Available online at www.sciencedirect.com

SCIENCE DIRECT®

Chemometrics and Intelligent Laboratory Systems 76 (2005) 45–54

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemolab

The use of CART and multivariate regression trees for supervised and unsupervised feature selection

F. Questier[a], R. Put[a], D. Coomans[a,b], B. Walczak[a,1], Y. Vander Heyden[a,*]

[a]*ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Department of FABI, Laarbeeklaan 103, 1090 Jette, Belgium*
[b]*Statistics and Intelligent Data Analysis Group, School of Mathematical and Physical Sciences, James Cook University, Townsville Q4814, Australia*

Received 16 February 2004; accepted 13 September 2004
Available online 11 November 2004

Abstract

Feature selection is a valuable technique in data analysis for information-preserving data reduction. This paper describes Classification and Regression Trees (CART) and Multivariate Regression Trees (MRT)-based approaches for both supervised and unsupervised feature selection. The well-known CART method allows to perform supervised feature selection by modeling one response variable (**y**) by some explanatory variables (**x**). The recently proposed CART extension, MRT can handle more than one response variable (**y**). This allows to perform a supervised feature selection in the presence of more than one response variable. For unsupervised feature selection, where no response variables are available, we propose Auto-Associative Multivariate Regression Trees (AAMRT) where the original variables (**x**) are not only used as explanatory variables (**x**), but also as response variables (**y**=**x**). Since (AA)MRT is grouping the objects into groups with similar response values by using explanatory variables, this means that the variables are found which are most responsible for the cluster structure in the data. We will demonstrate how these approaches can improve (the detection of) the cluster structure in data and how they can be used for knowledge discovery.

Keywords: Supervised; Unsupervised; Feature selection; CART; MRT; AAMRT; Clustering; Auto-associative; Multivariate regression trees

1. Introduction

A common problem of data analysis is that the large number of features obscures the patterns that are present in the data. Therefore, one often tries to reduce the number of features by applying a feature selection approach. While such methods are relatively well known in supervised data analysis, they are much less so in unsupervised applications. In supervised applications, such as classification (with categorical response variables) and calibration (with numerical response variables), the response variable(s) are modeled by the explanatory variables in the data set. Many feature selection methods exist for supervised classification (overviews can be found in Refs. [1,2]), but no standard approach is available for feature selection in unsupervised clustering. The measured features often are not all equally informative: some of them may be redundant or irrelevant for the problem. Often many candidate features are included since the relevant features are unknown a priori. Unfortunately, the presence of additional features does not always result in a better revelation of the hidden natural patterns in the data. On the contrary, the presence of irrelevant features can mask the underlying natural patterns in the data as was demonstrated by Milligan [3], who showed that by adding "masking variables" to strongly clustered data, the recovery of the underlying clusters deteriorated rapidly for a wide variety of clustering methods. Feature selection addresses this concern by the selection of a (minimally sized) subset of features that are

* Corresponding author. Tel.: +32 2 4774733; fax: +32 2 4774735.
E-mail address: yvanvdh@vub.ac.be (Y.V. Heyden).
[1] On leave from Silesian University, Katowice, Poland.

0169-7439/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2004.09.003

relevant for the problem. Knowledge about the most important features and the way they interact can help with the interpretability of the problem.

In unsupervised learning (clustering e.g.), there are no response variables available. This severely restricts the number of applicable feature selection methods in this field. We proposed approaches based on rough set theory [4] and on genetic algorithms [5] for feature selection for (hierarchical) clustering, with as criterion that the clustering structure found with all features would be preserved as much as possible after reduction of the number of features. These methods therefore cannot lead to improved, i.e. possibly different, clustering results. In this article, we propose a new unsupervised feature selection method based on Multivariate Regression Trees (MRT), recently proposed by De'ath [6]. MRT is an extension of the Classification and Regression Trees (CART) method to allow more than one response variable. We will demonstrate different types of applications on different data sets, schematically represented in Fig. 1. The typical supervised feature selection approach based on CART is represented in Fig. 1a. One response variable (**y**) is modeled by some explanatory variables (**x**). MRT can handle more than one response variable (Fig. 1b). We propose Auto-Associative Multivariate Regression Trees (AAMRT), where the original variables are not only used as explanatory variables, but also as response variables (Fig. 1c). Since (AA)MRT is grouping the objects into groups with similar response values by using explanatory variables, this means that the variables are found which are most responsible for the cluster structure in the data. We will demonstrate how this can improve the detection of the cluster structure in data and how this can be used for knowledge discovery.

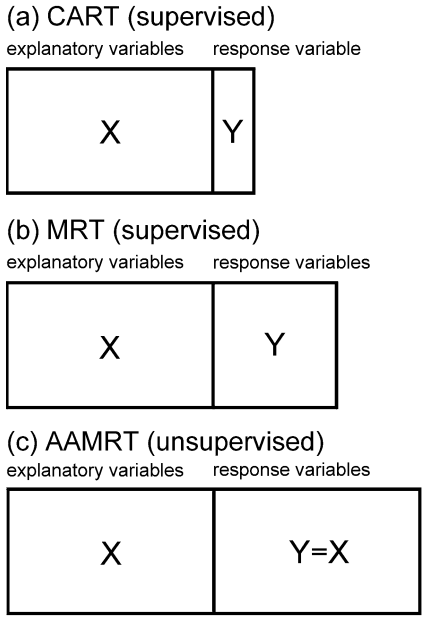

Fig. 1. Summary of CART, MRT and AAMRT.

