A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 1(311), 2015
http://dx.doi.org/10.18778/0208-6018.311.02

*Katarzyna Stąpor*[*]

# BETTER ALTERNATIVES FOR STEPWISE DISCRIMINANT ANALYSIS

**Abstract.** Discriminant Analysis can be best defined as a technique which allows the classification of an individual into several distinctive populations on the basis of a set of measurements. Stepwise discriminant analysis (SDA) is concerned with selecting the most important variables whilst retaining the highest discrimination power possible. The process of selecting a smaller number of variables is often necessary for a variety number of reasons. In the existing statistical software packages SDA is based on the classic feature selection methods. Many problems with such stepwise procedures have been identified. In this work the new method based on the metaheuristic strategy tabu search will be presented together with the experimental results conducted on the selected benchmark datasets. The results are promising.

**Key words**: discriminant analysis, stepwise procedures, feature selection, metaheuristic, tabu search.

## 1. INTRODUCTION

Discriminant analysis (DA) (Krzyśko 1990) is a multivariate technique for classifying study instances into groups (predictive discriminant analysis, PDA) and/or for describing group differences (descriptive discriminant analysis, DDA). Discriminant analysis is widely used in many areas such as biomedical studies, banking environment (for credit evaluation), financial management, bankruptcy prediction, marketing, and many others. DA is broken into a two–step process:

– computation of a set of discriminant variables and testing their significance,

– classification.

DA computes an optimal transformation (projection) by minimizing the within–class distance and maximizing the between–class distance simultaneously, thus achieving maximum class discrimination. The optimal transformation in DA can be readily computed by applying an eigendecomposition on the $W^{-1}B$. (W and B are the so–called scatter matrices: between and within – groups covariance matrices – for details see for example (Krzyśko 1990).

[*] Full Professor, Institute of Computer Science, Silesian Technical University.

For a new observation $x$ to classify, the linear combinations $y_i = v_i^T x$, called the discriminant variables ($i$ = 1,…,s) are first computed (where $v_1,...,v_s$ are the strictly positive eigenvectors of $W^{-1}B$, scaled so that $v_i^T W v_i = 1,\ i = 1,...,s$). There are several tests of significance of discriminant variables. The multivariate Wilk's lambda test statistic is used most frequently (for details see for example (Krzyśko 1990). Then, the Fisher classifier assigns the new observation x to the group k (c is the number of groups) if:

$$D_k(x) = \min_{j=1,...,c} \quad D_j(x) \tag{1}$$

where

$$D_j^2(x) = [V^T(x-\mu_j)]^T[V^T(x-\mu_j)] - 2\log\pi_j, \qquad V = (v_1,...,v_s) \tag{2}$$

is the so–called $j$–th Fisher discriminant score, which measures the (Euclidean) distance of the observation x to the j–th group center in the new discriminant space.

Stepwise discriminant analysis **(**SDA**)** procedures are common analytic procedures used to reduce the number of variables. In the existing, commercial statistical software packages (for example one of the most popular – STATISTICA), SDA is realized as a sequential forward or backward method while looking at the Wilk's lambda for each variable, i.e. variables are entered in a stepwise fashion using Wilk's lambda criterion. In the first step of forward stepwise procedures, each variable is entered into a separate analysis, and the variable with the best univariate discrimination (lowest Wilk's lambda) is selected. Next, each remaining variable is paired with the first and entered into a separate analysis. The variable which, when paired with the first provides the best multivariate discrimination (again, the lowest Wilk's lambda) is selected next. The third step matches each remaining variable with the first two, and so on. This process is continued until either all variables are selected or the decrease in Wilk's lambda is insufficient to warrant further variable selection, as determined by the F–ratio. Despite the frequency of their use, the SDA procedures entail a number of problems which can lead to misleading and inaccurate results, especially for predictive purposes (Huberty 1989). The following group of problems has been revealed:

- Variable selection procedures. Stepwise procedures do not always select the best subset of variables of a given size. By entering variables one at a time, stepwise procedures do not include all of the information supplied jointly by two or more variables not already included in the analysis.

- Capitalization on sampling error. Stepwise procedures are especially suspect to sampling error due to the fact that they select the variable with the lowest Wilk's lambda to be entered, no matter how small the difference.
- Selection criteria. Stepwise procedures are designed not for PDA, but for DDA. This distinction is important – in DDA a completely worthless variable would be given a weight zero while in PDA it would contribute "noise" to the prediction analysis, making group prediction less accurate.

