Não preencher

Combining Models in Supervised Classification: New developments

Ana Sousa Ferreira[1], Anabela Marques[2], Margarida Cardoso[3]
[1]*LEAD, Faculdade de Psicologia - UL, CEAUL e UNIDE, asferreira@fp.ul.pt;*
[2]*ESTBarreiro, Instituto Politécnico de Setúbal, anabela.marques@estbarreiro.ips.pt;*
[3]*UNIDE – ISCTE-IUL Business School, mgsc@iscte.pt.*

Abstract

In Discrete Discriminant Analysis dimensionality problems often occur. In this context, we propose a combining models approach, taking profit from several potential models. In the bi-class case, a single combination coefficient is considered and estimated using several strategies. In the multi-class case, the decomposition into several bi-class problems embedded in a binary tree is implemented. New developments of this approach are presented and their performances assessed on real or simulated data.

Key-words: Combining models; Dependence Trees Model; Discrete Discriminant Analysis**;** First-order Independence Model; Full Multinomial Model.

1. Introduction

In Supervised Classification the major goal is to derive a classification rule to assign sampling objects whose class membership is unknown.

This work focuses on Discrete Discriminant Analysis (DDA) in the small or moderate samples size setting where, often we are faced with problems of dimensionality due to the relative large number of parameters to be estimated. This dimensionality problem leads to poor performances, particularly when the classes analyzed are barely separated.

In a discrete problem of supervised classification, it is assumed that each object becomes from one of K exclusive classes, C_1, C_2,...,C_K, with prior probabilities $\pi_1, \pi_2, \ldots, \pi_K$, $\pi_l \geq 0$, and **(**$\sum_{k=1}^{K} \pi_k = 1$**).**

Each object is characterized by a multivariate vector **x** of *P* discrete variables and we dispose of a n-dimensional training sampling of multivariate observations X = ($\mathbf{x}_1$, ..., $\mathbf{x}_n$).

The Bayes classification rule assigns an individual vector **x** into C_k if

$$\pi_k P(\mathbf{x} \mid C_k) \geq \pi_l P(\mathbf{x} \mid C_l) \quad \text{for } \begin{array}{l} l = 1,\ldots,K \\ l \neq k \end{array}$$

where $P(\mathbf{x}|C_l)$ denotes the conditional probability function for the l-th class. Usually, the conditional probability functions are unknown and are estimated on the basis of the training sample.

For discrete data, the most natural model is to assume that the conditional probability function $P(\mathbf{x}|C_k)$, $k=1,...,K$ are multinomial probabilities. In this case, the conditional probabilities are estimated by the observed frequencies. Goldstein and Dillon (1978) call this model the Full Multinomial Model (FMM).

One way to deal with the dimensionality problem consists of reducing the number of parameters to be estimated. This is accomplished by the First-order Independence Model (FOIM) that assumes that the *P* discrete variables are independent within each class C_k, $k = 1, ..., K$.

In many situations several models are in competition for the same supervised classification problem, aiming to obtain the best decision rule. Usually, based on some validation criteria only a single model is selected and useful information about the supervised classification problem is lost. Besides, misclassification objects can be different for different DDA models (Brito et al, 2006).

In this context, we have proposed a new methodological approach based on a combining model (Sousa Ferreira, 2000).

The aim of this work is to present new developments of this combining models approach (Sousa Ferreira, 2000, 2010; Marques et al., 2008) and to assess their performances from numerical experiments.

2. Combining models

The idea of combining models currently appears in an increasing number of papers. The aim of this strategy is to obtain more robust and stable models. Sousa Ferreira (Sousa Ferreira, 2000) proposed a combining model approach to classification problems with binary predictors based on combining FMM and FOIM, using a single coefficient β ($0 \leq \beta \leq 1$) to weight these models.

This combining models approach revealed to be a good alternative in the bi-class case and for the multi-class case a Hierarchical Coupling model (HIERM) has been defined, decomposing the multi-class problems in several bi-class problems.