2. Theory

2.1. Classification and Regression Trees (CART)

Classification and Regression Trees (CART), introduced by Breiman et al. [7], is a statistical technique that can select from a large number of explanatory variables (**x**) those that are most important in determining the response variable (**y**) to be explained. This is done by growing a tree structure, which partitions the data into mutually exclusive groups (nodes) each as pure or homogeneous as possible concerning their response variable. Such a tree starts with a root node containing all the objects, which are divided into nodes by recursive binary splitting. Each split is defined by a simple rule based on a single explanatory variable.

The CART steps can be summarized as follows:

(1) Assign all objects to root node.
(2) Split each explanatory variable at all its possible split points (that is in between all the values observed for that variable in the considered node).
(3) For each split point, split the parent node into two child nodes by separating the objects with values lower and higher than the split point for the considered explanatory variable.
(4) Select the variable and split point with the highest reduction of impurity.
(5) Perform the split of the parent node into the two child nodes according to the selected split point.
(6) Repeat steps 2–5, using each node as a new parent node, until the tree has maximum size.
(7) Prune the tree back using cross-validation to select the optimal sized tree.

For regression trees (with a numerical response variable), the impurity calculated at step 4 can be defined as the total sum of squares of the response values around the mean of each node [7]. For a node with n objects, the impurity is then defined as:

$$\text{impurity} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

For classification trees (with a categorical response value), the impurity is defined with, e.g. the Gini index of diversity [7]. The Gini index of a node with n objects and c possible classes is defined as:

$$\text{Gini} = 1 - \sum_{j=1}^{c} \left(\frac{n_j}{n}\right)^2$$

where n_j is the number of objects from class j present in the node.

Many proposals were made to get a right-sized tree by defining stopping rules that can be used during the tree growing. Breiman et al. [7] pointed out that "looking for the

right stopping rule was the wrong way of looking at the problem". Different parts of the tree might need different depths. Stopping too early might fail to uncover interactions between explanatory variables. Therefore, a tree is generally first grown to its maximal size, that is until all terminal nodes are either small (one object or not more than a predefined number of objects) or pure (all objects in the node have the same response value or class). This tree of maximum size is usually overfitted. It fits the noise and every idiosyncrasy in the learning data set, which are unlikely to be present with the same pattern in future data sets. In a next step, the tree is gradually shrunk by pruning away branches that lead to the smallest decrease in accuracy compared to pruning other branches. For each subtree T, a cost-complexity measure $R_\alpha(T)$ is defined as:

$$R_\alpha(T) = R(T) + \alpha|T|$$

where $|T|$ is the complexity of the subtree T (number of terminal nodes), α is called the complexity parameter, and $R(T)$ is the resubstitution error (overall misclassification rate for classification trees or the total residual sum of squares for regression trees). For each value of α, there is a unique smallest tree that minimizes the cost complexity measure $R_\alpha(T)$. Gradually increasing the complexity parameter α starting from 0, results in a nested sequence of trees decreasing in size. Since each of these trees is the best of its size, choosing the best tree can then be redefined as choosing the best size. This optimal tree size is determined by *cross validation*. The data set is randomly divided into N (usually 10) subsets. One of the subsets is used as independent test set while the other $N-1$ subsets are combined and used as learning data set. The tree growing and pruning procedure is repeated N times, each time with a different subset as test set. For each size of the tree, the prediction error is calculated, averaged over all subsets. This prediction error is calculated as the misclassification error for classification trees or as the sum of squared differences between the observations and predictions for regression trees. The prediction error obtained for each subtree on the cross-validation subset is matched with the subtrees of the complete data set using the α values. The optimal sized tree is the one with the lowest cost-complexity measure, but Breiman et al. [7] proposed to select the smallest tree, such that its cost complexity measure is within one standard error of the cost complexity for the tree with the minimum cost complexity. This rule, known as the 1 S.E. rule, usually allows selection of a much smaller tree whose accuracy is still comparable to the optimal one. This is demonstrated in Fig. 2. The most right solid circle represents the tree with the lowest cost complexity measure and has 40 nodes. The other solid circle represents the smallest tree (only 26 nodes) with a cost complexity measure within one standard error of the one of the first tree.

At each node, one can determine a primary variable, alternative variables and surrogate variables. The splits used in these variables are called primary split, alternative splits and surrogate splits, respectively. The best variable selected at each node of the tree is called (first) primary variable. The competing best variables are called alternative variables. Note that in some software packages these alternative

