Receiver operating characteristic (ROC) and other curves measuring discriminability of classifiers′ ensemble for asthma diagnosis

Jurkowski P[1], **Ćwiklińska-Jurkowska M**[2], **Doniec Z**[3], **Szaflarska-Popławska A**[4]

[1] Department of Informatics and Research Methodology; Collegium Medicum; Nicolaus Copernicus University, Poland
[2] Department of Theoretical Background and Medical Informatics; Collegium Medicum; Nicolaus Copernicus University, Poland
[3] Institute of Tuberculosis and Lungs Diseases, Rabka, Poland
[4] Department of Pediatrics, Allergology and Gastroenterology; Collegium Medicum; Nicolaus Copernicus University, Poland

Abstract

Purpose: **The aim was studying the discriminability by ROC curves and gain charts for simple fixed combining of constituent classifiers, for asthma severity diagnosis, and also for bagging and boosting.**

Material and methods: **ROC shows a performance over a range of relative costs and probabilities a priori. Area under ROC curve (AUC) is the measure of separability of two probability distributions, for example of classifying functions. We examined ROC curves of different discriminant methods such as logistic regression, classification trees and neural networks. Next we combined these constituent classifiers and compared the obtained curve with curves of constituent classifiers. The analogous analysis was made on other methods of classifiers' ensemble: bagging and boosting. Besides ROC-in the same way we examined also another curves, measuring discriminability cumulative and non-cumulative lift charts. Social and simple clinical data of 439 patients from three groups of children, hospitalized at the Institute of Pulmunology in Rabka, were used to find classification functions for existing and severity of asthma. We studied also two-group classification problems: asthmatic and non asthmatic children to elaborate automatic predicting of asthma.**

Results: **We found out features with biggest discriminant properties in the differentiation groups of existing and severity of asthma. The improvement of performance after combining classifiers was proved by examining errors of classification and curves measuring discriminability.**

Conclusions: **Performance of ensemble method can be visualized in one graph and compared with joint graph of constituent classifiers.**

ADDRESS FOR CORRESPONDENCE:
Piotr Jurkowski
Department of Informatics and Research Methodology
Collegium Medicum; Nicolaus Copernicus University,
ul. Techników 3; 85-801 Bydgoszcz, Poland
Fax: 052-52193;
e-mail: jurkomal@atr.bydgoszcz.pl

Key words: aiding medical diagnosis, childhood asthma, discriminability, ROC, gain chart, combining classifiers.

Introduction

Receiver Operating Characteristic curve (ROC) is an indicator of performance of two classification rules. The purpose of the work was examining the discriminability by ROC curves and gain charts for combining of constituent classifiers for asthma severity diagnosis and comparing the usefulness of them.

Material and methods

Assessment of performance [1] can be made by:
1) discriminability (error rates: apparent=resubstitution, true=actual=conditional, Bayes error rate; holdout estimate, where set is divided into train and test subsets, cross-validation, jack-knife, bootstrap);
2) reliability (imprecision)
3) ROC curves (only for two classes).

The Receiver Operating Characteristic (ROC) curves can be used as indicators of performance for two-populations classification rules. ROC is a plot of the sensitivity as the function of (1-specificity). ROC curves are also called Lorentz diagrams. Charts with reversed axes are called ODS – Ordinal Dominance Curves.

The positive likelihood ratio is the slope. For discriminant Bayesian rule:

$$slope= c_2\Pi_2/c_1\,\Pi_1 \text{ where:}$$

c_1, c_2 are costs of misclassifications to groups 1 and 2;

c_1 and c_2 are probabilities a priori for groups 1 and 2.

Straight lines of constant costs (iso-performance lines) are such that gradients are equal to the slope. Minimum loss is obtained where line of loss contour are tangential with ROC. In practice precise costs are not known – ROC shows a performance over a range of relative costs. AUC – Area Under Curve does not depend on the relative costs of misclassifications [1]. Area under ROC curve (AUC) is the measure of separability of two probability distributions, for example of two classifying functions: excellent for AUC values bigger than 0.9, good for 0.8-0.9, fair for 0.7-0.8, poor for 0.6-0.7 and fail for values smaller than 0.6. ROC shows a performance over a range of relative costs and probabilities a priori.

In multivariate normal case AUC has the simple and clear probabilistic interpretation – it is a probability that classification function to one group is stochastically larger than classification function to the other group [2]. ROC curve is concave, if densities in two groups have a monotone likelihood ratio [3]. During last years some publications on modelling the classifying functions based on the area under ROC curves have appeared. For example, Ferri et al. [4] propose a novel splitting criterion in decision trees, which chooses the split with the highest local area under curve.

