Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 33 (2007) 49–60
Expert Systems with Applications
www.elsevier.com/locate/eswa

A hybrid approach for feature subset selection using neural networks and ant colony optimization

Rahul Karthik Sivagaminathan, Sreeram Ramakrishnan *

Department of Engineering Management and Systems Engineering, 1870 Miner Circle, University of Missouri – Rolla, Rolla, MO 65409, USA

Abstract

One of the significant research problems in multivariate analysis is the selection of a subset of input variables that can predict the desired output with an acceptable level of accuracy. This goal is attained through the elimination of the variables that produce noise or, are strictly correlated with other already selected variables. Feature subset selection (selection of the input variables) is important in correlation analysis and in the field of classification and modeling. This paper presents a hybrid method based on ant colony optimization and artificial neural networks (ANNs) to address feature selection. The proposed hybrid model is demonstrated using data sets from the domain of medical diagnosis, yielding promising results.

Keywords: Feature subset selection; Ant colony optimization; Neural networks

1. Introduction

1.1. Pattern classification

Pattern classification is the task of classifying any given input feature vector into pre-defined set of classes of patterns (Kulkarni & Vidyasagar, 1997) where as pattern recognition is the task of making important decisions based on complex patterns of information (Ripley, 1996). A detailed discussion on definition and various tools for pattern classification can be found in (Kulkarni, Lugosi, & Santosh, 1998). These methods include Artificial Neural Networks (ANNs), nearest neighbor, kernel and histogram methods, and support vector machines. Researchers in this area focus on characterizing problems to determine if a particular problem can be learned or not, the amount of data required for learning, and then developing the necessary algorithms for learning. Among the existing methods, ANNs have attracted many researchers and has emerged as the most popular tool for pattern recognition and classification. One domain where such applications have found significant utility is the analysis of medical data sets.

Certain problems in the medical diagnosis domain can be considered as a problem of pattern recognition and classification. The use of ANNs is not new in medical diagnosis. For example, in Lanzarini and Giusti (2000), ANNs were used to recognize patterns in medical images. In Zhou, Jiang, Yang, and Chen (2002), an automatic pathological diagnosis procedure named Neural Ensemble based Detection (NED) is proposed, which utilizes an ANN ensemble to identify lung cancer cell images. Unlike other researchers employing back propagation neural networks, in Desai, Lin, and Desai (2001) a neural network based on Kohonen's Linear Vector Quantization is used for the diagnosis of prostate cancer. In this paper, data sets from medical diagnosis are used to demonstrate the feature reduction method using ant optimization.

1.2. Importance of feature selection in classification methods

Many practical pattern classification tasks (Blum & Langley, 1997) (e.g., medical diagnosis) require learning

* Corresponding author. Tel.: +1 573 341 6787; fax: +1 573 341 6567.
E-mail address: Sreeram@umr.edu (S. Ramakrishnan).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.04.010

of an appropriate classification function that assigns a given input pattern (typically represented by using a vector of feature values) to one of a set of classes. The choice of features used for classification could have an impact on the accuracy of the classification function, the time required for classification, training data set requirements, and implementation costs associated with the classification.

The accuracy of the classification function that can be learned using an inductive learning algorithm such as ANNs depends on the set of input features. The attributes or features used to describe the pattern implicitly define a correlation. If the correlation is not accurate enough it would fail to capture the information that is necessary for classification and hence regardless of the learning algorithm used the accuracy of the classification function would be limited. In their paper (John, Kohavi, & Pfleger, 1994) explained the importance of identifying relevant and irrelevant features. It should also be noted that the time and the size of training data set(s) needed for learning a sufficiently accurate classification function increases for more complex patterns with many features (Punch et al., 1993).

The cost for measuring a feature is a critical issue to be considered while selecting a subset. In case of medical diagnosis the features may be observable symptoms or diagnostic tests. Each clinical test is associated with its own diagnostic value, cost and risk. The challenge is in selecting the subset of features with minimum risk, least cost yet which is significantly important in the determining its class/pattern. In Gorunescu, Gorunescu, Darzi, and Gorunescu (2005) a probabilistic neural network with heuristics is used for feature selection in cancer diagnosis.

As can be seen from the above discussion, the issue of feature subset selection in automated design of pattern classifiers is an important research issue. The feature subset selection problem refers to the task of identifying and selecting a useful subset of features to be used to represent patterns from a larger set of often mutually redundant, possibly irrelevant, features with different associated measurement costs and/or risks. Examples of feature subset selection problem include large-scale data mining, power system control, and medical diagnosis (Yang & Honavar, 1998).

1.3. Background

The existing literature in this domain is rich with different solution techniques. Initial methods included exhaustive search in which all combinations of subsets were evaluated. The method guarantees an optimal solution, but finding the optimal subset of features is NP hard. For large number of features, evaluating all states is computationally non-feasible (Boz, 2002) necessitating the need for heuristic search methods. As in Doak (1992), these methods can be classified as exponential, sequential or randomized methods.

The "exponential method" includes methods such as "branch and bound" (Narendra & Fukunaga, 1977) which starts from a full set and removes features using a first depth strategy. The method guarantees an optimal solution under the monotonic assumption that the children of the nodes whose objective function values are lesser than the current best will not contain a better solution and so these features will not be further explored. The other method in this category includes beam search (Doak, 1992). In this method, the features are arranged in a queue with the best states placed at the head of the queue. At each iteration, beam search evaluates all possible states that result from adding a feature subset.

Sequential search algorithms (SSA), also known as stepwise methods (Pudil, Novovicova, & Kittler, 1994), have a relatively lower complexity and use the "hill climbing" strategy to find an optimal solution. Depending upon the different starting points SSA is classified in to sequential forward selection (Devijver & Kittler, 1982) starting with an empty set, Sequential Backward Selection starting from the complete feature set. Meta-heuristic methods are generally considered as random search methods. Some popular meta-heuristic algorithms include genetic algorithm (Leardi, Boggia, & Terrile, 1992; Yang & Honavar, 1998) and simulated annealing (Debuse & Smith, 1997).

This paper presents as ant colony optimization (ACO) approach for feature selection problems using data sets from the field of medial diagnosis. This paper presents a novel approach for heuristic value calculation, which will reduce the set of available features. The rest of this paper is organized as follows. In the next section, an introduction on ACO applications in feature selection problems is discussed. The different methods for feature selection problems (based on the existence of classification function) are presented in Section 3. In Sections 4 and 5, the proposed hybrid methodology is discussed, followed by a discussion on the experimental setup, datasets used and the results.

2. Ant colony optimization

Ant algorithm was first proposed by Dorigo and Gambardella (1997) as a multi-agent approach for difficult combinatorial optimization problems such as traveling sales man problem (TSP) and the quadratic assignment problem (QAP). From then, researchers have applied ACO to many discrete optimization problems (Bonabeau, Dorigo, & Theraulaz, 1999; Corne, Dorigo, & Glover, 1999).

ACO is a meta-heuristic approach which has been applied to various NP hard problems such as static/dynamic combinatorial optimization. ACO applications in static combinatorial optimization problems include job shop scheduling (Blum & Sampels, 2002; Colorine, Dorigo, & Maniezzo, 1994), flow shop (Stützle, 1998), open shop (Blum, 2003), group shop (Sampels, Blum, Mastrolilli, & Rossi-Doria, 2002), vehicle routing (Bullnheimer, Hartl, & Strauss, 1998), sequential ordering (Gambardilla & Dorigo, 1997), graph coloring (Costa & Hentz, 1997) and shortest common super sequences (Micheal & Middendorf, 1999). ACO application to dynamic combinatorial optimi-