Applied Soft Computing 31 (2015) 91–102

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Pattern Matching based Classification using Ant Colony Optimization based Feature Selection

N.K. Sreeja [a,*], A. Sankar [b]

[a] Department of Computer Applications, Sri Krishna College of Technology, Coimbatore 641042, India
[b] Department of Computer Applications, PSG College of Technology, Coimbatore 641004, India

ARTICLE INFO

Article history:
Received 15 October 2012
Received in revised form 18 January 2015
Accepted 25 February 2015
Available online 5 March 2015

Keywords:
Classification
Pattern matching
Feature selection
Ant Colony Optimization

ABSTRACT

Classification is a method of accurately predicting the target class for an unlabelled sample by learning from instances described by a set of attributes and a class label. Instance based classifiers are attractive due to their simplicity and performance. However, many of these are susceptible to noise and become unsuitable for real world problems. This paper proposes a novel instance based classification algorithm called Pattern Matching based Classification (PMC). The underlying principle of PMC is that it classifies unlabelled samples by matching for patterns in the training dataset. The advantage of PMC in comparison with other instance based methods is its simple classification procedure together with high performance. To improve the classification accuracy of PMC, an Ant Colony Optimization based Feature Selection algorithm based on the idea of PMC has been proposed. The classifier is evaluated on 35 datasets. Experimental results demonstrate that PMC is competent with many instance based classifiers. The results are also validated using nonparametric statistical tests. Also, the evaluation time of PMC is less when compared to the gravitation based methods used for classification.

1. Introduction

Machine learning is the development of algorithms that allow computers to learn based on empirical data. The goal of Machine learning is to build computer systems that adapt and learn from their experience. Machine learning can be either supervised or unsupervised. An example of supervised learning is classification. It is defined as the task of learning from instances described by a set of features (attributes) and a class label. The result of learning is a classification model that is capable of accurately predicting the class label of unlabelled samples.

Several algorithms such as artificial neural networks [28], decision tree, support vector machines (SVMs) [42], instance based learning methods [1] and nature-inspired techniques such as genetic programming [10] have been proposed in literature for classification. Among these, decision tree, back-propagation network (BPN) and SVM classifiers are popular, and can be applied to various areas [7,21]. However, choosing the best kernel function is necessary for SVM. The usually preferred kernel function is Radial Basis Functions (RBF). RBF gives optimal performance only when the parameters are set properly. Lin et al. [23] adopted a simulated annealing approach and a particle swarm optimization approach for parameter setting and feature selection for SVM. However, SVM and BPN classifier does not handle missing values effectively [35].

Linear Discriminant Analysis (LDA) is a commonly used classification method. It can provide important weight information for constructing a classification model. LDA often suffers from the small sample size problem when the number of dimensions of the data is much greater than the number of data points. Lin et al. [34] have proposed a particle swarm optimization (PSO) method to enhance the classification accuracy of LDA. However this method is sensitive to parameter settings.

Although many classification algorithms exist in literature, instance based methods are attractive due to its simplicity. The Nearest Neighbor (NN) algorithm [6] is an instance based method which employs a simple classification procedure. Neighborhood based methods are attractive primarily due to their simplicity and good performance. However, the major problem with these methods is that they severely deteriorate with noisy data or high dimensionality, their performance becomes very slow, and their accuracy tends to deteriorate as the dimensionality increases, especially when classes are not separable or they overlap [22].

In recent years, new instance based methods have been proposed to overcome the drawbacks existing in NN classifiers. One

* Corresponding author. Tel.: +91 9659425507.
E-mail addresses: sreeja.n.krishnan@gmail.com (N.K. Sreeja), dras@mca.psgtech.ac.in (A. Sankar).

http://dx.doi.org/10.1016/j.asoc.2015.02.036
1568-4946/

such approach is Data Gravitation based Classification (DGC) proposed by Peng et al. [32,40,44]. The basic principle of DGC algorithm is to classify data samples by comparing the data gravitation between the different data classes [32]. The algorithm creates data particles using the maximum distance principle. However, the drawback of this method is that it reduces the accuracy, especially in the area away from the data particle centroid and along the border between classes [2]. Cano et al. [2] have proposed another instance based method called Weighted Data Gravitation based Classification (DGC+) and is proved to achieve greater classification accuracy than DGC approach [32]. However, the computational complexity of DGC+ is considerably higher. To overcome the drawbacks existing in NN classifiers and gravitation based models, a simple instance based algorithm based on pattern matching is proposed.

In this paper, a novel instance based algorithm called Pattern Matching based Classification (PMC) is proposed to classify unlabelled samples based on the similarity between the feature values of the instances in the dataset and the unlabelled sample. PMC classifies the unlabelled samples by matching the features of the unlabelled sample with that of the features of the instances in the dataset. The instances in the dataset having the maximum number of matching features are grouped. PMC votes for the majority class label in the group to classify the unlabelled sample. A probabilistic approach is used to predict the target class of the unlabelled sample when more than one class label have the same majority. To improve the classification accuracy of PMC algorithm, an Ant Colony Optimization based Feature Selection approach based on the idea of PMC is used.

