ELSEVIER

Chemometrics and Intelligent Laboratory Systems 61 (2002) 123–132

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Feature selection in principal component analysis of analytical data

Q. Guo [a], W. Wu [b], D.L. Massart [a,*], C. Boucon [c], S. de Jong [c]

[a] *ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b] *Safety Assessment, SmithKline Beecham Pharmaceuticals, The Frythe, Welwyn, Hertfordshire, AL6 9AR, UK*
[c] *Unilever Research Vlaardingen, Olivier van Noortlaan 120, 3133 AT Vlaardingen, Netherlands*

Received 8 November 2000; received in revised form 27 September 2001; accepted 12 October 2001

Abstract

A feature selection method is proposed to select a subset of variables in principal component analysis (PCA) that preserves as much information present in the complete data as possible. The information is measured by means of the percentage of consensus in generalised Procrustes analysis. The best subset of variables is obtained by applying a genetic algorithm (GA) to optimise the consensus between the subset and the complete data set in order to avoid exhaustive searching. The method was evaluated on a standard data set known as the Alate data, and on a high-dimensional industrial gas chromatography (GC) data set. The results showed that the proposed method successfully identified structure-bearing variables in both data sets and that it leads to a better subset of variables than other studied feature selection methods.

Keywords: Feature selection; Principal component analysis; Genetic algorithm; Generalised procrustes analysis; Data mining; Gas chromatography

1. Introduction

In industrial practice, it becomes more and more often necessary to handle huge sets of measurement data. We are interested in the analysis of large data sets, which may consist of over a thousand gas chromatograms each sampled at thousands of retention times. Even when the raw chromatogram is converted into a peak table, we have to deal with more than a hundred peaks. Such a large number of variables leads to multicollinearity, and to redundancies among the variables. Consequently, it makes it more difficult to reveal patterns in the data. Advanced data mining techniques that can deal with such difficulties become more and more important [1,2]. In this article, we develop methods that allow us to better understand the structure of large sets of chemical data.

Principal component analysis (PCA) has been widely applied in data mining to investigate data structure. In PCA, new orthogonal variables (latent variables or principal components) are obtained by maximising variance of the data. The number of the latent variables (factors) is much lower than the number of original variables, so that the data can be visualised in a low-dimensional PC space. While PCA greatly reduces the dimensionality of the space, it does

* Corresponding author. Tel.: +32-2-477-47-37; fax: +32-2-477-47-35.
E-mail address: massart@vub.vub.ac.be (D.L. Massart).

0169-7439/02/$ - see front matter
PII: S0169-7439(01)00203-9

not reduce the number of the original variables, as it uses all the original variables to generate the new latent variables (principal components). For interpretation purposes or future investigations, it would often be very useful to reduce the number of variables. Feature (variable) selection can be achieved either by choosing informative variables or discarding redundant variables. Several methods exist and most of them perform feature reduction using stepwise forward and/or backward techniques. Jolliffe [3–5] compared a number of methods, mainly working on preserving most of the variation of the data. McCabe [6] developed methods to keep as much information as possible by optimising four numerical criteria. Malinowski [7] selected a key set of variables being as orthogonal as possible. Rännar et al. [8] chose variables that span the original space as well as possible by a combination of PCA and PLS. In data mining, it is of importance to select a small subset of variables that can reproduce as closely as possible the structure of the complete data. Krzanowski [9] developed such a method based on Procrustes analysis. As the method seeks variables by a stepwise procedure (backward elimination), there is no guarantee to obtain the best global subset. Moreover, with hundreds or thousands variables, as is often the case in data mining, intensive calculation is needed to perform PCA in each elimination step. In this paper, a method is presented that applies a genetic algorithm to search for the best subset instead of a backward elimination procedure. The percentage of consensus in generalised Procrustes analysis is introduced to facilitate interpretation and comparison of data structures.

2. Theory

2.1. Notation

$\mathbf{X}$	$n \times p$, a data matrix with n objects (rows) and p variables (columns)
$\mathbf{X}_s$	$n \times q$, a subset data matrix with n objects and q selected variables
$\mathbf{S}$	$n \times k$, the PC score matrix corresponding to/ calculated from $\mathbf{X}$
$\mathbf{S}_s$	$n \times k$, the PC score matrix corresponding to/ calculated from $\mathbf{X}_s$
$\mathbf{Y}$	$n \times k$, the pre-scaled matrix of $\mathbf{S}$ (total variance of $\mathbf{S}' = 50$)
$\mathbf{Y}_s$	$n \times k$, the pre-scaled matrix of $\mathbf{S}_s$ (total variance of $\mathbf{S}_s' = 50$)
$\mathbf{Y}''$	$n \times k$, the resultant matrix from Procrustes analysis on $\mathbf{Y}$ or $\mathbf{S}$
$\mathbf{Y}_s''$	$p \times k$, the resultant matrix from Procrustes analysis on $\mathbf{Y}_s$ or $\mathbf{S}_s$
$\mathbf{\Sigma}$	$k \times k$, the diagonal singular-value matrix of $\mathbf{S}' \mathbf{S}_s$

2.2. Feature selection methods to preserve variation

A common way to locate important original variables is based on loadings in PCA. For any factor, high loadings in absolute value indicate that corresponding variables contribute more to the factor than other variables. In PCA, usually the first few, e.g. k, PCs are regarded as significant. For each of the first k PC factors, one or a few variables with the largest loadings are selected. The correlation coefficients between PC scores and the original variables can also be used to find subset of variables. Jolliffe [3–5] proposed two such methods: (i) find one variable having closest association with each of the last $p-k$ PC factors, and delete those variables; (ii) associate one variable with each of the first k PC factors, and retain these variables.

