A Decision Model for Fuzzy Clustering Ensemble

Yanqiu Fu Yan Yang Yi Liu

School of Information Science & technology, Southwest Jiaotong University, Chengdu 610031, China

Abstract

Recent researches and experiments have showed that clustering ensemble approaches can enhance the robustness and stabilities of unsupervised learning greatly. Most of them focused on crisp clustering combination. However, in this paper, we offer a decision model based on fuzzy set theory for fuzzy clustering ensemble. Firstly, obtain the optimal partition called "expert" from H individual fuzzy partitions generated by fuzzy c-means algorithm. Then, use fuzzy voting scheme to generate the majority judger. Finally, the two matrixes are combined by Decision Model. Experimental results show the effectiveness of the proposed method comparing to the results based on crisp clustering.

Keywords: Fuzzy Clustering Ensemble, Fuzzy Valid Function, Similarity Measure, Comprehensive Decision Model

1. Introduction

Clustering algorithms provide an important mean to explore structure within the unlabelled data by organizing it into groups or clusters. Many clustering algorithms have existed, but no single algorithm can adequately handle all sorts of cluster shapes and structures [1]. The exploratory nature of clustering tasks demand efficient methods which would benefit from the strengths of many individual clustering algorithms. Inspired by the success of combining of multiple classifiers in improving the quality of data classifications, the same idea applied in integration of multiple clustering, which is considered to be an example to further broaden and stimulate new progress in the area [2], the basic process of clustering combination is illustrated in Figure 1.

In a recently review, there are many approaches proposed to cope with these problems [1]-[6], and most of researches interest in two main fields: generate different multiple clusterings and design the different consensus function [3]. Be noted of the related literatures above, few approaches focused on the combination of fuzzy clustering. Here below, we show some researches on this filed.

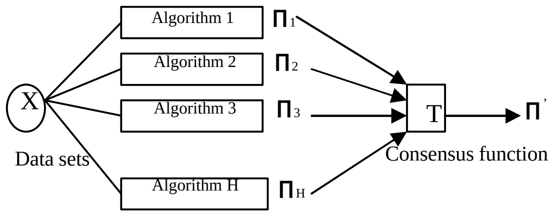


Fig.1: The process of clustering combination.

Finding a fuzzy consensus clustering or generate fuzzy co-association matrix are both feasible way in fuzzy clustering ensemble [7]. Fuzzy co-association matrix based on three fuzzy similarity between data points is generated in Ref[7], Ref[8] integrates fuzzy c-means (FCM) algorithm and fuzzy k-nearest neighbors (FKNN) algorithm through utilizing selected good results produced by FCM to teach the FKNN algorithm. Other related works are interesting in the application of fuzzy clustering ensembles [9], [10]. In this paper, we try to present a novel way for fuzzy consensus clustering, that is called *Decision Model*. In section 2, we introduce fuzzy valid functions, fuzzy voting scheme, similarity measure and the complete decision process for combining fuzzy clusterings. Section 3 is experimental analysis. Finally in section 4, we give the conclusion and discuss the future work.

2. A Dicision Model

In the field of fuzzy clustering ensemble, we can't use the general clustering ensemble method directly because of the vagueness an object belongs to any cluster. Thus, we should find an appropriate way to suit for fuzzy clustering combination. Here, we present a novel *decision model* for fuzzy clustering ensemble. To express the core idea, we take competition in our life for example. Sometimes as we know both experts and majority judgers' opinions should be taken into consideration for selecting the best ones. In this paper we extend the same concept to our fuzzy clustering ensemble. First, we choose an optimal partition as *expert* from *H* partitions generated by FCM algorithm, then produce *majority judger* by using fuzzy majority voting scheme. Finally, the clustering results are given through our *Decision Model*. The whole process is illustrated in Table 1. The detailed algorithm will be posed in **2.1, 2.2** and **2.3** sections:

2.1. FCM algorithm and Option of fuzzy validity functions

Suppose $X=\{X_1,X_2\cdots,X_n\}$ is a set of N data samples (objects) in d-dimensional space, where $X_i=\{x_{i1},x_{i2}\cdots,x_{id}\}$ represents the ith sample for i=1,2,$\cdots$,n. We use FCM algorithm to generate H fuzzy partitions matrix $\Pi=\{p_1,\cdots,p_H\}$ by randomly selecting c cluster centers and initial cluster prototype model. For the ith partition p_i, there is a membership degree m_{ij} ($\in$ [0,1])indicating with what degree the sample x_i belongs to the cluster center. These H partitions have some diversity from each other since the different settings of the initial conditions and parameters.

