Computer and Automation Institute Hungarian Academy of Sciences
KÖZLEMÉNYEK 28/1982

STATISTICAL AND PATTERN RECOGNITION PROGRAM PACKAGES FOR MEDICAL DIAGNOSIS AND PROGNOSIS AT THE SEMMELWEIS UNIVERSITY OF MEDICINE

J. BAK

Semmelweis University of Medicine,
Computing Group

Medical diagnosis and prognosis based on laboratory tests and physical measurement are one of the most intriguing task of the biomedical application of pattern recognition techniques, discriminant and cluster analysis methods. These methods are in the field of mathematical statistics and a lot of statistical program packages have some classification algorithms themselves.

At Semmelweis University of Medicine we adapted a general purpose statistical program package, BMDP [1] /Biomedical Computer Programs, 1977/, for an ESZR-1020 computer. This program package contains statistical programs in the first place, however, among multivariate methods there are discriminant analysis and cluster programs for clustering cases and variables too. Among the programs there are some methods for data preprocessing, data reduction or transformation. These methods may be very important from the point of view of feature extraction. /It is to be noted that the methods the features are obtained by, are often intuitive and empirical, and their efficiency depends on the designer's knowledge and experience in the problem./

Besides BMDP we often apply SSP [2] /Scientific Subroutine Package, IBM/. It contains about 250 subroutines, including factor analysis and discriminant analysis programs which are essential for pattern recognition.

These program packages usually have no supervised learning algorithms and these are only a few programs for unsupervised learning methods too.

In mathematical pattern recognition we want to get a decision rule which can classify examples of patterns quickly. A pattern recognition problem thus begins with class definition and labelling samples of the classes in some workable representations. The problem is solved when a decision rule has been derived to assign a unique label to new patterns.

For the statistical pattern recognition techniques we developed a program package, GALAXY [3]. It contains linear discriminant function, committee-machine algorithm /piecewise linear discriminant function/, first k-nearest neighbour decision rule, and potential function methods. The goals of such techniques are to identify the parameters /if any/ which can qualitatively distinguish the known groups and to select, if possible, a classification rule for identifying the known groups.

Besides learning algorithms, GALAXY contains some basic statistical programs for computing mean value, standard deviation, kurtosis,skewness, correlation coefficients, regression and empirical distribution. We don't require a rigorous data structure, i.e. data may be given on cards, magnetic tape or disc and there are no restrictions for the format of the data either.

This program can be operated with control cards.

GALAXY has been effectively applied for medical diagnosis and prognosis tasks, such as the prognosis for myocardial infarction patients [8], classification of sleep stages [9,10], and studying of acute cerebrovascular diseases.

We modified and adapted a nonhierarchical cluster program, DIDAY [4] developed at SZÁMKI /Budapest/ for an ESZR-1020 computer.

Most analysis techniques assume the homogenity of the variables, whereas real data sets often have mixed variables. Our methods are applicable for continuous variables only. We have no methods for discrete, binary or dichotomous variables, although cases are frequently described by such variables.

/For instance, classifying the cases on the basis of medical records or giving parameters by coded data./

That is why we would like to complete our program packages with discrete discriminant analysis methods /W. Goldstein, [5]/ for solving problems of discrimination between groups with discrete multivariate observations, and with some hierarchical clustering algorithms for binary variables /SYN-TAX, [7]/. We have some experience with the medical applications of the program package ARTHUR [6] too, what was developed originally for chemical applications by L. Kowalsky.

REFERENCES

[1] Dixon, W.J.; Brown, M.B.: Biomedical Computer Programs, P-Series, University of California Press, 1977, Berkeley.

[2] System/360 Scientific Subroutine Package, Version III, Programmer's Manual /IBM/, 1968.

[3] Bak, J.; Szadeczky-Kardoss, G.: Program Package for Classification Algorithms. Dokumentation /GALAXY/. March, 1977, SOTE.

[4] Diday, E.; Schroeder, A.: The Dinamic Cluster Method in Pattern Recognition, Information Processing 74.

[5] Goldstein, M.; Dillin, W.R.: Discrete Discriminant Analysis. John Wiley and Sons, New York, 1978.