2.3. Principal variables methods

The criterion in the above methods is to keep as much variance of the data in the subset. McCabe [6] recommended four criteria to select variables which he called principal variables, namely (i) maximising the determinant of the covariance matrix of the selected variables, (ii) minimising the trace of the covariance matrix of the variables eliminated, (iii) minimising the squared norm of the conditional covariance matrix of the variables eliminated, and (iv) maximising the squared sum of the canonical correlations between the variables eliminated and those selected.

For the first criterion, i.e. maximising the determinant of the information matrix ($\mathbf{X}_s' \mathbf{X}_s$), McCabe [6] developed an algorithm which enabled evaluation of all possible subsets provided the number of variables

(p) is not greater than 20. In this study, our aim is to develop a methodology for high-dimensional multivariate data. However, when dealing with such data, an exhaustive search becomes inapplicable. Therefore, a forward selection [10] and the use of a genetic algorithm [11,12] were investigated.

2.4. Key set factor analysis

Key Set Factor Analysis (KSFA) was developed by Malinowski [7], and applied to problems in nuclear magnetic resonance, gas–liquid chromatography, and mass spectrometry [7,13,14]. The method performs forward selection. The variable with PC1 loading closest to zero is selected as the first key variable. The second variable is the one which is most orthogonal to the first selected variable. The third is the most orthogonal to the plane defined by the first two selected variables, etc. This selection procedure is repeated until a desired number (q) of variables is attained.

2.5. PCA and Procrustes analysis

Krzanowski [9] combined PCA with Procrustes analysis to select variables to preserve multivariate data structure, i.e. distances between objects should be preserved as much as possible. The basic principle of the method is to keep the structure of the subset as close as possible to that of the whole data set in PC space. The similarity/dissimilarity of the compared structures is quantified by a Procrustes criterion, i.e. the sum of squares of differences between two data sets after optimal matching. The number of dimensions (k) in the PC space can be determined by various methods, of which cross-validation [15], Malinowski indication function [16], reduced eigenvalue [16], permutation test [17], scree-plot, [17] and considerations of practical interpretation are commonly used [9]. Fig. 1a shows the procedure for the estimation of the Procrustes criterion, where $\mathbf{X}_s$ denotes a subset of $\mathbf{X}$ ($n \times p$) with q selected variables ($k \leq q < p$). $\mathbf{S}$ and $\mathbf{S}_s$ are the score matrices with the first k factors after PCA of $\mathbf{X}$ and $\mathbf{X}_s$, respectively. To measure the discrepancy between the two configurations, $\mathbf{S}$ and $\mathbf{S}_s$ are subjected to Procrustes analysis [18], with $\mathbf{S}$ as the target matrix. After the two matrices undergo translation, rotation, and reflection, output matrices

