ-emerging pattern (ρ-EP or simply EP) from D_1 to ,  >=

When D_1 is clear from the context, an EP e from D_1 to D_2 is called an EP of , the support of e in D_2 , $s_{D_2}(e)$, is simply denoted as the support of e, s(e), and its growth rate from D_1 to , $gr(e)$. ,is denoted as growth rate of e, As stated above assume that we have a set of n training instances $T = \{i_1, i_2, \ldots i_n\}$ classes, $C = \{C_1, C_2 \ldots C_k\}$  . We have a set of EPs mined for each class, such that E_{cj} is a set of EPs related to class . The support of an EP $e \in E_{Cj}$ j is $s_{Ci}(e)$. The growth rate of an EP $e \in E_{Cj}$ is $gr_{Cj}(e)$. The strength of an EP $e \in E_{Cj}$ in class C_j, $\alpha_{Cj}(e)$, is defined as follows:

where $\alpha_{Cj}(e_i)$ represents the contribution of $e \in E_{Cj}$ in class C_j. This contribution is proportional to both the growth rate (discriminating power) of C_j and its support in the home class (C_j). $sup_{Cj}(e)$ .Notice that an EP might have a high growth rate and a low support in its home class and, as a result, its strength will be low. Alternatively, an EP might have a low growth rate and a high support in its home class, again resulting in low strength. That is, in order for an EP to be strong, it has to have both high growth rate and high support.

The overall contribution of EPs contained in an instance $i \in T$ of class C_j, $\beta_{Cj}(i)$ is found by aggregating
the contributions of these EPs [10].

The aggregated value, $\beta_{Ci}(i)$, presented in above equation cannot be directly used as a weight for a training instance. The reason behind this argument is that the number of EPs may differ from one class to another. As a result, the class with the largest number of EPs will have the highest aggregated value. To overcome this problem, the aggregated values of instances in a class are divided by the median aggregated value in the same class. This division balances the aggregated values of an instance in the different classes. That is, a large number of EPs in a class will not substantially bias the final weight toward this class. The weight of a training instance $i \in T$ in class C_j, $\omega_{Ci}(i)$, is defined as follows:

where $Median_{Cj}$ is the median of the aggregated values above equation in class C_j.. The weight is calculated for each training instance in each class. The weights of each training instance are normalized so that their sum is equal to 1. The normalized weight of a training instance $i \in T$ in class C_j, $\delta_{Cj}(i)$ is defined as follows

The normalized weight represents the strength of the relation between an instance and a class. That is, it
represents the weighted class for this instance.

Weighted Decision tree are constructed by this weighting scheme. After applying the above weighting scheme on data instances, these instances will change from crisp classes where every instance is assigned completely to one class to weighted instances where the weight of each instance is distributed among different classes

Featured ranking procedure

Existing feature selection/ranking techniques are mostly suitable for classification problems, where the range of the output is discrete. These techniques result in a ranking of the input feature (variables). The approach exploits an arbitrary fuzzy classifying of the model output data. Using these output classes, similar feature ranking methods can be used as for classification, where the membership in a cluster (or class) will no longer be crisp, but a fuzzy value determined by the classification. The Sequential Backward Selection (SBS) search method is proposed to determine the feature ranking by means of different criterion functions.

Feature selection methods are of two main types: Feature selection and ranking methods [1]. The methods of the former type determine which input features are relevant in a given model, whilst the ones of the latter type result in a rank of importance. Feature ranking methods can be considered as preprocessing of feature selection, because relevant features can be selected by taking the first k elements of the head of the feature ranking, and then, by optimizing the number of k, e.g. by a trial-and-error procedure. The method aims at providing a reliable feature ranking method for weighted classifiers.

A fuzzy classification method divides the clustered space into various regions, called clusters, and determines a vector of membership degrees for each data, which indicates the grade to which the particular data belongs to the clusters. Because clustering is only in the way of one dimensional output, the shape of the clusters (e.g. spherical or ellipsoid) is irrelevant, due to the fact that in our case clusters are interval.

The feature ranking on fuzzy clustered output (FRFCO) algorithm [11]

1. Let
2. For all .
   (a) Let ,and update matrix X by deleting temporarily its th row, and vectors by deleting temporarily th element.
   (b) Calculate matrices and determine .
3. Let i.e. where J attains its minimal value. The final is obtained by deleting permanently the variable from , and then update expressions X, vi and x appropriately.
4. If then back to step 2, else stop.