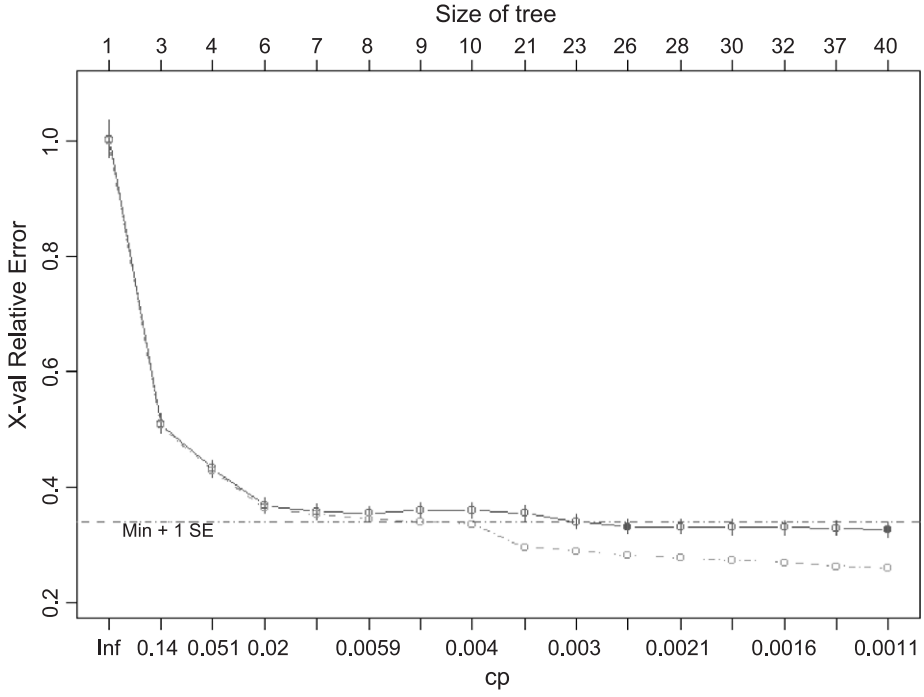

Fig. 2. Error graph for flavour data set (full line is cross-validation relative error, dashed line is resubstitution error).

variables are also called primary variables. Surrogate variables are defined as the variables that most accurately predict the action of the primary variable. This means that surrogate variables will lead to a similar partitioning into two nodes as the primary variable. Surrogate variables are used when the data contains missing values. The best surrogate will often, but not always, give the best alternative split.

Since the primary variables can be considered to have the best predicting power, these can be used as the best feature subset in a feature selection approach. Note that using the n first listed variables of the so-called "variable importance ranking" (often available in software packages) does not lead to the best subset of n features, since they contain correlated variables too. Inspection of the alternative and surrogate variables can lead to a better understanding of the data. When a selected primary variable is not satisfying, e.g. because of high cost, it might be replaceable with a good alternative or surrogate variable. Sometimes one could try to reduce further the number of features by replacing a primary variable by its strong competing alternative variable, if this alternative variable is already used elsewhere in the tree as splitting variable.

Results can be presented in a tree such as in Fig. 3. Above each split, the splitting rule is written. Under each (terminal) node, the number of objects can be written. Typically, the length of the vertical lines represent the strength of each split, this is the proportion of the total sum of squares explained by the split for regression trees or the misclassification rate for classification trees.

2.2. Multivariate Regression Trees (MRT)

Classification and Regression Trees, as introduced by Breiman et al. [7], are univariate regression trees, which means they can handle only a single response variable. Multivariate Regression Trees (MRT), as introduced very recently by De'ath [6], are an extension of CART in order to handle several response variables. (Note that the few other articles discussing "multivariate regression trees" are most often about multivariate splits, which are splits formed not by a single explanatory variable but by a combination of several explanatory variables.) These MRT multivariate regression trees are constructed in the same way as in the classic CART, but the impurity is defined as the total sum of squares of the response values around the multivariate mean of the nodes. Geometrically, this is simply the squared Euclidean distance of objects around the node centroid.

For a node with n objects, each having p variables, the impurity is then defined as:

$$\text{impurity} = \sum_{i=1}^{n} \sum_{j=1}^{p} (y_{ij} - \bar{y}_j)^2$$

The other concepts and practices of CART, such as methods for determining tree splits and the optimal tree selection based on cross-validation carry over to MRT. Because the multivariate response introduces extra complexity, additional tools were developed for the interpretation of the MRT analysis. For instance, at each node of the

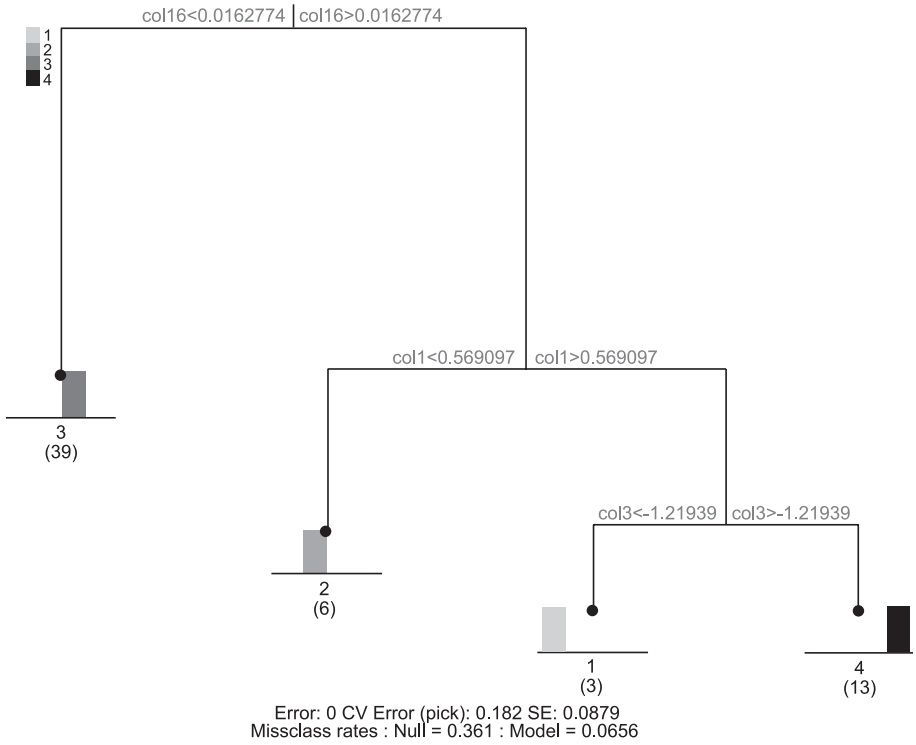

Fig. 3. CART Tree for viruses.

tree, a bar plot shows the distribution of the response variables in that particular node (see Fig. 4).

