COENOSES 5 (1): 17-21, 1990

COMPARISON OF FUZZY CLASSIFICATIONS

J. Podani*, International Centre for Earth & Environmental Sciences, I.C.S.-I.C.T.P., via Beirut 7, I-34100 Trieste, Italy

Keywords: Consensus, Distance, Minimum length sequence metrics, Partitions

Abstract. Results of fuzzy clustering are influenced by the degree of fuzziness arbitrarily specified, so the comparative evaluation of fuzzy partitions is useful. A distance coefficient based on finding the best correspondence of clusters is suggested for such comparisons. Two algorithms are described for calculating consensus fuzzy partitions. Determination of the best overall correspondence of clusters results in a more exact solution but requires more computation than the heuristic algorithm of centroid clustering. The actual and artificial examples demonstrate that 1) the methods provide a useful means for the evaluation of community classifications, and 2) the methods equally apply to conventional and fuzzy classifications, allowing their simultaneous comparison.

Introduction

Since the introduction of the concept of fuzzy sets in mathematics (Zadeh 1965), increasing attention has been paid to clustering methods that produce fuzzy clusters. Whereas most traditional classification techniques depict some group structure even if it is in fact not present in the data, fuzzy clustering is less prone to produce such artefacts. Fuzzy clustering methods calculate membership weights for every object to express its affinity to each cluster, so that the first scrutiny of results gives more information on the existence of groups than that of conventional dendrograms and partitions. A recent paper by Bezdek (1987) gives a thoroughful description of fuzzy clustering algorithms and lists many useful references for better orientation in the mathematical foundations of methods. However, as apparent from this review and from the inspection of relevant literature, the number of ecological applications is relatively few. Roberts (1986) applied fuzzy sets for ordination of vegetation by defining fuzzy sets on the basis of continuous variables and by generalizing the ordination method of Bray and Curtis (1957). Fuzzy cluster membership weights were proposed by Feoli and Zuccarello (1986, 1988) and used by Banyikwa, Feoli and Zuccarello (1990) for constructing classification-based ordinations. Marsili-Libelli (1989) presented some algorithmic details of fuzzy clustering together with ecological examples. These studies sufficiently illustrate the power of fuzzy clustering for the ecologist and suggest that this approach will receive wider application in community studies in the near future.

Even though fuzzy clustering seems to have some advantages over the traditional 'deterministic' methods, or 'hard' clustering, these two families of methods share some properties. For example, the user has to decide on the number of clusters for both fuzzy and hard partitions. Moreover, generation of fuzzy partitions requires specification of the degree of fuzziness as well. In addition to choices related to a particular algorithm, the specific problems of the field of application also require many decisions to be made by the investigator. In a recent review I listed many examples showing that the selection of data type, subset of variables, and the parameters of sampling design, etc., may greatly influence classifications and ordinations, and these problems call for comparative evaluation of results (Podani 1989a). Obviously, this statement applies to fuzzy classifications as well.

Comparison of classifications may follow two basic strategies. In the first, two classifications are compared at a time and their distance or similarity is expressed by a single number. Given p alternative classifications for the same set of objects, the pairwise distance coefficients are written into a symmetric matrix which may be subjected to multivariate analyses to reveal trends in the original results ('multiple comparisons'). The other approach is concerned with the construction of a new classification which emphasizes agreements among the p alternatives. This new result is often termed as an 'average' or 'consensus' classification. In this paper I describe methods for evaluating fuzzy partitions based on both approaches. The utility of these methods is illustrated by artificial data and by examples taken from plant community studies.