The order of the deleted variables gives their rank of importance.

Remark1. Note that can contain more than one variable. In such a case we delete all of them at a time. The Feature ranking algorithm is an instance of SBS search method. In our application, SBS method applies

the interclass separability criterion function. This method has two advantages. Firstly, it has more stability and faster convergence due to fuzzy clustering; secondly, it improves the accuracy of the classifier using the selected features.

In the ith step, temporarily a variable is deleted , so that feature set and input matrix , where the starting feature set is , and then calculate the matrices and to be used in the criterion functions. This procedure is repeated for all the variables in . By means of an appropriate criterion function, the expression attains its minimum when the deviation between Qb and Qw is the least, i.e. when the most important variable is deleted. Then we remove the selected feature permanently, and then restart the algorithm with the updated feature set . The algorithm ends when the cardinality of the feature set is 1.

Experimental Evaluation

Weighted classifiers

In this section, weight is assigned to dataset using emerging pattern, and then weighting scheme is applied to classifiers. Then Weighted Decision classifier (C4.5), and Weighted DT with Emerging Pattern, Weighted DT with Emerging Pattern are compared. The experimental evaluation is on two datasets from UCI repository of machine learning databases. .The accuracy is obtained by evaluation these algorithms respectively on datasets.

Fuzzy feature ranking weighted classifier

Feature ranking is applied on Weighted C4.5 classifiers. The accuracy of classifiers is compared.

Accuracy comparison between Weighted Decision tree (WDT), Weighted Decision tree Emerging Pattern

(WDT-EP) and Weighted Decision tree Emerging Pattern Feature Ranking (WDT-EPFR)

Table 1. Breast Cancer dataset taken from UCI Repository has 9 Attributes, 286 Instances

No of Records	WDT	WDT-EP	WDT-EPFR
50	97.6	97.8	98.1
100	98.1	98.3	98.5
150	98.8	98.9	99.0
200	99.4	99.5	99.7

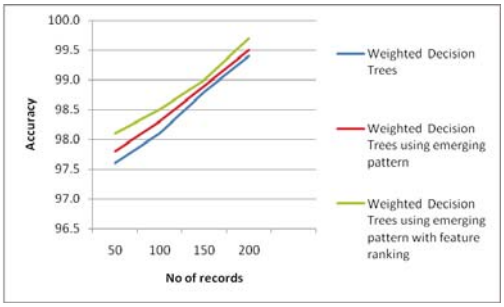


Figure 2. comparison of Classifiers accuracy WDT=99.4, WDT-EP=99.5, WDT-EPFR=99.7 (Beast Cancer)

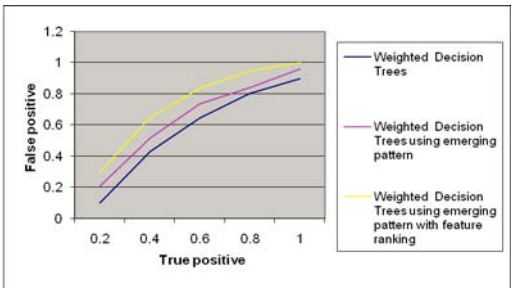


Figure 3. ROC Curves of WDT,WDT-EP and WDT-EPFR (Beast Cancer)

Table 2. Breast Cancer Wisconsin dataset taken from UCI repository has 32 attributes, 569 instances

No of Records	WDT	WDT-EP	WDT-EPFR
100		78.2	79.3
80.2			
200		79.3	80.4
81.3			
300		80.2	81.3
82.2			
400		81.3	82.2
83.0			
500		82.1	83.0
83.7			

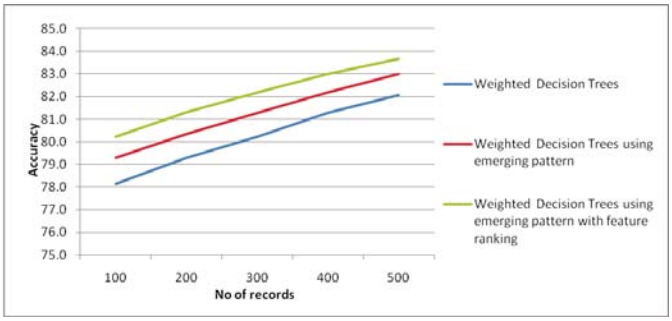


Figure 4. Comparison of Classifiers Accuracy
WDT=82.1, WDT-EP=83.0, WDT-EPFR=83.7 (Breast cancer Wisconsin)