2.3. Auto-Associative Multivariate Regression Trees (AAMRT)

CART is often used for feature selection. However, only supervised feature selection is possible since a response variable is needed. For our main interest, clustering, which is a form of unsupervised learning, such a response variable is typically not available. Therefore, we propose an approach based on multivariate regression trees (MRT) [6]. With MRT, one can use more than one response variable, but this still leaves us with the question what to use as response variables in unsupervised learning. We propose to use the original variables, not only as explanatory variables, but also as response variables. Since (AA)MRT is grouping the objects into groups with similar response values by using explanatory variables for splitting rules, this means that the variables are found which are most responsible for the cluster structure in the data. All concepts and practices of CART and MRT, such as methods for determining tree splits and the optimal tree selection based on cross-validation carry over to AAMRT.

The Auto-Associative Multivariate Regression Trees can somehow be compared with auto-associative neural networks (also referred to as identity mapping) [8], where the input is also used as desired output. By going through a "bottleneck" layer of smaller dimension than input and output, the data is compressed. The fewer features of this bottleneck will represent the most significant patterns of the data. In contrast with Auto-Associative MRT, these bottleneck features are not selected from the original input features, but they are new features (combinations of the original features). Thus, auto-associative neural networks are feature *reduction* methods, but not feature *selection* methods, in which we are interested.

3. Software

The trees were constructed in the S-Plus 2000 Environment (Statistical Sciences) with the Trees++ add-on module, developed by De'ath and Coomans [6]. This S-Plus-based library can be downloaded from Ref. [9] and is based on (the univariate) RPART [10]. In contrast to the built-in S-Plus 2000 module for trees, this Trees++ add-on module allows to use cross validation and multivariate regression trees. Most other CART software implementations, such as CART® (Salford Systems, San Diego, CA) and the Tree module standard available in S-Plus 2000 are univariate.

Additional calculations (e.g. hierarchical clustering) were performed under MATLAB 5.3 (Mathworks).

4. Data sets

4.1. Synthetic data set

This data set was constructed to evaluate the ability of the methods to find the relevant features in the presence of

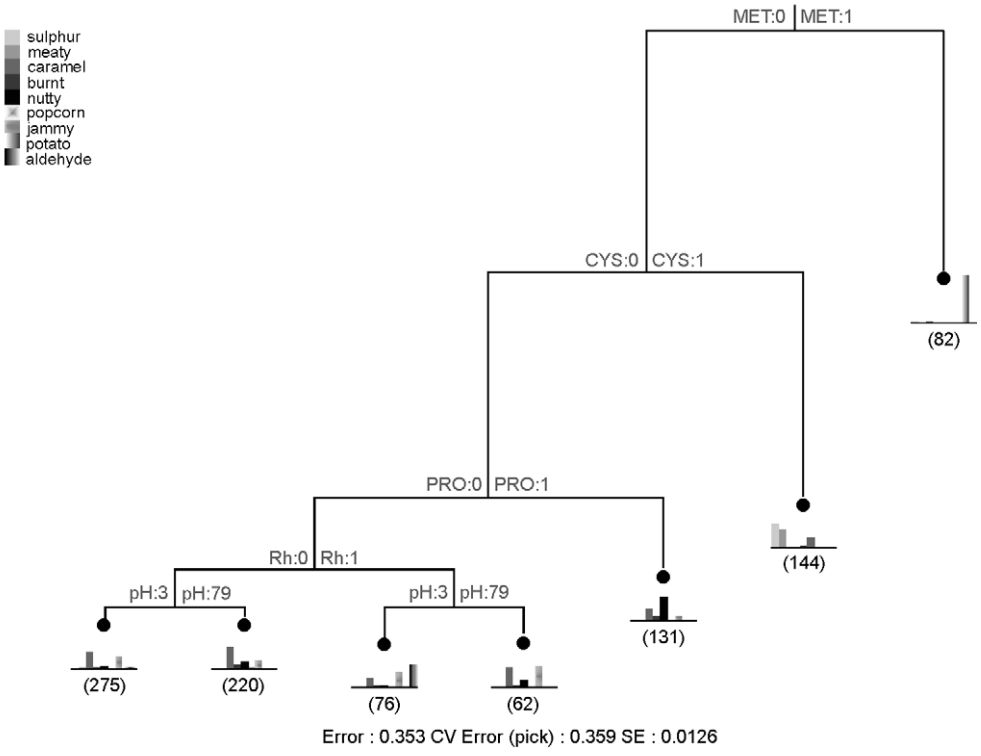

Fig. 4. MRT Tree for flavour data set.

redundant and irrelevant features. The first two features were constructed to contain four separated clusters. Features 3, 4 and 5, 6 were constructed to be perfectly correlated with the first two features, in order to introduce redundancy. Four irrelevant features with random data were added. The resulting data set has size 40×10.

4.2. Viruses data set

This data set is described in the book Pattern Recognition and Neural Networks by Ripley [11] and is available at his website [12]. The data set consists of 61 viruses (3 Hordeviruses, 6 Tobraviruses, 39 Tobamoviruses and 13 'furoviruses') with rod-shaped particles, affecting various crops (tobacco, tomato, cucumber and others). There are 18 measurements on each virus, the number of amino acid residues per molecule of coat protein.