In order to avoid the possible loss of information, classifiers can be pooled. Diversity among individual classifiers of the team is expected to be important for effectiveness in classifier combination. The recognition rate of a combination is usually better than that of each individual classifier. Multiple classifier systems have been attempted in a variety of pattern recognition fields [1].

We examined simple fixed combining method of different constituent classifiers and bagging and boosting fusion.

Social and simple clinical data of 439 child patients from three groups of children, hospitalized at the Institute of Pulmunology in Rabka, were used to find classification functions for existing and severity of asthma. We studied two classification problems:

Classification A

group 1= non asthmatic children (n=101)

group 2= mild and moderate asthma (n=62)

group 3= severe asthma (n=176)

Classification B

group 1= non asthmatic (n=101)

group 2= asthma (n=338).

The social information was collected from family of hospitalized patients by the questionnaire.

We applied the real medical data with mixed variables and known disease diagnosis (children's asthma) as the training set for aiding diagnosis for new patients. We examined ROC curves of different discriminant methods such as: logistic regression, classification trees and neural networks with comparing theirs AUCs to the global classification errors.

Next we combined above constituent classifiers and compared the obtained curve with the curves of constituent classifiers. The analogous analysis was made on other methods of classifiers' ensemble: bagging and boosting [1].

In a lift chart [5], also known as a gains chart, for a nonbinary grouping variable, all observations from the scored data set are set in order from highest expected profit to lowest expected profit. For a binary target, the scored data set is sorted by the posterior probabilities of the event level in diminishing order. Then the individuals are grouped into deciles. Patients with actual profit values greater than the cutoff value are classified as responders (for binary targets: individuals with a posterior probability of the event level greater than or equal to 0.5). If the model has high predictive power, then the positive responses are concentrated in the highest deciles.

We analysed percent response and the percent captured response. To compute the exact model for the grouping variable, observations are sorted in diminishing order by actual profit. The exact model quickest captures all of the responses.

We compared obtained plots with the baseline (corresponding to the random classification) and with the exact model plot.

Results

The data set consisted of much incomplete information. From 90 variables we rejected the most incomplete ones, obtaining 49 variables. Next we performed single imputation of missing data and we found out features with biggest discriminant properties in the differentiation of groups: home contact with a dog in the past or actually, passive smoking by child in the past, active or passive smoking by mother during pregnancy, number of children in the family (or the number of pregnancy or the number of childbirth), pregnancy week when birth, days of staying in the hospital after birth, age of home building, kind of water heating.

For the described in presented paper asthmatic data we obtained the satisfying results. Classification effectiveness was high for different discriminant methods. Kernel discrimination with prior probabilities proportional to group sizes and with normal kernel attained the smallest assessment for the new trial error by cross-validation method. Simple fixed combining some of not such effective methods and of different construction character among them improved effectiveness of classification comparing to constituent classifiers errors.

Examples of charts for simple fixed combining and bagging or boosting are presented in figures of ROC and gain charts. We observed the improvement of combining by examining errors of classification and also curves measuring discriminability.

Discussion

Improvement of effectiveness for fusion of classifiers is confirmed by classification errors and by graphs measuring discriminability. Performance of ensemble method can be visualized in one graph, such as ROC or gain chart, and compared with the joint graph of constituent classifiers. For the case of more then 2 groups we can use other than ROC curves, measuring discriminability, such as gain charts. Similarly to ROC, cumulative lift chart for a model with good predictive power is superior over the line for a model with bad predictive power. In the contrary to ROC, they can be applied for more then 2 groups. However, the interpretation of those curves is different.

References

1. Webb AR. Statistical pattern recognition. Wiley &Sons; 2002.
2. Krzyśko M. Linear discriminant functions which maximize the area under the ROC curve. Discussiones Mathematicae. Algebra and Stochastic Methods, 1999; 19: 335-44.
3. Hsieh F, Turnbull B. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. The Annals of Statistics, 1996; 24: 25-40.
4. Ferri C, Flach P, Hernández-Orallo J. Learning Decision Trees Using the Area Under the ROC Curve. [In:] Claude Sammut and Achim Hoffmann, editors, Proceedings of the 19th International Conference on Machine Learning; Morgan Kaufmann, July 2002: 139-146.
5. SAS Enterprise Miner for Windows NT Release 4.1. SAS Institute Inc. Cary, NC, USA; 2002.