This work proposes the new method for variable selection in discriminant analysis based on the metaheuristic strategy tabu search which can cope with the problems of stepwise procedures. The use of tabu search for feature selection in classification has already been reported, for example in (Zhang 2002), but there are very few key references (Pacheco 2006) on the selection of variables for their use in discriminant analysis. In comparison with Pacheco (2006) our method utilizes different representation of a solution as well as different definition of a neighborhood, the two important concepts of tabu search metaheuristic.

The remainder of this paper is organized as follows. Section II presents our solution to the mentioned problems of SDA. The results of the tests on the proposed method are presented in section III, and a short conclusion follows in section IV.

## 2. THE SOLUTION APPROACH: TABU SEARCH ALGORITHM

Stepwise discriminant analysis can be formulated as a feature selection problem in pattern recognition (Stąpor 2011; Kohavi 1997) which is the process of selecting a subset of relevant features for their use in the model construction. An exhaustive approach to feature selection problem would require examining all possible subsets of the feature set which grows exponentially. For bigger values of feature set size d, the explosive computational cost makes the exhaustive search impracticable. Thus, the main stream of feature selection research was directed towards suboptimal, but efficient methods. For an overview of feature selection methods see for example (Kohavi 1997). To conclude, although some progress has been obtained, the available feature selection techniques for large feature sets are not yet completely satisfactory. They are either computationally feasible but far from optimal, or they are optimal or almost optimal but cannot cope with the computational complexity of feature selection problems of realistic size.

In this paper, we introduce the use of metaheuristic tabu search method (Glover 1989) for feature selection in discriminant analysis. Metaheuristics (Blum 2003) are a new kind of approximate algorithms that try to combine basic heuristic methods in higher level frameworks aimed at efficiently and effectively exploring a search space. In our algorithm the short–term memory is implemented as a tabu list to record and guide the process of the search, i.e. that keeps track of the most recently visited solutions and forbids moves toward them.

Each solution (i.e. the feature subset of size p) in our algorithm is represented as a vector of length d with the 0/1 element in a position i, indicating that a feature $i$ ($i = 1, 2,\ldots, d$) is not/is included in a subset. The neighbourhood $N(x)$ of a solution $x$ is a set of solutions which are generated through adding randomly one feature on $x$ meanwhile removing one feature. The neighborhood of the current solution is restricted to the solutions that do not belong to the tabu list. The initial solution is generated randomly, but it must have exactly the required number of features. The objective function value $f(x)$ of a solutions (i.e. a feature subset) is defined as a percentage of hits on a given dataset obtained through the features of s with Fisher's classifier. The termination condition is a predefined number of iterations.

The functioning of our complete tabu search algorithm for feature selection is outlined as follows:

***

**Tabu search algorithm for PDA**

(1) <u>Initialize</u>.

Generate an initial solution $x$. Let $Sb = x$, $k = 1$, $TL = \varnothing$

/* $Sb$ – the best solution obtained so far */

(2) <u>Generate neighborhood</u>

Generate neighborhood $N(x)$ of $x$

(3) <u>Move</u>

a) If $N(x) = \varnothing$ go to step (2), otherwise find out the best solution $y$ in $N(x)$.

b) If $y$ is in tabu list and $f(y)$ is not better than $f(Sb)$, let $N(x) = N(x) - \{y\}$, go to 3a), otherwise let $x = y$, $Sb = y$ if $y$ is better than $Sb$

(4) <u>Output.</u>

If termination condition is reached, stop and output $Sb$, otherwise add the new solution $x$ to the tail of tabu list and if the length of the list exeeds a predefined size, remove the head of a list, let $k = k + 1$, go to (2).

***

## 3. EXPERIMENTAL RESULTS

To check the efficacy of the proposed feature selection method, an experiment was run with the selected datasets from the well–known data repository of the University of California, UCI: *hepatitis,Indian–diabetes, liver–disorders, spectf–heart* and *spambase* datasets.

Simple preprocessing including translating values of ordered and categorical valued attributes was performed first on these datasets. The conducted experiment consisted of comparing the performance of the Fisher classifier on a subset of features selected by the proposed tabu search–based feature selection algorithm with the performance obtained using stepwise procedures as implemented in the well known statistical package STATISTICA (i.e. *stepwise forward* procedure in the *Discriminant Analysis* module of STATISTICA).

The proposed feature selection algorithm was run for the predefined values of the dimensionality $p$, starting with $p = 1$ until the number $d$ of features in a given dataset. For each value of $p$ we have noted down the best performance of Fisher classifier obtained through the predefined number of iterations, in our case 100. Tabu length list was set to $l = 30$.