4.3. Flavour data set

This data set consists of the sensorically evaluations of Maillard reaction product mixtures. The Maillard reaction plays an important role in the development of brown colour and flavours during thermal heating of food products. It involves very complex reactions between reducing sugars and amino acids. Even the simplest possible prototype system of one sugar reacting with one amino acid gives rise to hundreds of reaction products since intermediate products may react with their precursors or with one another. Each sample was obtained as Maillard reaction products of mixtures of 1 or 2 sugars (maltose, lactose, fructose, glucose, xylose or rhamnose) and 1 or 2 amino acids (alanine, cysteine, glutamate, glycine, lysine, asparagine, glutamine, arginine, threonine, methionine or proline) at 2 different pH conditions (3 and 7.9). The sugars and amino acids are absent or present, coded in separate binary variables. These samples were evaluated sensorically by a trained panel of three persons. Each sample was given a score from 0 (absent) to 4 (very strong) for nine smells (overall, sulphur, meaty, caramel, burnt, nutty, popcorn, jammy, potato, aldehyde).

The data set consists of 992 samples, 18 explanatory variables (sugars, amino acids and pH) and 10 response variables (smells). Two samples had missing response values.

4.4. Bacteria data set

For a taxonomic study focusing on poultry spoilage microorganisms (mainly *Pseudomonas* flora) [13], 60 bacteria strains (isolated from broiler skins in slaughter houses) and 36 reference bacteria were tested. For each bacterium, 126 phenotypic features were tested: the BIOLOG GN identification system, based on oxidative capacity of microorganisms; API 20NE using several biochemical and assimilation tests; and some other non-automated classical tests (morphological, physiological and biochemical). The tests are coded in a discrete way: 1 for positive results, 0 for negative results and 0.5 for doubtful results. Each reference bacterium was tested twice to check for and cope with the variability, which is typically large for such biological tests. In the present study, these double measurements on the reference bacteria were used as data set (72×126). More details about this data set can be found in our previous article [4], where another feature selection approach on this data set was described.

5. Results

5.1. CART example: viruses data set

CART is typically used in a supervised way, with class labels (or with a numerical response for regression trees). To demonstrate this typical approach, we applied CART with Gini index on the viruses data set. The four known virus classes were used as response variable. A decision tree (Fig. 3) with 0 misclassifications can be constructed using three variables: 16, 1 and 3. Variable 16 is used to split off class 3. Variable 1 is used to split off class 2. Variable 3 is used to further distinguish between class 1 and 4. The bar plot under each node represents the number of objects in that node that belong to each class.

While the real need of feature selection for this data set might be questionable, we consider it as a nice example where three features can be enough for a good classification tree.

5.2. CART and AAMRT examples: synthetic data set

Both with CART (supervised approach can be used since we know which objects belong to which of the four clusters) and AAMRT method, two features (1 and 2) are found for the optimal tree. These are indeed the two variables in which space the studied data has clustering tendency. The correlated and irrelevant variables are not selected in the main tree. The correlated variables are found as alternative variables and surrogates. The objects are correctly assigned to four groups by both the CART and the AAMRT trees with zero misclassifications.

Fig. 5a shows a hierarchical clustering of the data set with the complete feature set, while Fig. 5b shows a clustering based on only the two relevant features. It can be seen that cluster separability is seriously improved. With only two relevant features, it is much more obvious that the natural number of clusters in the data set is four.

To determine the optimal subset of features or the optimal number of features to select, one can use the (number of) features as suggested by cross-validation for the optimal tree.