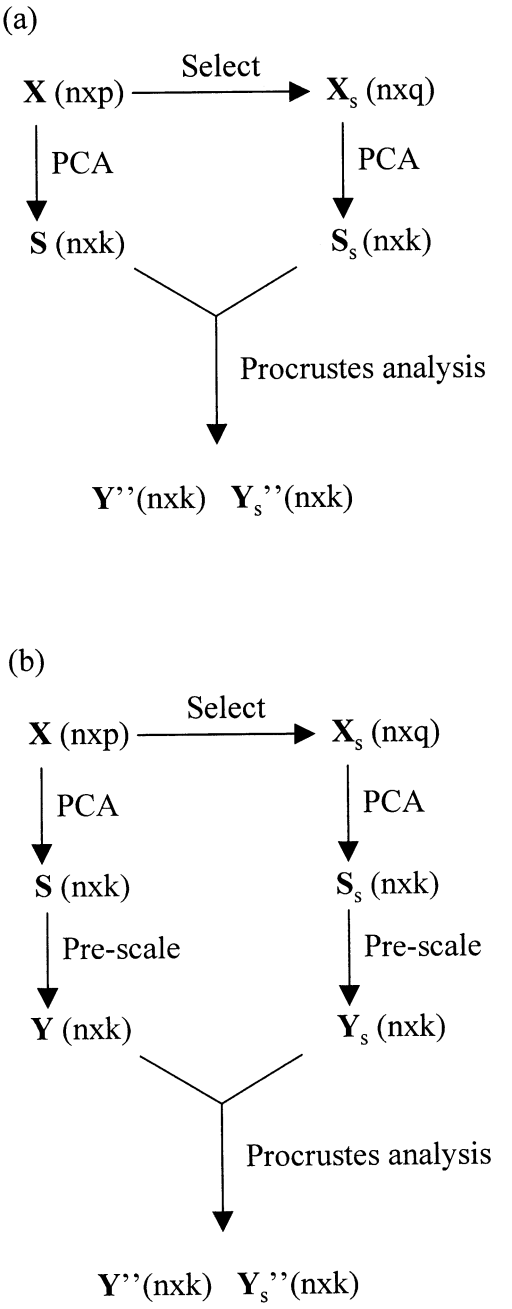

Fig. 1. Flow charts of (a) the Krzanowski method and (b) the proposed method.

$\mathbf{Y}''$ and $\mathbf{Y}_s''$, respectively are produced for the purpose of optimal match. The residual sum of squared differences between the corresponding two configurations

$\mathbf{Y}''$ and $\mathbf{Y}_s''$ is used as the criterion to measure the difference in data structure when the q variables substitute all p variables. To obtain the residual sum of squared differences efficiently, the following formula [9] was suggested avoiding calculation of $\mathbf{Y}''$ and $\mathbf{Y}_s''$:

$$D = \text{trace}(\mathbf{S}'\mathbf{S} + \mathbf{S}_s'\mathbf{S}_s - 2\mathbf{\Sigma}) \quad (1)$$

where $\mathbf{\Sigma}$ is the diagonal singular-value matrix of $\mathbf{S}'\mathbf{S}_s$.

The best subset is the one giving the lowest D among all possible subsets of q variables. In practice, exhaustive search of all possible subsets may not be computationally feasible, especially with large data sets. Krzanowski [9] designed an efficient algorithm to perform backward elimination. The procedure is summarised as follows:

1. apply PCA on $\mathbf{X}$ to obtain $\mathbf{S}$ for k significant factors;
2. delete a variable from $\mathbf{X}$ to obtain $\mathbf{X}_s$, and calculate $\mathbf{S}_s$ of $\mathbf{X}_s$;
3. calculate D by Eq. (1);
4. repeat steps 2–3 until each variable in $\mathbf{X}$ has been excluded once;
5. delete the variable in which exclusion resulted in the smallest D;
6. go back to step 2 with $p-1$ variables in $\mathbf{X}$, and repeat steps 2–5 iteratively until eventually, only q variables are left in $\mathbf{X}$.

To improve the computation speed, Krzanowski applied the updating algorithms by Bunch et al. [19,20].

2.6. Proposed method

In Krzanowski's method, the Procrustes criterion represents loss of the structural information in a candidate subset, so that different subsets from the same studied data set can be compared to decide the best subset. However, it does not explain how much structural information has been preserved by each candidate subset. For instance, there are three candidate subsets with Procrustes criteria of 239, 258, and 267, respectively. One can conclude that the first subset is the best and the third is the worst, but one does not know the percentage of structural information preserved by the first subset. It is possible that it only retains 20% of the structural information, and as a result, this best subset still does not contain enough structural information. A specific criterion is required to judge whether the best subset represents enough structural information. In the proposed method, the consensus concept from generalised procrustes analysis (GPA) [21,22] is adopted. An additional prescaling step (Fig. 1b) is introduced to pretreat $\mathbf{S}$ and $\mathbf{S}_s$, so that both configurations ($\mathbf{Y}$ and $\mathbf{Y}_s$) have variances equal to 50, and the total variance is 100. The total variance is then decomposed into two parts: variance of consensus and variance of difference between the two configurations ($\mathbf{Y}''$ and $\mathbf{Y}_s''$),

$$V_{\text{total}} = V_{\text{consensus}} + V_{\text{difference}} \quad (2)$$

where $V_{\text{consensus}}$ is the percentage of agreement between the two matrices $\mathbf{Y}''$ and $\mathbf{Y}_s''$. $V_{\text{consensus}}$ measures the structural information preserved by the candidate subset. The consensus can be calculated as

$$V_{\text{consensus}} = 2(\lambda_1 + \lambda_2 + \ldots + \lambda_k) \quad (3)$$

where λ_1, λ_2, …, λ_k are the singular values of the matrix $\mathbf{Y}'\ \mathbf{Y}_s$. $V_{\text{consensus}}$ is a relative value expressed as a percentage. In the extreme case where the subset preserves all the structural information of the original data, the $V_{\text{consensus}}$ is equal to 100. Compared with the Krzanowski's Procrustes criterion (D), the proposed criterion ($V_{\text{consensus}}$) is easier to interpret since it is expressed by a percentage. D, which is equivalent to $V_{\text{difference}}$, measures the loss of the structural information in a candidate subset, while $V_{\text{consensus}}$ represents the remains of the structural information. Both criteria do not affect the subset selection since total structural information is constant and they rank candidate subsets in an equivalent way.