Input: *n d*-dimensional samples; *H*-number of partitions;
Output: Decision results of clustering combination
Initialization: Set c>=2, *m*>1, e>0;
Steps:
1. Produce *H* data partitions:
   For *j* = 1 to *H* do
   1.1 Randomly choose *c* cluster centers and initial cluster prototype model;
   1.2 Run the *FCM* Algorithm and produce *H* data partitions;
2. Obtain the optimal partition U_{ij}* called “*expert*” from the *H* partitions:
3. Generate *majority judgerFV*$_{ij}$:
   For each fuzzy partition matrix, count the times one sample belongs to the same cluster using *cut-set* theory.
4. Decision Model:
   Design a fuzzy comprehensive decision model:
   M_{ij} = U_{ij}* ? FV_{ij}:
   If the *expert* and the *majority judger* have consistent results, join the samples in the decided cluster, else the samples will be assigned to their appropriate cluster using similarity measure.

Table 1: A Decision Model for Fuzzy Clustering Ensembles.

It is expected to choose an optimal clustering which most closes to the natural structure of the given data set from these different *H*-partitions. To cope with this problem, we introduce the definition of cluster-validity analysis. The process of choosing a partition fits with the unknown structure of the input data set most using some kind of cluster validity functions is called cluster-validity analysis [11].

As for cluster validity functions which are often used to evaluate the performance of clustering in different indexes and even two different clustering methods, a lot of them were proposed during the last 10 years. Among the functions, there are two important types for FCM. One is based on the fuzzy partition of sample set, whose representative functions are partition coefficient (F_{pc}) and partition entropy (F_{pe}), the other is on the geometric structure of sample set, the representative functions for this type are N.Zahid-M.Limouri function (F_{FS}) and the Xie–Beni function (F_{xb}). A brief summary of these 4 functions are illustrated in Table 2.

Validity function	Functional description	Optimal partition
Partition coefficient	$F_{pc}(U;x)=\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2$	Max(F_{pc})
Partition entropy	$F_{pe}(U)=-\frac{1}{n}\left\{\sum_{i=1}^{c}\sum_{j=1}^{n}[\mu_{ij}\log\mu_{ij}]\right\}$	Min(F_{pe})
N.Zahid-M.Limouri function	$F_{FS}(U,V;x)=\sum_{i=1}^{c}\sum_{k=1}^{n}[(\mu_{ik})^m\lVert x_k-v_i\rVert^2]-\sum_{i=1}^{c}\sum_{k=1}^{n}[(\mu_{ik})^m\lVert v_i-\bar{v}\rVert^2]$	Min (F_{FS})
Xie-Beni function	$F_{xb}(U,V;X)=\frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2\lVert x_j-v_i\rVert^2}{n(\min_{i\neq k}\lVert v_i-v_k\rVert^2)}$	Min(F_{xb})

Table 2: A brief summary of four selected validity functions.

Based on the some experiences of other literatures, we select N.Zahid-M.Limouri function (F_{FS}) to be our validity function. [10,12] In fact, F_{FS} takes into account simultaneously the properties of the fuzzy membership degrees and the structure of the data itself, it is defined as:

$$F_{FS}(U,V;x)=\sum_{i=1}^{c}\sum_{k=1}^{n}[(m_{ik})^m\|x_k-V_i\|^2]$$
$$-\sum_{i=1}^{c}\sum_{k=1}^{n}[(m_{ik})^m\|V_i-\bar{V}\|^2]$$
$$=J_m(U,V;X)-K(U,V;X) \quad (1)$$

where $\bar{V}$ is the mean of the whole data set, J_m is a compactness measure and K_m is a separation measure between clusters centers and the mean $\bar{V}$. The matrix U_{ij}^* which has the minimum value of F_{FS} is recognized to be the optimal partition and called *expert* in this paper.