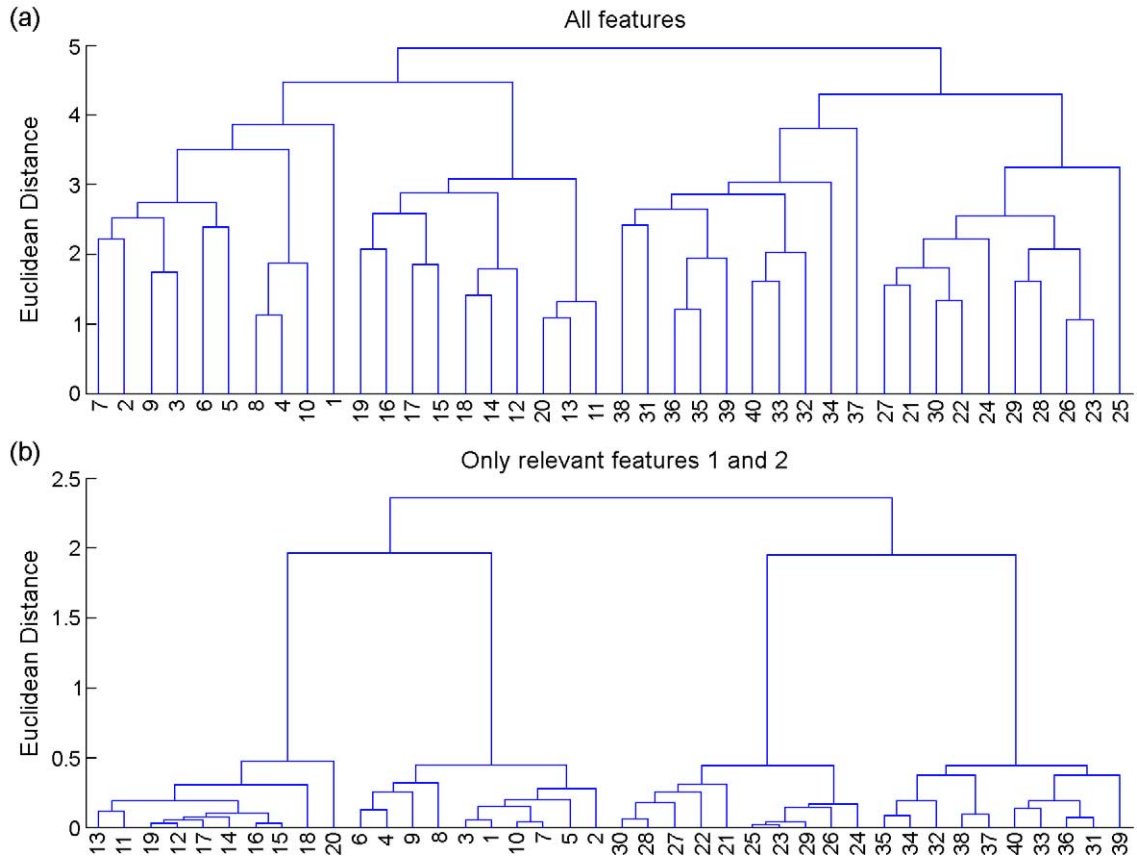

Fig. 5. UPGMA clustering on synthetic data set containing four clusters–based on (a) complete feature set, (b) feature subset.

5.3. Supervised MRT example: sensory data set

After a first exploratory analysis, we simplified the analysis by removing the overall (strength) smell, which we considered as less important and which seems—not surprisingly—strongly influenced by unwanted smells like, e.g. sulphur. Cross-validation suggests us a tree with 26 terminal nodes (see error graph in Fig. 2), but we start with investigating a smaller one, easier to investigate. The tree with seven terminal nodes (Fig. 4) has almost the same estimated error. Note that in MRT, the bar plot represents the average value for each of the response values.

The first split is determined by methione. The presence of methione results in a very strong potato flavour with almost no other flavours. This right branch with methione is strongly homogeneous and is not further splitted. The next split is determined by cysteine. The presence of cysteine results in a sulphur or meaty flavour. The next split is determined by proline, whose presence results in a nutty flavour. The next split is determined by rhamnose, whose presence would result in an aldehyde smell (pH 3) or a rather jammy and caramel flavour (pH 7.9). The bottom left split—based on pH in absence of rhamnose—is the least important one as can be seen from the length of its vertical lines. This split is harder to explain as differences in level of caramel and nutty flavour.

By growing the tree larger, one can uncover more complex interactions, e.g. how to distinguish between the meaty flavour and the unwanted sulphur flavour. It seems that for the production of a meaty smell without sulphur smell, one can use a combination of cysteine and/or proline, with any sugar except rhamnose, at pH 7.9.

The analysis was repeated on a subset of the data, containing only samples with maximum one sugar and maximum one amino acid, to see whether this simplification would lead to extra conclusions, but this was not the case.

5.4. Auto-Associative MRT example: bacteria data

With standard settings, the optimal tree as suggested by cross-validation and the 1 S.E. rule has 10 nodes and uses nine variables. The hierarchical clustering results based on this subset of nine variables still make sense (see Figs. 6 and 7). However, many objects became indistinguishable from each other. This is due to the limited number of possible values for each variable (0, 0.5 and 1). Our previous studies [4] on this data set showed that almost half of the variables are needed to retain high similarity with the original