The backward elimination procedure generally works effectively, but can sometimes lead to a local optimum. Moreover, when p becomes large, intensive computations are required. In order to efficiently search for the best subset, a genetic algorithm (GA) is applied.

GA is a very powerful means to search for a global optimum in a high-dimensional space [23,24]. We applied the algorithm developed by Leardi et al. [11], which was originally designed to select the best subset of variables in multiple linear regression and was used for feature selection in sequential projection pursuit

[12]. The algorithms are detailed in Refs. [11,12]. The description here focuses on how it is adapted for feature selection in PCA. In GA, each subset of variables is represented as a string vector, which contains the same number of elements as the number of total variables. Each element is coded as 1 if the variable was chosen and 0 if otherwise. The initial population of solutions consists of a number of strings (e.g. 30), which were randomly created and evaluated by measuring their fitness. The fitness of the string indicates the amount of structural information preserved by the related subset of those chosen variables, evaluated by Eq. (3). Pairs of parent strings are selected according to a probability value, which is proportional to the quality of their fitness. Then they undergo 'reproduction' by applying a crossover operator (e.g. probability 0.5) and, to a lesser extent, a mutation operator (e.g. probability 0.02). This results in two offspring strings for each pair of parents. They become new candidate subsets and join the population to construct a new generation. This procedure is repeated a certain number of times (e.g. 500 epochs) specified by generations. In GA [11], one should input the maximal number of variables in the subset. In the output generation, there are subsets with different sizes. Since large-size subsets tend to have high consensus, the best subset always has the maximal number of variables. In the proposed algorithm, only small subsets of q variables are involved in the calculation of the PC scores of $\mathbf{X}_s$, so that updating algorithms [19,20] are not necessary.

3. Experimental

3.1. Data

3.1.1. Alate Adelges data

The data set was first given by Jeffers [25] as a case study of PCA. Later, it was used by Jolliffe [4,5] to compare different feature selection methods. Krzanowski [9] also applied it to testify his new method. The data consist of 19 variables measured on 40 alate adelges (winged aphids). The details are described in Ref. [25]. The full data matrix (40×19) can be found in Ref. [9]. The aim of the investigation was to diagnose variants of aphids. Jeffers [25] concluded that there are four major groups in the data.

3.1.2. GC data from food chemistry

A set of gas chromatography (GC) data [2] of product mixtures from a complex chemical reaction was measured. The reaction was carried out at two different pH levels. Reactions were carried out on one of six sugars: fructose, glucose, lactose, maltose, rhamnose, and xylose, and one or two of nine amino acids: alanine, asparagine, glutamine, glycine, threonine, arginine, cysteine, lysine, and glutamate. The data set comprises 228 chromatograms and peak areas for 199 peaks.

4. Results and discussion

4.1. Alate Adelges data

Subjecting the standardised data to the cross-validation technique [15] showed that four PC factors are necessary to model the data. They account for 92% of the total variance. As in all other feature selection methods, the number of selected variables must be predefined. Usually, it is decided according to experience. Alternatively, one may try to decide it according to a penalty function. In order to compare with the results in the literature [4,5,9], the number of four variables are selected to represent all 19 variables. Table 1 lists the results for different feature methods. Based on PC loadings, two subsets are selected. One

Table 1
The subset of selected variables and consensus value for different feature selection methods of the Alate Adelges data

Method	Four variables	Consensus
PC loading (1st)	13, 17, 11, 5	88.77
PC loading (:, 1:2)	13, 12, 17, 5	91.43
$\mathbf{X}_s' \mathbf{X}_s$ (forward)	10, 11, 17, 19	88.36
$\mathbf{X}_s' \mathbf{X}_s$ (GA)	9, 11, 17, 19	87.44
Jolliffe (i)	5, 8, 11, 14[a]	91.80
Jolliffe (ii)	5, 11, 13, 17[a]	88.77
McCabe (i)	9, 11, 17, 19[a]	87.44
McCabe (ii)	5, 9, 11, 18[a]	88.80
McCabe (iii)	6, 11, 17, 19[a]	87.52
McCabe (iv)	5, 8, 11, 18[a]	88.29
Krzanowski	5, 12, 14, 18[a]	94.47
Key set	18, 8, 11, 5	88.29
Proposed	3, 4, 14, 16	95.09
All-subset	3, 4, 14, 16	95.09

[a] From Refs. [4,5,9].

of those was obtained by selecting one variable on each of the first four PC and another by selecting two variables on each of the first two PCs. The subsets obtained from Jolliffe's (Section 2.2), McCabe's (Section 2.3), and Krzanowski's methods are quoted from Refs. [4,5,9]. The results show that all these methods selected some identical variables. Certain variables are frequently chosen by different methods. In general, all those subsets give quite high consensus ranging from 87% to 95%, indicating that most of the structural information is preserved. The method proposed by us gives the highest consensus. Since there