[6] Harper, A.M.; Duewer, D.L.; Kowalski, B.R.: ARTHUR and Experimental Data Analysis: The Heuristic Use of a Polyalgorithm. Chemometrics: Theory and Application, 1970.

[7] Podani, J.: Computer Program Package for Cluster Analysis in Ecology, Phytosociology and Taxonomy. Abstracta Botanica, Tomus VI., 1980, Budapest.

[8] Jánosi, A.; Bajkai, G.; Bak, J.: Prognosis of Patients with Myocardial Infarction Treated in an Intensive Coronary Care Unit. Cardiologica Hungarica, 1979.

[9] Pál, I.; Bak, J.; Halász, P.; Rajna, P.; Kundra, O.: Wide-band Spectra of Different SWS2 Stages are Different. In: Sleep, 1978. L. Popovics /Ed./, S. Karger, Basel. 4th Europ. Congr. of Sleep Research.

[10] Bak, J.; Pál, I.; Halász, P.; Rajna, P.: Cluster Analysis of Broad-Band EEG Spectra in Identical Sleep Stages. Magyar EEG Társ. XXIV. évi tud. ülése, Debrecen, 1980. /XXIV. Congress of the Hungarian EEG Society, 1980, Debrecen/

ÖSSZEFOGLALÁS

J. Bak

Az orvostudomány számos területén, akár a differenciál diagnózis témakörében, akár a diagnóziskészitès folyamatának vizsgálatában felmerülnek olyan osztályozási vagy csoportba sorolási problémák, melyek igénylik az alakfelismerési módszerek alkalmazását. A SOTE-n kidolgoztunk egy alakfelismerési programokat tartalmazó rendszert, mely elsősorban tanuló algoritmusokat tartalmaz. Cluster analizis témakörből mások által fejlesztett programokat vettünk át és alkalmaztunk gyakorlati feladatokra. Néhány évvel ezelőtt adaptált statisztikai programrendszereink, mint a SSP, vagy BMDP is tartalmaznak olyan többváltozós módszereket /hierarchikus cluster algoritmusok változókra és esetekre, diszkiriminancia és faktoranalizis/, melyek ezirányu munkáinkat nagymértékben elősegitik. A módszerek gyakorlati alkalmazásait illetően a prognózis területén a myocardiális infarctus lefolyását vizsgáltuk a kórházi időtartam /28 nap/ alatt, a diagnosztika területén pedig az akut cerebrovascularis kórképeket tanulmányoztuk. Jelananalizis témakörben az alvásfázisok automatikus felismerése és az alvás folyamat alatti fáziskülönbségek kimutatása jelentette sikeres alkalmazását az osztályozási módszereknek.

ПАКЕТЫ СТАТИСТИЧЕСКИХ ПРОГРАММ И ПРОГРАММ РАСПОЗНАВАНИЯ ОБРАЗОВ ДЛЯ МЕДИЦИНСКОЙ ДИАГНОСТИКИ И ПРОГНОЗА В МЕДИЦИНСКОМ УНИВЕРСИТЕТЕ им. И. СЕММЕЛЬВЕЙСА

Содержание

Во многих областях медицины, в том числе в области дифференциальной диагностики или при исследовании диагностического процесса возникают проблемы классификации и группировки объектов, требующие применения методов распознавания образов. В Медицинском Университете им. И.Семмельвейса разработана система программ распознавания образов, содержащая в первую очередь программы, реализующие обучающиеся алгоритмы. В области кластер-анализа применили для решения практических задач программы, взятые у других в готовом виде. Многомерные методы (как иерархический кластер-анализ переменных и событий, дискриминантный, факторный анализ) содержатся также в пакетах статистических программ /SSP, BMDP/ , адаптированных у нас в последние годы, что значительно облегчает нашу работу в этом направлении. На практике эти методы были применены в решении следующих задач: в области прогноза исследовали течение инфаркта миокарда во время пребывания больного в стационаре (28 суток), а в области диагностики исследовали различные формы острых цереброваскулярных заболеваний. В области анализа биосигналов методы классификации успешно применились при автоматическом распознавании фаз сна, и при выявлении различий между фазами сна.