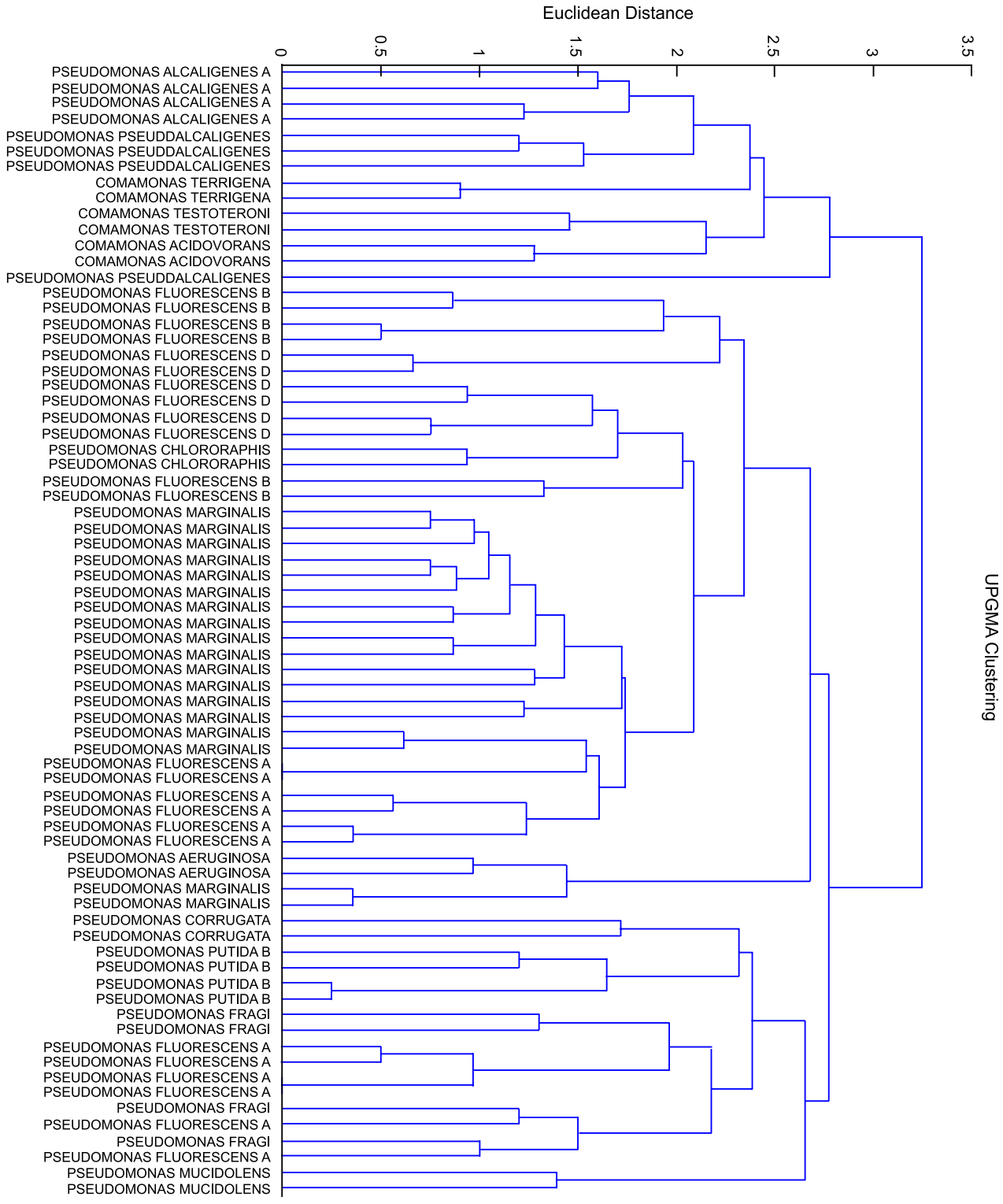

Fig. 6. Hierarchical clustering of bacteria data set, based on full feature set.

hierarchical clustering results up to a high detail level. The fact that some of the objects become indistinguishable from each other is not problematic since those objects seem to be most often from the same family and since, generally, the primary aim of clustering is to group objects. The contrast between these groups of similar objects and the rest is even better. While it is difficult to clearly pinpoint distinct groups in the hierarchical tree based on the complete feature set,

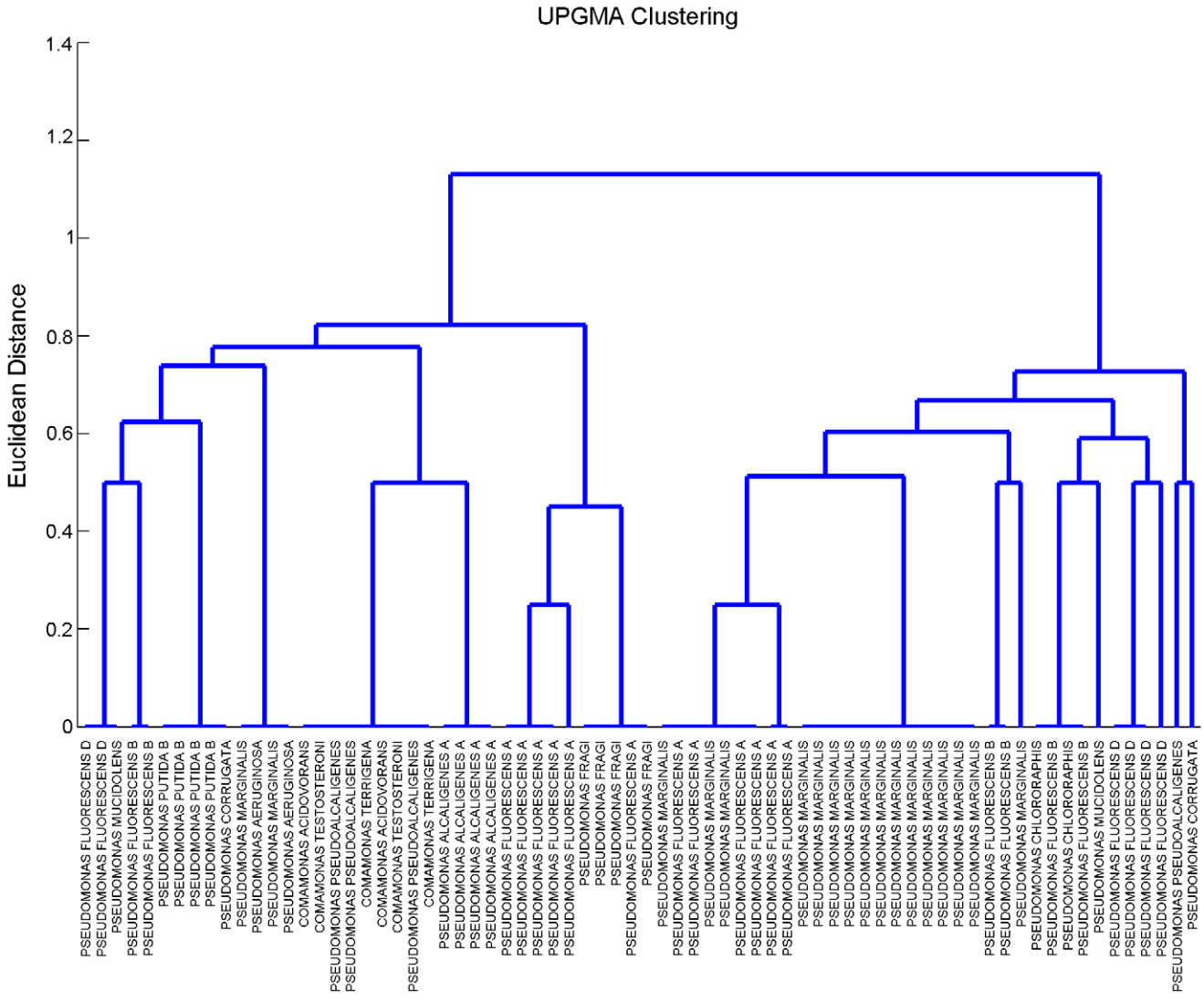

Fig. 7. Hierarchical clustering of bacteria data set, based on feature subset.