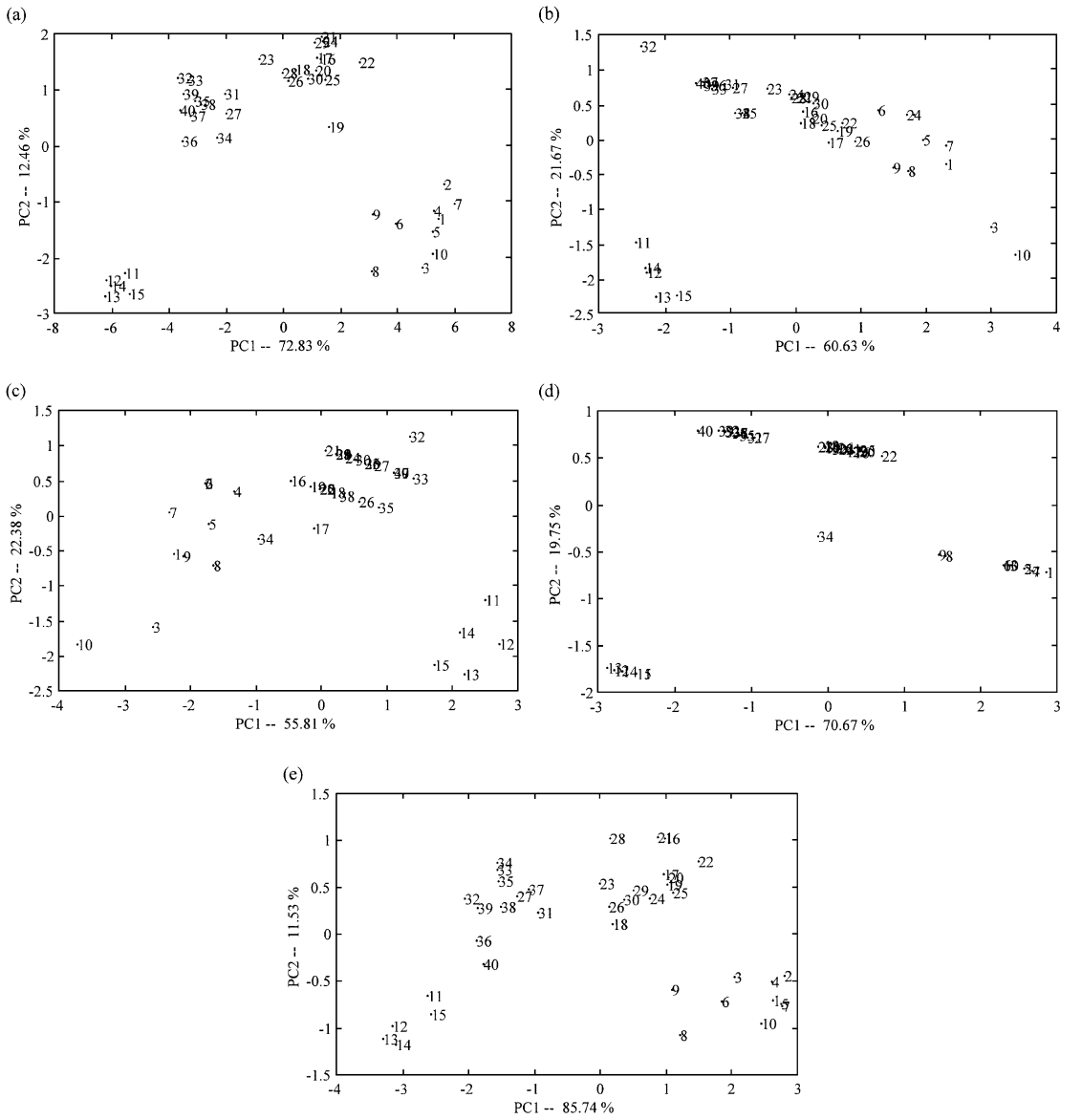

Fig. 2. Score plot of the Alate data using (a) all 19 variables, (b) the four variables (5, 8, 11, and 14) selected by the Jolliffe (i) method, (c) the four variables (5, 9, 11, and 18) selected by the McCabe (ii) method, (d) the four variables (5, 12, 14, and 18) selected by the Krzanowski method and (e) the four variables (3, 4, 14, and 16) selected by the proposed method.

Table 2
The consensus value for different feature selection methods of the GC data

Method	Five variables	10 variables	20 variables
PC loading (1st)	66.56	60.39	69.16
PC loading (:, 1:5)	66.56	74.02	86.47
$\mathbf{X}_s' \mathbf{X}_s$ (forward)	57.28	67.12	75.95
$\mathbf{X}_s' \mathbf{X}_s$ (GA)	51.44	63.98	80.35
Key set	51.44	43.32	60.37
Proposed	71.54	81.23	89.68

are only 19 variables, the all-subset section method was also used and the selected subset is the same as with the proposed method. The second best is Krzanowski's method, and the worst is McCabe's method using the determinant criterion. The method based on the largest loading of the first four PCs gives the same subset as the Jollife (ii) method. The former is based on loadings and the latter on correlation coefficients with PC scores. For PCA, both methods are in fact equivalent. For the determinant criterion, the same subsets are obtained by the McCabe (i) method and GA. In the method of McCabe (i), all possible ($16 \times 17 \times 18 \times 19 = 5814$) subsets are evaluated to decide the most representative one, while GA successfully finds the best solution with a smaller number of evaluations.