Since the *expert* has been selected from all the partitions in this section, one important judger is generated. However, considering the information only from the *expert* is not fully ideal and unilateral, fuzzy majority voting scheme of clustering individuals is needed and will be proposed in the next section.

2.2. Fuzzy majority voting scheme

The idea behind fuzzy majority voting is the judgment of a group is superior to those of individuals. This conception has been successfully explored in combining hard clusters to account the co-association values, However, in fuzzy clustering, we can't account values like in the general cluster ensemble method by using statistic of how often each pair of samples appears in the same cluster because we just obtain the different degrees of membership that every sample belongs to its own cluster. As a solution of this problem, the conception of *cut-set level* based on fuzzy set theory is introduced in this paper. The process is described as following:

Let *F(X)* be a fuzzy set, if $A \in F(X)$ and $\forall a \in [0,1]$, the *cut-set level* is a distinct set defined by $A_a = \{x | x \in X, m_A(x) \geq a\}$. For any partitions

$$p_t = \begin{bmatrix} m_{11}^t & m_{12}^t & \cdots & m_{1c}^t \\ m_{21}^t & m_{22}^t & \cdots & m_{2c}^t \\ \cdots & \cdots & m_{ij}^t & \cdots \\ m_{n1}^t & m_{n2}^t & \cdots & m_{nc}^t \end{bmatrix}, \; t=1,2,\cdots,H,$$

compute each row's *cut-set level*, that is $A_{ak} = \{x | x \in p_t, m_{ij}(x) \geq a\}$ (k=1,2,⋯,c), a can be fixed by 0.5. According to the *characteristic function* which is defined as $c_A(x_i) = \begin{cases} 1, x_i \in A \\ 0, x_i \notin A \end{cases}$, the probability of each sample belongs to a certain cluster can be counted by the equation of $fv_{ij} = \frac{\sum_{t=1}^{H} c_{At}^{j}(x_i)}{H}$, i=1,2,⋯,n; j=1,2,⋯,c, (2) which results in the fuzzy voting matrix FV_{ij}, where Each (i, j) is therefore a vote towards their gathering in a cluster. Now, another measure called *majority judger* is also produced.

2.3. Decision model

To take into account simultaneously the opinion of the *expert* and the attitude of *majority judger*, a *decision model* based on fuzzy set theory has been proposed in this section. A brief specification of the deciding process is firstly given as follows:

As the data set $X = \{X_1, X_2 \cdots, X_n\}$ given before, for each sample x_i (i=1,2,⋯,n), its cluster is fixed by the following description that is for each row, obtain the max value of *expert* $m_{ik} = \max_{1 \leq j \leq n}\{m_{ij}\}$ and the *majority judger* $fv_{it} = \max_{1 \leq j \leq n}\{fv_{ij}\}$, if $k=t$, join the x_i to the kth (or tth) cluster, else the cluster label is uncertain, which seems like a contradiction exists between two *judgers*. In this case, some more discussion should be taken as the next step. Noting that the *expert* is constant since it's one of the data partitions while the *majority judger* is variable because it depends on the performance of partitions, the *expert* considered to be the authority when the result of the *majority judger* is not ideal, for instance, every cell in *majority judger* matrix is equal, which suggests the voting is fail or the diversity of *majority judger* value for some samples is low, which can be represented by $\max_{1 \leq j \leq n}\{fv_{ij}\} - \min_{1 \leq j \leq n}\{fv_{ij}\} < 0.1$. Some other examples should be further studied in the future work. As for the remaining unlabeled samples, we can use the similarity measure to deal with this problem. Now, there are many ways to define the similarity measure, such as the related coefficient and Euclidean distance, etc. Motivated by simplicity and easy-manipulation, we choose the related coefficient measure because it's most frequently used [14]. The similarity measure r_{ij} is defined as follows:

$$r_{ij} = \frac{\sum_{k=1}^{m} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{m} (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^{m} (x_{jk} - \bar{x}_j)^2}},$$

where

$$\bar{x}_i = \frac{1}{m}\sum_{k=1}^{m} x_{ik}, \quad \bar{x}_j = \frac{1}{m}\sum_{k=1}^{m} x_{jk} \tag{3}$$