Distance between fuzzy partitions

The techniques for comparing hard partitions (reviewed by Podani 1986 for applications in community studies) cannot be used for fuzzy partitions, as they require a deterministic description of classes (for example, in terms of cross-classification tables). The minimum length sequence metrics (Day 1981), however, although

* Permanent address: Department of Plant Taxonomy and Ecology, ELTE, Kun Béla tér 2, Budapest, Hungary, H-1083.

also inapplicable directly, provide a starting point for the development of an appropriate distance measure. These metrics require to find the best or optimum correspondence of clusters for the two classifications being compared, and then to count the number of elementary operations necessary to transform one partition to the other. The most commonly used type of these metrics, abbreviated as MINDMT, counts the MINimum number of Divisions, Mergences and Transfers (see Day 1981, for terminology) of single elements; this sum in fact gives the number of objects to be relocated to obtain one partition from the other. The optimization involves the rearrangement of cross classification tables so as to maximize the sum of diagonal values. This requires examination of all possible permutations of the rows of the table such that the columns remain unchanged. Then, the sum of non-diagonal values yields the distance required. In the sequel, an analogous procedure is described for fuzzy partitions.

A fuzzy partition of m objects into t clusters is usually described in terms of a matrix $U=\{u_{jk}\}$ of membership weights, where $0 \leq u_{jk} \leq 1$ represents the degree to which object j is assigned to cluster k, with the constraint that

$$\sum_{k=1}^{t} u_{jk}=1 \qquad (1)$$

for all j (see e.g., Marsili-Libelli 1989). Given a profile $F=\{F_1, F_2, ..., F_h, F_i, ..., F_p\}$ of p fuzzy partitions of the same set of m objects, we have the corresponding matrices of membership weights, i.e., $U_1, U_2, ..., U_h, U_i, ..., U_p$. The comparison of F_h and F_i from the profile should involve some comparison of U_h and U_i. An obvious measure is the minimum sum of squared changes of membership values in U_h necessary to transform F_h into F_i. The optimum correspondence of groups in F_h with the groups of F_i is found by checking all permutations of columns in U_h over the columns of U_i that remain unchanged, such that the quantity

$$d_{hi}^2=\sum_{j=1}^{m}\sum_{k=1}^{t}(u_{hjk}-u_{ijk})^2 \qquad (2)$$

is minimized. The minimum, i.e., the squared distance of F_h and F_i will be denoted by δ_{hi}^2. If the number of groups in F_h and F_i is not equal, the same function may be used by adding empty classes to the partition with the lower number of classes. The number of permutations to be examined is t!, so the minimization of (2) requires much computer time even for moderately small values of t.

It is easily seen that MINDMT is closely related to the above technique; for hard partitions δ_{hi}^2 equals twice the number of relocations to be made to transform F_h into F_i. Thus, the upper bound of δ_{hi}^2 is twice the upper bound of MINDMT. Note, however, that although the maximum contribution of one object to δ_{hi}^2 is 2, $\sup(\delta_{hi}^2)<2m$, because there are some unavoidable agreements between F_h and F_i. I have investigated this problem thoroughly (Podani 1986) and found a formula for defining the upper bound of MINDMT for the case when the number of clusters (t) is the same for both partitions:

$$\sup(\text{MINDMT})=\sup(\delta_{hi}^2/2)=m-t(\text{int}(m/t^2))-\alpha,$$

where

$$\alpha=\begin{cases}\text{int}(m-t^2(\text{int}(m/t^2))/t), & \text{if mod}(m-t^2(\text{int}(m/t^2))/k)=0;\\ \text{int}(m-t^2(\text{int}(m/t^2))/t)+1, & \text{otherwise.}\end{cases}$$

If the number of clusters is not the same, the upper bound can be determined by a simple computer algorithm operating on cross-classification tables.