The results based on the consensus criterion may be biased since the criterion is optimised by the proposed method. The PC score plot and correlation coefficient are used to compare different methods. The score plot of PC1–PC2 obtained from all 19 variables (Fig. 2a) shows four clusters. The four clusters are distributed as a trapezoid, the two upper clusters being closer than others. The best result of the two Jolliffe methods, i.e. the Jolliffe (i) method (Fig. 2b), shows only one cluster in the lower left clearly separated from the others. The separations between the other three clusters become vague compared with Fig. 2a. Fig. 2c is the best result among the four McCabe methods. Three clusters are somewhat merged, and only one cluster in the lower right is well represented by the subset. The results indicate that none of these subsets captures well the structure of the complete data. In Fig. 2d, four clusters are retained by Krzanowski's method, but three outliers appear, namely objects 34, 8, and 9, which did not exist in Fig. 2a. In

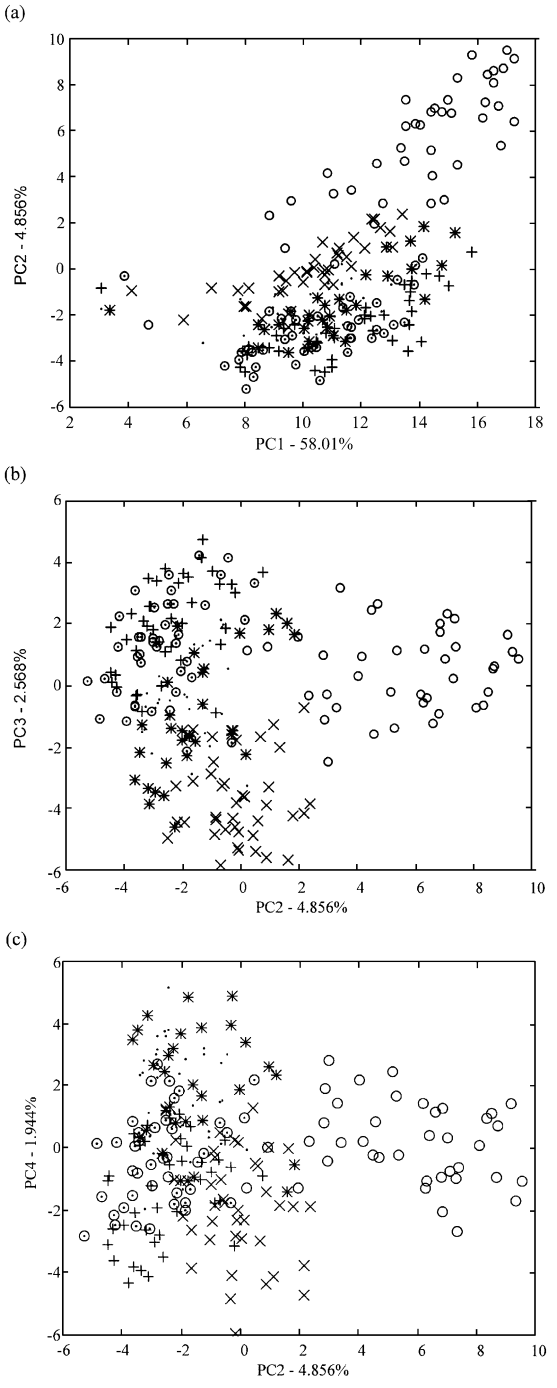

Fig. 3. Score plot using all 199 variables for the GC data; (a) PC1–PC2, (b) PC2–PC3, and (c) PC2–PC4; '*' denotes fructose, '.' glucose, '+' lactose, '⊙' maltose, 'o' rhamnose, and '×' xylose.

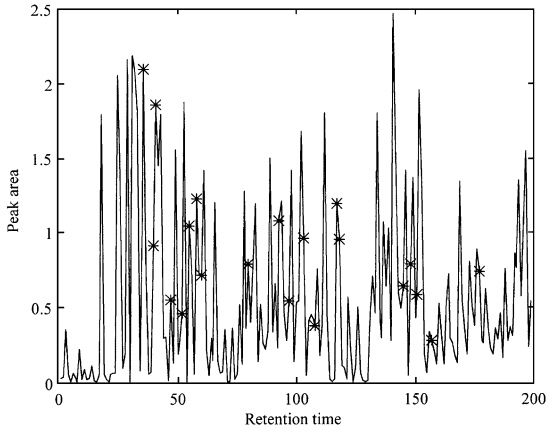

Fig. 4. The 20 variables selected by the proposed method.

