Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biomedical Informatics 38 (2005) 456–468
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

GMDH-based feature ranking and selection for improved classification of medical data

R.E. Abdel-Aal *

Physics Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

Received 7 February 2005
Available online 16 April 2005

Abstract

Medical applications are often characterized by a large number of disease markers and a relatively small number of data records. We demonstrate that complete feature ranking followed by selection can lead to appreciable reductions in data dimensionality, with significant improvements in the implementation and performance of classifiers for medical diagnosis. We describe a novel approach for ranking all features according to their predictive quality using properties unique to learning algorithms based on the group method of data handling (GMDH). An abductive network training algorithm is repeatedly used to select groups of optimum predictors from the feature set at gradually increasing levels of model complexity specified by the user. Groups selected earlier are better predictors. The process is then repeated to rank features within individual groups. The resulting full feature ranking can be used to determine the optimum feature subset by starting at the top of the list and progressively including more features until the classification error rate on an out-of-sample evaluation set starts to increase due to overfitting. The approach is demonstrated on two medical diagnosis datasets (breast cancer and heart disease) and comparisons are made with other feature ranking and selection methods. Receiver operating characteristics (ROC) analysis is used to compare classifier performance. At default model complexity, dimensionality reduction of 22 and 54% could be achieved for the breast cancer and heart disease data, respectively, leading to improvements in the overall classification performance. For both datasets, considerable dimensionality reduction introduced no significant reduction in the area under the ROC curve. GMDH-based feature selection results have also proved effective with neural network classifiers.

Keywords: Abductive networks; Neural networks; Feature ranking; Feature selection; Dimensionality reduction; Classification accuracy; ROC characteristics; Medical diagnosis; Breast cancer; Heart disease

1. Introduction

Machine learning classification techniques provide support for the decision-making process in many areas of health care, including screening, diagnosis, prognosis, monitoring, therapy, survival analysis, and hospital management. Tools used include Bayesian and nearest-neighbor classifiers, rule induction methods, decision trees, fuzzy logic, artificial neural networks, and abductive networks [1] based on the group method of data handling (GMDH) algorithm [2]. Compared to neural networks, abductive networks allow easier model development and provide more transparency and greater insight into the modeled phenomena, which are important advantages in medicine. Medical applications of GMDH-based techniques include modeling obesity [3], analysis of school health surveys [4], drug detection from EEG measurements [5], medical image recognition [6], and screening for delayed gastric emptying [7]. Accuracy is very important in classifiers used for medical applications. A high percentage of false negatives in screening systems increases the risk of real patients not receiving the attention they need, while a high false

* Fax: +966 3 860 4281.
E-mail address: radwan@kfupm.edu.sa.

1532-0464/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2005.03.003

alarm rate causes unwarranted worries and increases the load on medical resources. In quest for higher classification accuracies, feature subset selection has been used for data reduction in areas characterized by high dimensionality due to the large number of available features, e.g., in remote sensing [8], seismic data processing [9], speech recognition [10], drug design [11], and image segmentation [12]. This approach attempts to select a small subset of optimum features that ideally is necessary and sufficient to describe the phenomenon being modeled [13]. Feature subset selection is expected to improve classification performance, particularly in situations characterized by the high data dimensionality problem caused by relatively few training examples compared to a large number of measured variables. This situation arises frequently in medicine where considerations of risk, time, difficulty, cost, and inconvenience may limit the number of training examples, while the number of disease markers increases rapidly over the years [14]. Even if no significant improvements in classification performance are achieved, feature reduction has many practical advantages in reducing the number of measurements required, shortening training and execution times, and improving model compactness, transparency, and interpretability. Fewer model inputs result in simpler models that train and execute faster, and allow training on smaller datasets without the risk of overfitting. Reducing the number of attributes to be measured for model implementation makes screening tests faster, more convenient, and less costly. Simpler models with fewer inputs are also more transparent and more comprehensible, providing better explanation of suggested diagnosis, which is an important requirement in medical applications. Discarding irrelevant and redundant features reduces noise and spurious correlations with the output, and avoids the problems of colinearity between inputs, e.g., instability of least squares estimates and removal of solution uniqueness [15]. Feature reduction has been applied to several areas in medicine, including: classification of EEG signals for operating brain–computer interfaces [16], classification of hepatic lesions from computed tomography images [17], detection of mass lesions in digital mammograms [18], segmenting digital chest radiographs [19], processing of ECG signals for the detection of obstructive sleep apnea [20], classification of ultrasound liver tissues using the wavelet transform [21], and detection of seizure events in newborn children using EEG data [22].

Techniques for feature subset selection can be classified into three main categories: embedded, filter (open-loop), and wrapper (closed-loop) techniques [23]. With embedded techniques, feature selection is performed as part of the induction learning itself. By testing the values of certain features, decision tree algorithms seek to split the training data into subsets, each containing a strong majority of one class. Both filter and wrapper techniques perform feature selection as a preprocessing step prior to the modeling application, with the objective of selecting an optimum feature subset that serves as an input to the learning algorithm. Filter techniques do not use the learning mechanism for feature selection. They filter out undesirable and redundant features through checking data consistency and eliminating features whose information content is represented by others. Examples of filter techniques for feature selection include Relief [13], which ranks individual features according to a feature relevance score. The correlation-based feature selection (CFS) technique [24] scores and ranks subsets of features, rather than individual features. It uses the criterion that a good feature subset for a classifier contains features that are highly correlated with the class variable but poorly correlated with each other. Information theoretic measures, such as the mutual information criterion, were used for feature selection to avoid mistakes introduced by linear measures such as correlation [25]. The Bhattacharyya probabilistic distance and other statistical measures were used to select feature subsets that maximize class separability [26]. Since filter methods do not use the learning algorithm, they are fast and therefore suitable for use with large databases. Also, resulting feature selections are applicable to various learning techniques. Wrapper techniques [27] search for an optimal feature subset through testing the performance of candidate subsets using the learning algorithm. As the learning algorithm is called repeatedly, wrapper methods are slower than filter methods and do not scale up well to large, high-dimensional datasets, particularly with neural networks, which require long training times. To overcome this limitation, feature subset evaluation could use a simpler learning algorithm, e.g., nearest-neighbour classifier, that is closely related to the target neural network architecture [28]. Wrapper feature selections are unique to the learning algorithm used, and the process should be repeated for a different learning algorithm. Strategies used for searching the feature space include sequential feature selection (SFS) methods [29], either with forward sequential search (FSS) or backward sequential search (BSS). FSS starts with an empty set, adding single features that best improve performance criteria. BSS starts with the full feature set and sequentially removes features whose removal leads to maximum gain in performance. Genetic algorithm (GA) search methods have been used with both filters [12] and wrappers [28]. Feature selection techniques based on the rough set theory have also been proposed [30].

This paper describes a novel technique for feature ranking and selection with GMDH-based abductive network classifiers. The method relies on the property of the GMDH learning algorithm [1,2] of automatically selecting optimum predictors [31] at various levels of model complexity specified by the user. Information gathered in this way is used to rank the available