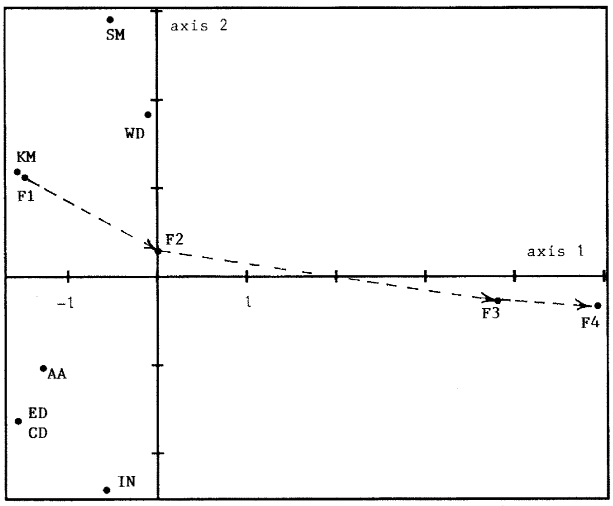

Fig. 1. Principal coordinates ordination of 11 partitions of 80 sample plots from the Sashegy Nature Reserve, Hungary. For abbreviations, see text. Arrows indicate increase of fuzziness.

Consensus fuzzy classifications

Of the several ways of constructing consensus objects (cf. Day 1988), the concept of median consensus will apply most directly to fuzzy partitions. The median consensus for p classifications is defined as a new classification for which the sum of squared distances from the input classifications is the minimum. As long as no exact solution is available for minimization, the two algorithms described below may be used to obtain good approximations to the optimum.

Method 1: Best overall correspondence of clusters

For p classifications each with t clusters there are exactly $t!^{p-1}$ different ways of matching the clusters. For each permutation an average partition is computed by taking the mean vector of weights, and the sum of squared distances from this average is calculated. Then the minimum of these sums is found to select the most optimal correspondence of clusters. Average cluster membership weights for the optimum match give a consensus classification F_c, represented by U_c, equal to or approximating the absolute optimum. In other words, we minimize the functional

$$s^2 = \sum_{i=1}^{p} \sum_{j=1}^{m} \sum_{k=1}^{t} (u_{ijk} - u_{cjk})^2, \quad (3)$$

where

$$u_{cjk} = \sum_{i=1}^{p} u_{ijk}/p. \quad (4)$$

The minimum of (3) for a given profile will be denoted by σ^2.

Method 2: Centroid clustering

Examination of all possible matches of clusters is a time-consuming procedure even for relatively small values of p and t. The heuristic algorithm described below provides a consensus fuzzy partition within a reasonably short time for large number of classifications and clusters, but usually gives a less accurate approximation to the optimum than does Method 1. The underlying strategy is a standard agglomerative centroid clustering algorithm whose final result is further improved if possible.

(1) Calculate δ^2_{hi} between all possible pairs of p partitions. The distance coefficients are written into matrix Δ.

(2) Identify the closest pair of partitions by finding the smallest non-diagonal value in Δ. Let these partitions be, say, F_h and F_i; $h<i$. Find the best match of U_h over U_i, and update U_h according to

$$u_{hjk} = (n_h\, u_{hjk} + n_i\, u_{ijk})/(n_h + n_i),$$

where n_h and n_i represent the number of partitions formerly joined in F_h and F_i, respectively. Update n_h according to $n_h = n_h + n_i$. Remove partition F_i from the profile.

(3) If $n_h < p$ then calculate the new distance between the new partition F_h and the remaining partitions, and update matrix Δ. Go to step (2). If $n_h = p$ then go to the final step.

(4) By minimizing function (2) find the best match of each original partition to F_c using formula (4).

It is always case-dependent if any improvement can be achieved in step (4).

Consensus efficiency and percentage contributions

After creating a consensus partition by Method 1 or 2 the question arises: how well represents F_c the profile F. σ^2 provides a means for expressing absolute consensus efficiency. Zero value indicates complete agreement among the alternative partitions, and a consensus partition identical to them. The maximum is influenced by p, m and t in a very complex manner (one reason is already mentioned in connection with the maximum of d^2_{hi}). It is left to future investigations to find a normalizing constant for σ^2. Without normalization, sum of squares coming from different consensus situations cannot be compared directly.