Observing the equation, it is easy to see the similarity measure well express the association between each pair of samples, with the idea that a higher value indicates greater similarity. Since similarity measure r_{ij} is called for each x_{ij} can uniformly distribute in [0, 1]. However, most real data sets may not meet the requirement of uniform distribution in [0, 1]. To meet the need of fuzzy similarity measure, the following transform is necessary:

$$x'_{ik} = \frac{x_{ik} - \min_{1 \leq i \leq n}\{x_{ik}\}}{\max_{1 \leq i \leq n}\{x_{ik}\} - \min_{1 \leq i \leq n}\{x_{ik}\}} \tag{4}$$

where k=1, 2, ⋯,m, i=1,2,⋯,n, each x_{ik} is a cell in the given data set $X = \{X_1, X_2 \cdots, X_n\}$.

We assume that neighboring samples within a "natural" cluster measured by similarity are very likely to be located in the same group, which implies that the samples within one cluster are more similar (r_{ij}? 1) and dissimilar samples are more separate (r_{ij}? 0), Therefore, we can join the unlabeled sample to the same cluster in which one labeled sample who has maximization of similarity degrees with it. Repeat this process until all the unlabeled samples are assigned.

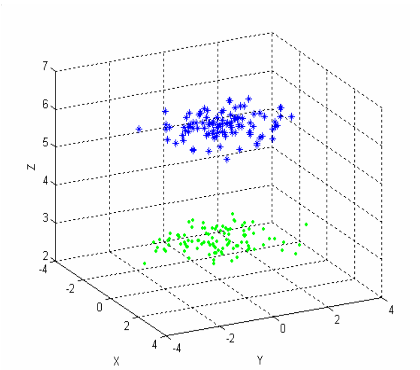
3. Experimental results

A reasonable approach to measure the performance of the proposed decision model is to compute the F-measure values on various datasets since our experiments were conducted with datasets where true natural clusters are known. The datasets include artificial, UCI and web types, some characteristics of the datasets are showed in Table 3.

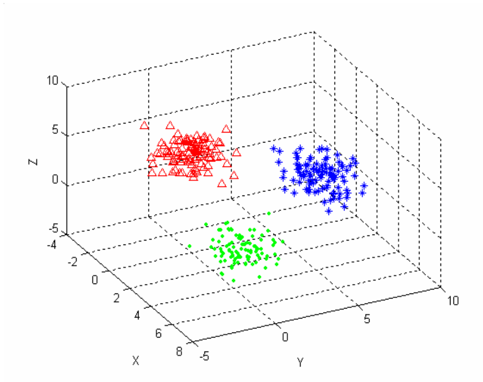
Datasets	Types	No.of features	No.of points	No.of clusters
3D-3C	Artificial	3	300	3
3D-2C	Artificial	2	200	2
Iris	UCI	4	150	3
Zoo	UCI	16	101	7
Web1	WEB	6	300	6
Web2	WEB	8	500	8

Table 3: Characteristics of the datasets.

The two artificial datasets of 3-dimensional points are artificially generated from Gaussian distributions with different means and co-variance matrices, and the sizes of them are as follows (100, 100, 100) and (100, 100). From Figure 2, we can easily see the distributions of the two datasets have satisfactory feature that is the samples within one cluster are similar and different centers are separate. Iris and Zoo datasets are available from the UCI machine learning repository at [13]. The Iris dataset consists of three classes (Setosa, Versicolor and Virginica), with 50 instances per class, represented by 4 features. The zoo dataset consists of 101 instances represented by 16 numeric attributes. Finally, we tested with two preprocessed web datasets which obtained from news documents downloaded from Yahoo English Website by RSS pattern.



(i)3D-2C



(?) 3D-3C

Fig 2: Distributions of the two artificial datasets (3D-2C and 3D-3C).

For comparison fuzzy clustering and crisp clustering, we perform FCM algorithm and K-means algorithm based on Euclidean Distance. However, FCM algorithm is used to generate *H* partitions while K-means algorithm is used to produce the final results directly. For sake of unification of comparison and based on Ref [12], we set *H*=20 for all the datasets in each of which the generated *H* partitions algorithm are combined by running the Decision Model independently. Then compute F-measure values of all the given datasets with the two methods respectively, which are reported in Table 4 and the comparative results are illustrated in Figure 3.