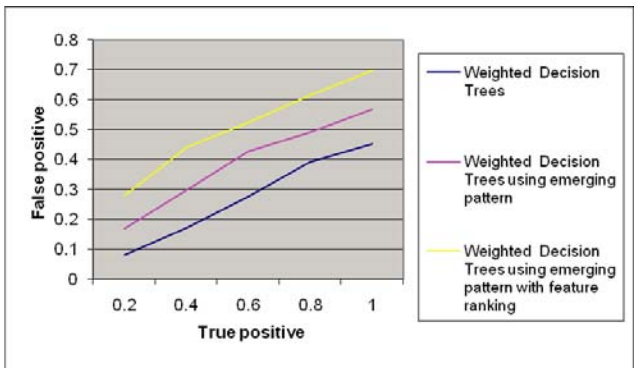


Figure 5. ROC Curves of WDT,WDT-EP and WDT-EPFR (Breast cancer Wisconsin)

Speed evaluation

Execution time for Fuzzy Featured-Weighted Decision tree using Emerging patterns classifier is measured in Table 3.

Table 3. Speed Measurement

Data Set	
Speed Breast Cancer	
17.65/ms Breast Cancer Wisconsin (prognostic) 62.96/ms	

Receiver operating characteristics (ROC)

Receiver Operating Characteristics (ROC) curves is a helpful method for visualizing the performance of classification. ROC curves are plotted on two-dimensional graphs. The X axis represents the true positive rate (TPR) and Y axis represents the false positive rate (FPR).The ROC curves show that WDT-EPFR has better performance than WDT-EP.

Conclusion

From the parameter comparison among two datasets it is concluded Weighted Decision tree with Emerging Pattern and feature ranking algorithm has better performance (accuracy) than Weighted Decision tree algorithm. In future, accuracy can be improved using partitioning algorithms.

References

1. Alhammady, H., & Ramamohanarao, K.. “Using Emerging Patterns to Construct Weighted Decision Trees” in
IEEE Transaction on Knowledge and Data Engineering, Vol 18, NO. 7, July 2006.
2. G. Dong, and J. Li. “Efficient Mining of Emerging Patterns: Discovering Trends and Differences”. In Proceedings of the 1999 International Conference on Knowledge Discovery and Data Mining (KDD’99), San Diego, CA, USA.
3. H. Alhammady, and K. Ramamohanarao. “The Application of Emerging Patterns for Improving the Quality of Rare-class Classification”. In Proceedings of the 2004 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’04), Sydney, Australia.
4. H. Alhammady, and K. Ramamohanarao. “Using Emerging Patterns and Decision Trees in Rare-class Classification”. In Proceedings of the 2004 IEEE International Conference on Data Mining (ICDM’04), Brighton, UK.
5. H. Alhammady, and K. Ramamohanarao. “Expanding the Training Data Space Using Emerging Patterns and
Genetic Methods”. In Proceeding of the 2005 SIAM International Conference on Data Mining (SDM’05),
6. New Port Beach, CA, USA.
7. H. Fan, and K. Ramamohanarao. “A Bayesian
8. Approach to Use Emerging Patterns for Classification”. In Proceedings of the 14th Australasian Database
Conference (ADC’03), Adelaide, Australia.

9. Guozhu D., Xiuzhen Z., Limsoon W., and Jinyan L. “CAEP: Classification by Aggregating Emerging Patterns”. In Proceedings of the 2nd International Conference on Discovery Science (DS’99), Tokyo, Japan.
10. Alhammady, H., & Ramamohanarao, K. (2005). “Mining Emerging Patterns and classification in data streams”. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI), Compiegne, France, pp. 272-275.
11. C. Blake, E. Keogh, and C. J. Merz. “UCI repository of machine learning databases”. Department of Information and Computer Science, University of California at Irvine, CA, 1999.
12. Domonkos Tikk, Tamas D. Gedeon, and Kok Wai Wong, “A Feature Ranking Algorithm for Fuzzy Modelling Problems”.