In Table 1 we can see the comparison of the obtained performances (in %) of the Fisher classifier with our tabu search and forward stepwise/STATISTICA feature selection methods respectively, on the train datasets. The column (perf.all) presents the classification performance of Fisher classifier obtained with all features, the column (perf. stepwise) – the performance of the subset obtained by forward stepwise procedure implemented in STATISTICA, the column (best perf tabu) – the performance of the best solution obtained with the proposed feature selection algorithm, the column (#best subset) – gives the associated number of features comprising the best subset.

Table 1. The comparison of our tabu search and forward stepwise/STATISTICA methods

| Data set | Perf. all | Perf. stepwise | Best perf. tabu | # best subset |
|---|---|---|---|---|
| hepatitis | 23.33 | 74.32 | 80.03 | 11 |
| Indian–diabetes | 74.68 | 75.12 | 78.32 | 5 |
| liver–disorders | 59.42 | 66.02 | 70.26 | 5 |
| spectf–heart | 50.03 | 71.41 | 76.91 | 22 |
| spambase | 59.26 | 78.00 | 84.89 | 27 |

Source: own elaboration.

From Table 1 it can be seen that our feature selection algorithm improves the solutions obtained with stepwise procedures for feature selection implemented in software package STATISTICA for any case.

## 4. CONCLUSIONS

This work approaches the problem of variable selection in discriminant analysis. In fact, the best known statistical packages continue to use classic selection methods like sequential forward/backward suffering from the nesting effect. Moreover, due to the criterion used for the evaluation of feature subsets – Wilk's lambda, they are designed for descriptive discriminant analysis only, not for predictive one.

We proposed the new feature selection algorithm based on metaheuristic tabu search that could be used instead of stepwise procedures for selecting input variables in discriminant analysis modules from the existing statistical packages. After performing some tests, it is found that our tabu search–based feature selection algorithm obtained better results than stepwise forward procedures implemented in STATISTICA package for stepwise discriminant analysis.
The presented feature selection algorithm based on tabu search could be further improved, for example by using more elaborated intensification for exploring the regions where the best solutions have been found up to this moment, as well as a diversification of the search, i.e. directing the search towards unexplored regions. This will be the subject of the future research.

### LITERATURE

Blum Ch., Roli A. (2003), *Metaheuristics in combinatorial optimization: overview and conceptual comparison*, ACM Computing Surveys, vol. 35, 3, p. 268–308.

Glover F. (1989), *Tabu Search*. Part I, ORSA Journal of Computing, v.1, p. 190–206.

Huberty C.J. (1989), *Problems with stepwise methods – better alternatives*, in: Thompson B. (ed.) Advances in Social Science Methodology, vol.1, 43–70, Greenwich, CI: JAI Press.

Kohavi R., John G. (1997), *Wrappers for feature subset selection*, Artificial Intelligence, vol. 97, 1–2, p. 234–273.

Krzyśko M. (1990), *Discriminant analysis*, WNT, Warszawa (in Polish).

Murphy P.M., Aha D.W. (1994), *UCI repository of machine learning*. University of California, Department of Information and Computer Science, http://www.ics.uci.edu/—~mlearn/ MLRepository.html.

Pacheco J. et al. (2006), *Analysis of new variable selection methods in discriminant analysis*, Computational Statistics&Data Analysis, vol.51, 3, p. 1463–1478.

STATISTICA – package documentation, (2005), StatSoft Inc.

Stąpor K. (20111), *Metody klasyfikacji obiektów w wizji komputerowej*, PWN, Warszawa.

Zhang H, Sun G. (2002), *Feature selection using tabu search method*, Pattern Recognition, 35, p. 701–711.

*Katarzyna Stąpor*

**ALTERNATYWNE, LEPSZE METODY W KROKOWEJ ANALIZIE DYSKRYMINACYJNEJ**

**Streszczenie**. Analiza dyskryminacyjna to jedna z metod umożliwiających klasyfikację obserwacji do jednej z predefiniowanych klas na podstawie wartości pomierzonych cech. Celem krokowej analizy dyskryminacyjnej (KAD) jest wybór podzbioru cech wejściowych przy zachowaniu możliwie dużej mocy dyskryminacyjnej. Zmniejszenie wymiarowości wejściowej przestrzeni cech jest konieczne z wielu powodów. W istniejących na rynku, komercyjnych pakietach do obliczeń statystycznych, KAD bazuje na klasycznych metodach selekcji cech. Metody te generują wiele problemów. W prezentowanej pracy zostanie przedstawiona alternatywna metoda wykorzystująca metaheurystykę przeszukiwania z tabu. Wyniki eksperymentalne na wzorcowych zbiorach danych są obiecujące.

**Słowa kluczowe**: analiza dyskryminacyjna, procedury krokowe, selekcja cech, meta–heurystyka, przeszukiwanie z tabu.