Fig. 2e, obtained with the proposed method, the four clusters are well represented and the distribution of the samples is quite similar to that in Fig. 2a. As the four clusters are mainly separated in PC1, the correlation coefficients between the scores on PC1 in Fig. 1a and the scores on PC1 in the other figures (Fig. 1b–e) could also give a more quantitative measures of the similarities between them. The correlation coefficients (0.96, 0.90, 0.98, and 0.99, respectively, for Fig. 1b–e) lead to the same conclusion that Fig. 1e retains the most similar cluster information to that in Fig. 1a. These results clearly indicate that for this data set, the proposed method captures more structural information than the other ones.

4.2. GC data from food chemistry

For this data set, five PC factors are considered significant by the reduced eigenvalue technique [16]. The six methods are applied to select 5, 10, and 20 variables. The resulting consensus values are listed in Table 2. Two approaches are used for the loading-based method. One is to select one variable on each of the first few PCs (e.g. 5, 10, and 20 PCs), and the other is to select two or four variables on each of the first five PCs. The consensus value for the subset selected by the latter approach is higher than that of the former approach. When five variables are selected, the proposed method contains 71.5% of the structural information of the complete data, which is higher than

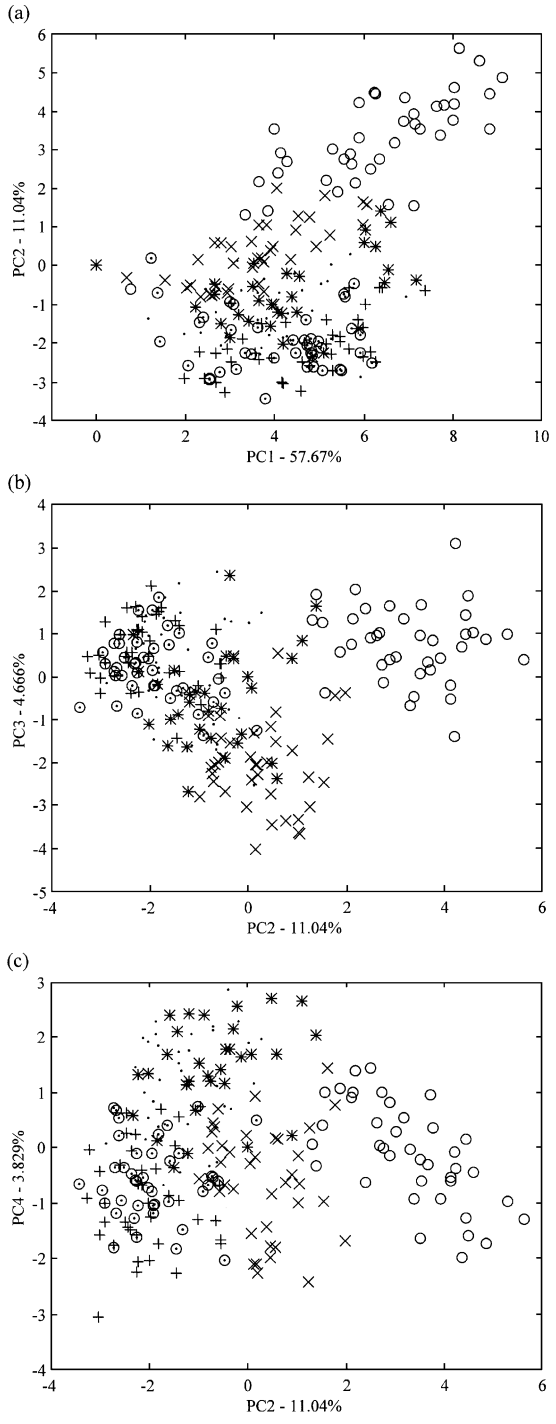

Fig. 5. Score plot using the 20 variables selected by the proposed method for the GC data; (a) PC1–PC2, (b) PC2–PC3, and (c) PC2–PC4; '*' denotes fructose, '.' glucose, '+' lactose, '⊙' maltose, '○' rhamnose, and '×' xylose.

all other methods. When a subset contains 10 variables, the proposed method also gives the highest consensus value (81.2%). The consensus value is 89.7% when 20 variables are selected by the proposed method. It is interesting to note that the 20 variables selected based on the first five PCs also preserve a high-consensus value (86.5%) compared with other methods.

Fig. 3a–c are the score plots obtained from the complete data. The 20 variables selected by the proposed method are marked in the average profile of the data (Fig. 4), and most of them are chromatographic peaks. The score plots (Fig. 5a–c) obtained with the best subset are compared with Fig. 3a–c. In the score plot of PC1–PC2 (Fig. 3a) from the complete data, most rhamnose-containing objects reside on the upper right of the plane, and are overlapped with the xylose group. There is a small cluster on the lower left with six objects, all containing the same amino acid (cysteine) and different sugars. The six objects are still situated on the lower left along one diagonal line from [0, 0] to [2, −3.2] in the score plot of PC1–PC2 (Fig. 5a). The rhamnose-containing objects are also separated from the others in the direction of PC2.