The evaluation of Sousa Ferreira's approach (Sousa Ferreira, 2000; Brito et al., 2006) showed that the coefficient derived for the combining models tends frequently to reduce FMM contribution, even when the observed frequencies are smoothed. According to these conclusions, a new combining model has been proposed (Marques et al., 2008), which is based on the models FOIM and the Dependence Trees Model (DTM).

DTM is an alternative model that takes the predictor relations into account. In this model one can estimate the conditional probability function, using a dependence tree that represents the most important predictor relations. DTM provides, for each class, an estimate of the conditional probability functions based on the idea proposed by Pearl (Pearl, 1988). To construct the dependence tree for each class, we rely on the algorithm of Chow and Liu (Celeux and Nakache, 1994; Pearl, 1988), where the length of each edge referred to one pair of variables and represents a measure of the association between the same variables, using the mutual information.

3. Hierarchical Coupling Model

The Hierarchical Coupling Model (HIERM) consist in decompose the multi-class problem in several bi-class problems, using a structure of a binary tree. HIERM demands two decisions in each level of the tree:

i. Selection of the decomposition in a bi-class problem among the possible $2^{K-1}-1$ forms;

ii. Choice of the model or combining models that gives the best classification rule for the chosen couple.

First, we consider the K classes corresponding to the training subsamples that we want to reorganise into two classes. Then, we propose either to explore all the hierarchical coupling solutions or to select the two new classes that are the most separable, which can be implemented. These most separable classes can be selected using the affinity coefficient (Matusita, 1955; Bacelar-Nicolau, 1985).

The process stops when a decomposition of classes leads to single classes.

The individual vector **x** is classified in the class determined by the tree branch chosen.

4. Numerical Experiments

In the present work, we propose to use combining models to address supervised classification problems in the small or moderate sample size setting.

We illustrate the performance of the two combining models proposed on real and simulated data sets, evaluating the misclassification error rate in a test subsample or estimated by two-fold cross validation.

References:

BACELAR-NICOLAU, H. (1985) The Affinity Coefficient in Cluster Analysis. *Meth.Oper. Res*., 53, 507-512.

BRITO, I., CELEUX, G. & SOUSA FERREIRA, A. (2006) Combining method in supervised classification: A comparative study on discrete and continuous problems. *REVSTAT - Statistical Journal*, Vol. 4(3), 201-225.

CELEUX, G. and NAKACHE, J.P. (1994) *Analyse Discriminante sur Variables Qualitatives*. G. Celeux et J. P. Nakache (Eds.), Polytechnica.

GOLDSTEIN, M., DILLON, W.R. (1978) *Discrete Discriminant Analysis*, Wiley and Sons.

MARQUES, A., SOUSA FERREIRA, A. & CARDOSO, M. (2008) Uma proposta de combinação de modelos em Análise Discriminante. IN OLIVEIRA, I. et al. (Eds.) *"Estatística - Arte de Explicar o Acaso", Ciência Estatística*, Edições S.P.E., 393-403.

MATUSITA, K. (1955) Decision Rules Based on Distance for Problems of Fit, Two Samples and Estimation. *Ann. Inst. Stat. Math*.,. 26(4), 631-640.

PEARL, J. (1988) *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Los Altos: Morgan Kaufmann.

SOUSA FERREIRA, A. (2010) A Comparative Study on Discrete Discriminant Analysis through a Hierarchical Coupling Approach. IN LOCAREK-JUNGE, H., WEIHS, C. (Eds.) *Classification as a Tool for Research,* Studies in Classification, Data Analysis, and Knowledge Organization. Springer-Verlag, Heidelberg-Berlin, 137-145.

SOUSA FERREIRA, A. (2000) *Combinação de Modelos em Análise Discriminante sobre Variáveis Qualitativas*, Tese de doutoramento, Univ. Nova de Lisboa.