some tight groups are suggested in the tree based on the subset, such as *Pseudomonas marginalis* and *Pseudomonas alcaligenes* A.

6. Discussion and conclusion

We have outlined the fundamentals of the well-known CART and its recently proposed extension, Multivariate Regression Trees (MRT), and we have shown how they can be applied for supervised feature selection. We proposed a new approach for unsupervised feature selection revealing data clustering tendency. Since this method is based on MRT, we refer to this method as Auto-Associative Multivariate Regression Trees (AAMRT). In this Auto-Associative MRT, the original variables are not only used as explanatory variables, but also as response variables. Since (AA)MRT is grouping the objects into groups with similar response values by using explanatory variables, this means that the variables are found which are most responsible for the cluster structure in the data.

The well-known advantages of CART [6,14] carry over to its extensions MRT and AAMRT: (1) nonparametric method (no assumptions are made regarding the underlying distribution of the data), (2) invariance to monotonic transformations of the explanatory variables (only the rank order of each explanatory variable is important), (3) fast and simple method (relatively little input is required from the analyst), (4) easy (graphical) interpretation, (5) ability to handle missing data (with surrogates), (6) robust to noisy explanatory and response variables, (7) robust to outliers, since they will be either quickly separated in a separate class, or either do not influence the prediction, (8) model selection by cross-validation.

To know the optimal subset of features or the optimal number of features to select, one can use the (number of) features as suggested by cross-validation for the optimal tree.

The presented tree-based feature selection methods are stepwise feature selection methods. One feature is added at a time, i.e. one at each splitting step. Combinations of variables are not tested, but all features are considered at each step, also those features already selected in previous steps. Often the same feature but with another split point will be sufficient to further split the data. These approaches do not only find features with an 'overall goodness', but also features which are important in only a region of the data space. The combination of such features might lead to a superior model.

Demonstrations on synthetic and real data sets showed that the methods can effectively be used for feature selection. While reducing the number of features, the most important cluster structure is preserved. The methods can also possibly lead to improved detection of the cluster structure by removing the redundant and irrelevant features. One of the important reasons for which feature selection is used, is knowledge discovery and interpretability. Where most feature selection methods will tell not much more than (1) which features are most responsible for the structure in the data, the discussed tree methods also reveal (2) the split points, (3) the resulting groups, (4) the interactions of the features and (5) the alternative and surrogate variables. This extra information can lead to improved interpretability.

Acknowledgements

The authors wish to thank Claire Boucon and S. de Jong from Unilever for the flavour data set and I. Rollier for the bacteria data set.

References

[1] R. Kohavi, G. John, Wrappers for feature subset selection, Artif. Intell. 97 (1–2) (1997) 273–324.
[2] B. Walczak, D.L. Massart, Chapter calibration in wavelet domain, in: B. Walczak (Ed.), Wavelets in Chemistry, Elsevier, Amsterdam, 2000, pp. 323–349.
[3] G.W. Milligan, An examination of the effect of six types of error perturbation on fifteen clustering algorithms, Psychometrika 45 (1980) 325–342.
[4] F. Questier, I. Arnaut-Rollier, B. Walczak, D.L. Massart, Application of rough set theory to feature selection for unsupervised clustering, Chemometr. Intell. Lab. Syst. 63 (2002) 155–167.
[5] F. Questier, B. Walczak, D.L. Massart, C. Boucon, S. De Jong, Feature selection for hierarchical clustering, Anal. Chim. Acta 466 (2) (2002) 311–324.
[6] G. De'ath, Multivariate regression trees: a new technique for modeling species–environment relationships, Ecology 83 (2002) 1105–1117.
[7] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth, Monterey, CA, USA, 1984.
[8] M.A. Kramer, Nonlinear principal components analysis using auto-associative neural networks, AIChE J. 37 (1991) 233–243.
[9] G. De'ath, Multivariate regression trees: a new technique for modeling species–environment relationships, Ecol. Arch. E083-017-S1 (2002) http://www.esapubs.org/archive/ecol/E083/017/.
[10] T. Therneau, RPART software from statlib@lib.statlib.cmu.edu, 1998.
[11] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, 1996.
[12] B.D. Ripley, Pattern recognition and neural networks (datasets), http://www.stats.ox.ac.uk/~ripley/PRbook/.
[13] I. Arnaut-Rollier, L. Vauterin, P. De Vos, D.L. Massart, L.A. Devriese, L. De Zutter, J. Van Hoof, A numerical taxonomic study of the Pseudomonas flora isolated from poultry meat, J. Appl. Microbiol. 87 (1) (1999) 15–28.
[14] G. De'ath, K.E. Fabricius, Classification and regression trees: a powerful yet simple technique for ecological data analysis, Ecology 81 (2000) 3178–3192.