In the score plot of PC2 against PC3 (Fig. 3b) from the complete data, one can see that objects containing rhamnose and xylose form two clusters. All other objects together form one big cluster. Overlapping occurs between the xylose cluster and the cluster with other objects. A similar separation is observed in the PC2–PC3 score plot (Fig. 5b) from the subset, between objects containing xylose, objects containing rhamnose, and all others.

In both score plots of PC2–PC4 (Figs. 3c and 5c), the fructose and glucose containing objects are discriminated from most maltose and lactose containing objects, and the separations in both plots are quite similar. All these results show that the best subset selected by the new method yields similar group structures as those presented by the whole data set.

5. Conclusion

A new method is proposed to select features from high-dimensional data in data mining. The Procrustes consensus criterion makes subset evaluation and interpretation more intuitive. A genetic algorithm is applied to search for the best subset in principal component analysis. The performance of the method was validated on the Alate data by comparison of other feature selection methods. The method was also applied to industrial GC data. The results show that the proposed method leads to a better subset of variables than Krzanowski's original method. It also outperformed the other studied feature selection methods by keeping more structural information of the whole data.

References

[1] A.W. Czarnik, Combinatorial chemistry, Anal. Chem. 70 (1998) 378A–386A.
[2] Q. Guo, W. Wu, F. Questier, D.L. Massart, C. Boucon, S. de Jong, Sequential projection pursuit using genetic algorithms for data mining of analytical data, Anal. Chem. 72 (2000) 2846–2855.
[3] I.T. Jolliffe, Discarding variables in a principal component analysis: I, Artificial data, Appl. Stat. 21 (1972) 160–173.
[4] I.T. Jolliffe, Discarding variables in a principal component analysis: II, Real data, Appl. Stat. 22 (1973) 21–31.
[5] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.
[6] G.P. McCabe, Principal variables, Technometrics 26 (1984) 137–144.
[7] E.R. Malinowski, Obtaining the key set of optical vectors by factor analysis and subsequent isolation of component spectra, Anal. Chim. Acta 134 (1982) 129–137.
[8] S. Rännar, S. Wold, E. Russell, Selection of spanning variables in PCA, in: S. Rännar, Many Variables in Multivariate Projection Methods, PhD Thesis, Department of Organic Chemistry, Umeå University, Sweden, 1996.
[9] W.J. Krzanowski, Selection of variables to preserve multivariate data structure, using principal components, Appl. Stat. 36 (1987) 22–33.
[10] S. Van Huffel, J. Vandewalle, Subset selection using total squares approach in collinearity problems with errors in variables, Linear Algebra and Its Applications, Elsevier Science Publishing Co., Inc., New York, 1987, pp. 695–714, 88/89.
[11] R. Leardi, R. Boggia, M. Terrile, Genetic algorithms as a strategy for feature selection, J. Chemom. 6 (1992) 267–281.
[12] Q. Guo, W. Wu, D.L. Massart, C. Boucon, S. de Jong, Feature selection in sequential projection pursuit, Analytica Chemica Acta 446 (2001) 85–96.
[13] F.V. Warren, B.A. Bidlingmeyer, Selection of representative wavelength sets for monitoring in liquid chromatography with multichannel ultraviolet-visible detection, Anal. Chem. 59 (1987) 1890–1896.
[14] F.V. Warren, B.A. Bidlingmeyer, Selection of wavelengths for absorbance ratio monitoring in liquid chromatography, Anal. Chem. 59 (1987) 1897–1907.

[15] S. Wold, Cross-validation estimation of the number of components in factor and principal components models, Technometrics 20 (1978) 397–405.
[16] E.R. Malinowski, Factor Analysis in Chemistry, 2nd edn., Wiley, New York, 1991.
[17] G. Dijksterhuis, W.J. Heiser, The role of permutation tests in exploratory multivariate data analysis, Food Qual. Preference 6 (1995) 263–270.
[18] J.C. Gower, Generalised procrustes analysis, Psychometrika 40 (1975) 33–51.
[19] J.R. Bunch, C.P. Nielsen, Updating the singular value decomposition, Numer. Math. 31 (1978) 111–129.
[20] J.R. Bunch, C.P. Nielsen, D.C. Sorensen, Rank one modification of the symmetric eigenproblem, Numer. Math. 31 (1978) 31–48.
[21] G. Dijksterhuis, P. Punter, Interpreting generalised procrustes analysis 'analysis of variance' tables, Food Qual. Preference 2 (1990) 255–265.
[22] G.B. Dijksterhuis, J.C. Gower, The interpretation of generalised procrustes analysis and allied methods, Food Qual. Preference 3 (1991/2) 67–87.
[23] D.E. Goldberg, Genetic Algorithms in Search, Optimisation and Machine Learning, Addison-Wesley Publishing, Reading, MA, 1989.
[24] C.B. Lucasius, G. Kateman, Understanding and using genetic algorithms: Part 1, Concepts, properties and context, Chemom. Intell. Lab. Syst. 19 (1993) 1–33.
[25] J.R.N. Jeffers, Two case studies in the application of principal component analysis, Appl. Stat. 16 (1967) 225–236.