Experiments have been carried out on 35 data sets collected from the KEEL [3] and UCI [12] repositories. The experiments have been carried out for different problem domains, number of instances, attributes, and classes. It is shown that PMC is competent with the recent instance based algorithms obtaining significantly better results in terms of predictive accuracy and Cohen's kappa rate [4,5]. The result of statistical analysis such as Iman and Davenport test [24] and Bonferroni–Dunn tests [9,14,33] show that there are significant differences in the results of the algorithms. Also, the computational complexity of PMC algorithm is less when compared to the gravitation based approaches such as DGC+ and DGC.

The paper is organized as follows. Section 2 presents the related work. Section 3 describes the proposed PMC algorithm. Section 4 describes the feature selection for PMC. Section 5 describes a case study. The experimental study is described in Section 6. Section 7 describes the results of the experiments. The time performances of PMC, DGC+ and DGC are compared in Section 8. Section 9 presents the discussion and Section 10 presents some concluding remarks.

2. Related work

This section presents an overview of the various instance based methods and the gravitation based methods developed recently for classification. The simplest amongst all nearest neighbor classifier is K-Nearest Neighbor (KNN). They perform a class voting among the k closest training instances. The drawback of standard KNN classifier is that it does not output meaningful probabilities associated with class prediction [26]. Therefore, higher values of k are considered for classification which provides smoothing that reduces the risk of over-fitting due to noise in the training data [26]. However, choosing higher value of k, leads to misclassification of unlabelled samples having an exact pattern as that of an instance in the dataset.

Dudani proposed the distance-weighted NN rule (DW-KNN) [8] for classification. It weights the evidence of a neighbor close to an unlabelled sample more heavily than the evidence of another neighbor which is at a greater distance from the unlabelled sample. It assigns the unlabelled sample to the class for which the weights of the representatives among the K NNs sum to the greatest value.

Gao and Wang proposed the center-based NN classifier (CNN) [13]. CNN classifies the unknown sample by computing the distance between the training instances and centers of their class to find how far the training instances are from the unlabelled sample. However, the performance of the algorithm deteriorates when the center of the data classes is overlapped. Wang *et al.* proposed an adaptive k NN algorithm (KNN-A) [41] which weights the distance from the training instance to its nearest instance belonging to a different class. Thus the instances near to the decision boundaries become more relevant.

Paredes and Vidal [29,30] proposed a class-dependent weighted dissimilarity measure in vector spaces to improve the performance of the NN classifier. It computes a dissimilarity measure such that the distances between points belonging to the same class are small while interclass distances are large. The accuracy of NN classifiers can further be improved by Prototype selection [15] and prototype generation [37]. Prototype methods aim to select a relatively small number of samples from a dataset which if well chosen can serve as a summary of the original dataset. Paredes and Vidal extended their NN model together with a prototype reduction algorithm [37], which simultaneously trains both a reduced set of prototypes and a suitable local metric for them.

Zhou and Chen [43] proposed a Cam weighted distance (CamNN) to improve the performance of classification by optimizing the distance measure based on the analysis of inter-prototype relationships. Triguero et al. [38] applied differential evolution to the prototype selection problem as a position adjusting of prototypes. SSMA–SFLSDE [38] combines a prototype selection stage with an optimization of the position of prototypes, prior to the NN classification. This enhances the performance of SSMA [16] + SFLSDE [27].

Cano et al. [2] have shown that although the NN classifiers have achieved fame by their simplicity, they are very sensitive to irrelevant, redundant, or noisy features because all features contribute to classification. Also, it is stated that the problem worsens for high dimensional datasets. To overcome these drawbacks, instance based methods such as Data Gravitation based approaches are proposed.

Peng et al. [32] proposed Data Gravitation based Classification (DGC) method to classify datasets. According to this model, a kind of "force" called data gravitation between two data samples are computed. Data from the same class are combined as a result of gravitation. The algorithm also employs a tentative random feature selection to calculate the weights of features by simulating the mutation operation in a genetic algorithm. DGC also achieved reasonably high accuracies, but fails to classify imbalanced datasets.

As an improvement of DGC, Cano et al. [2] proposed an algorithm called Weighted DGC (DGC+) that compares the gravitational field for the different data classes to predict the class with the highest magnitude. The proposal improves previous data gravitation algorithms by learning the optimal weights of the attributes for each class and solves some of their issues such as nominal attributes handling, imbalanced data performance, and noisy data filtering. However, the computational complexity of DGC+ algorithm is high.

3. Pattern Matching based Classification

PMC algorithm is used to classify an unlabelled sample based on the instances in the training dataset. Consider a dataset with p instances and n features. It can be represented as a data matrix