	3D-3C	3D-2C	Iris	Zoo	Web1	Web2
Decision Model	0.995	0.713	0.871	0.749	0.691	0.714
K-means	0.993	0.676	0.890	0.735	0.431	0.509

Table 4: F-measure values of K-means algorithm and the Decision Model.

Figure 4 includes six pie charts of six datasets clustering results produced by the proposed method, which specifically show which samples belong to their given clusters.

From the results, it is easy to see that the results from decision model are generally better than crisp one in the same conditions. In the case of “Web1” and “Web2” datasets, the F-measure values improve by bigger margin than other datasets, which suggests that the proposed method in this paper is more suitable for WEB datasets. The reason can be considered that the proposed method include more information than the crisp one since it consist of two judgers.

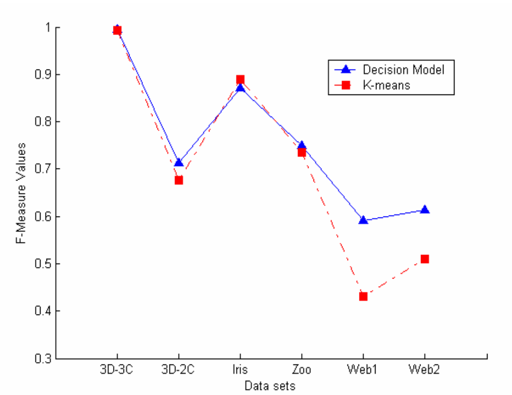
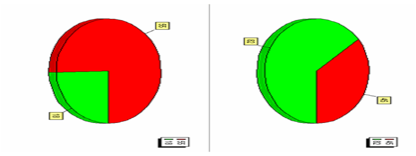
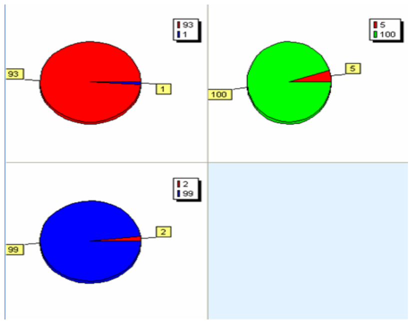


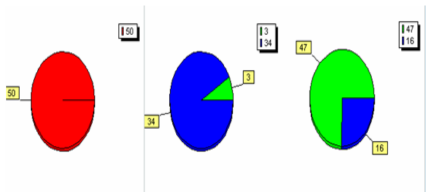
Fig 3: the comparative results of K-means algorithm and the Decision Model.



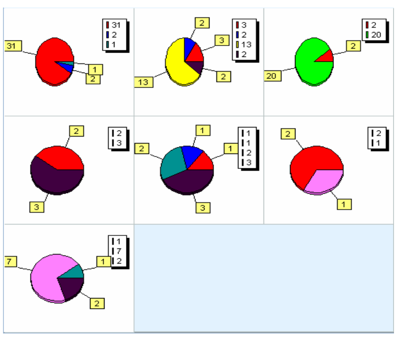
(?)3D-2C



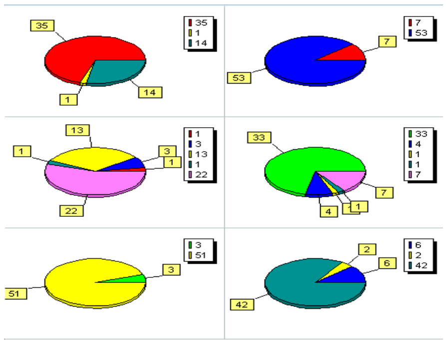
(?)3D-3C



(?) Iris



(?)Zoo



(?) Web1

(?) Web2

Fig 4: Six pie charts of six datasets clustering results produced by the Decision Model.

4. Conclusions and future work

The purpose of this paper is to present a novel decision model for fuzzy clustering ensemble. In our life, especially in some competitive activities, to select the optimal ones, both experts and the majority's opinion should be taken into consideration. Once the experts and the majority have consensus, final decision can be made, or some extra discussion should be taken. In this paper the same concept has been extended to combine data partitions which produced by fuzzy clustering algorithms. As we know, different clustering algorithms or a single clustering algorithm with different initializations will in general produce different partitioning results. In crisp clustering partitions, it is impossible to compare different crisp partitions with the validity functions based on the fuzziness of partitions, since the fuzziness of any crisp partition is zero, no ideal partition can be available. But in fuzzy clustering analysis, the optimal partition which called *expert* in this article can be easily chosen by the fuzzy cluster-validity criterion and opinions of the *majority* could be obtained through fuzzy voting scheme, finally, synthesizing the two matrixes results in the ultimate decision. If *expert* and the *majority* have differing opinions, fuzzy similarity measure would be adopted to solve this problem.

Based on theoretical analysis and experiments, we can conclude that the novel *decision model* is suitable for fuzzy clustering ensemble and it is better than the method which crisp clustering ensemble adopts. However, some cases when contradiction between *majority judgers* and *expert* happened should be discussed in the next step and ongoing work will be focused on designing a method or extrapolation of this methodology to fit for dealing with large data set in fuzzy clustering ensemble.

References

[1] A. Fred, Finding Consistent Clusters in Data Partitions, *Proceedings of the 2nd International Workshop on Multiple Classifier Systems, Volume 2096 of Lecture Notes in Computer Science*, Springet, pp.309-318, 2001.

[2] A Strehl, J Ghosh. Cluster Ensembles A Knowledge Reuse Frame work for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3:583-617, 2003.

[3] B. Minaei-Bidgoli, A Topchy, W F Punch. A Comparison of Resampling Methods for Clustering Ensembles. *Intl Conf on Machine Learning. Models, Technologies and Applications(MLMTA 2004)*, pp.939-945, 2004.

[4] T. Alexander, Anil K. Jain, and William Punch, Combining Multiple Weak Clusterings, *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)*, 2003.

[5] H. Ayad, M. Kamel, Finding Natural Clusters Using Multi-Clusterer Combiner Based on Shard Nearst Neighbors. *Proceedings of the 4th International Workshop on Multiple Classifier Systems(MCS'03), Volume 2709 of Lecture Notes in Computer Science*, Springer, pp.166-175, 2003

[6] B. Minaei-Bidgoli, A Topchy, W F Punch, A Comparison of Resampling Methods for Clustering Ensembles. *Intl Conf on Machine Learning Models, Technologies and Applications(MLMTA 2004)*, pp.939-945, 2004

[7] L. Y. Yang, H.R. Lv and W.Y. Wang, Soft Cluster Ensemble Based on Fuzzy Similarity Measure, *IMACS Multiconferenceon"CESA"*, pp.1994-1997, 2006.

[8] K. Josien, T. Warren Liao, Simultaneous grouping of parts and machines with an integrated fuzzy clustering method, *Fuzzy Sets and Systems*, 126:1-21, 2001.

[9] Julinda Gllavata, Ermir Qeli, and Bernd Freisleben, Detecting Text in Videos Using Fuzzy Clustering Ensembles, *Proceedings of the Eighth IEEE International Symposium on Multimedia (ISM'06)*, 2006.

[10] J. C. Bezdek, Cluster validity with fuzzy sets, *J. Cybernit.* 3:58-72, 1974.

[11] J. C. Bezdek, Numerical taxonomy with fuzzy sets, *J. Math.Biol.* 1:57-71, 1974

[12] G. Xinbo, Fuzzy Cluster Analysis and Application, *Xidian University Press*, pp.110-113, 2004.

[13] http://www.ics.uci.edu/mlearn/MLRepository.html.

[14] Y.Fukuyama, M. Sugeno, A new method of choosing the number of clusters for the fuzzy c-means method, *Proc.5th, Fuzzy Syst. Symp*, pp.247-250, 1989.

[15] J.M.F. Salido and S. Murakami, Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. *Fuzzy Sets and Systems*, 139:635-660, 2003.

[16] http://racr.shmtu.edu.cn/en_index.asp.

[17] H.R. Berenji and P.S. Khedkar, Using fuzzy logic for performance evaluation in reinforcement learning. *International Journal of Approximate Reasoning*, 18(3):131-144, 1998.