Assuming that the columns of each U_i have been rearranged for best match to those of U_c, the relative contribution of F_i to σ^2 is obtained as the percentage

$$x_i = 100 \sum_{j=1}^{m} \sum_{k=1}^{t} (u_{ijk} - u_{cjk})^2/\sigma^2 \quad (5)$$

Similarly, the relative contribution of object j to the sum of squares is given by

$$y_i = 100 \sum_{i=1}^{p} \sum_{k=1}^{t} (u_{ijk} - u_{cjk})^2/\sigma^2. \quad (6)$$

These percentages may be ranked in order to select most divergent classifications in F and to identify objects with the least consistent cluster membership weights. Removal of such partitions and objects from the data may considerably improve the efficiency of consensus generation. The percentages calculated according to

$$z_k = 100 \sum_{i=1}^{p} \sum_{j=1}^{m} (u_{ijk} - u_{cjk})^2/\sigma^2 \quad (7)$$

express the relative divergence of each consensus cluster from the original clusters matched to it. If the values of z differ with consensus classes considerably, it is advisable to try a new value of t in a repeated analysis.

Illustrative examples

The methods described above are demonstrated using data from the rock grassland communities of the Sashegy Nature Reserve, Budapest, Hungary. Detailed description of the site, communities and sampling design is not given here, this information is presented elsewhere (Podani 1985, 1986, 1989a). Suffice to mention here that different analyses have suggested the existence of three noda types without sharp community boundaries. The consensus methods will also be illustrated using a small set of artificial data. Computations were performed using programs of the SYN-TAX package (Podani 1988), and the new programs FCM for fuzzy c-means clustering (Bezdek 1987), written following Marsili-Libelli (1989), and FCOMP for multiple comparisons and consensus generation.

Example 1: multiple comparison of community classifications

Four fuzzy classifications were constructed based on the presence/absence data for 80 $4\times4m^2$ sample plots. The number of clusters was 3 in every case, whereas the coefficient of fuzziness was set to 1.05, 1.2, 1.5, and 1.8. These fuzzy partitions are abbreviated as F1, F2, F3 and F4, respectively. In addition, seven hard partitions are also considered for the comparison; six of them were obtained by 'cutting' dendrograms at the three-cluster level. The hierarchical clustering techniques were as follows: incremental sum of squares using Euclidean distance (ED), association-analysis based on mutual information between species (AA), furthest neighbor sorting from weighted dissimilarities (WD), minimization of average within-cluster similarity measured by the simple matching coefficient (SM), optimization of the increase of pooled entropy (IN), and chord distance with incremental sum of squares sorting (CD). For further details, see Podani (1985). The seventh hard partition was the most optimal 3-cluster solution selected from 10 results of k-means clustering (KM). The 11 partitions were compared in every pair, and then the distance matrix of partitions was subjected to principal coordinates analysis (PCoA) and nonmetric multidimensional scaling (NMDS) to obtain ordinations of partitions.

In the principal coordinates ordination of the 11 partitions, the percentages of variance accounted for by the first four components are 38%, 21%, 15% and 11%. The scattergram of the first two axes provides a well-interpretable configuration. The hard partitions are clustered on the negative side of axis 1. KM is the most similar to F1, and increases of the coefficient cause increased distances from FM and from the other hard partitions. The diagram supports the obvious relationship between c-means and k-means clustering: the lower the coefficient of fuzziness the closer are the two approaches. (Note that 1 cannot be used as coefficient of fuzziness because of singularity problems).

The best two-dimensional solution for the 11 partitions (Fig. 2) was obtained starting from the PCoA configuration (stress = 0.12). This ordination agrees well with the two-dimensional PCoA result. In view of the consensus generation (Example 3, see below), however, NMDS depicts the relationship of KM to CD, ED, AA and WD more faithfully than PCoA. The explanation is that the first two PCoA dimensions are not sufficient to portray such fine structural details, whereas NMDS was forced to summarize all information in two dimensions. The relative closeness of CD and ED to KM was expected anyway, because all the three methods use the same sum of squares criterion.

This example showed that comparison of fuzzy classifications with hard partitions allows for the assessment of the effect of coefficient of fuzziness. Together with the consensus generation approach (Example 3), such an analysis reveals the relative impact of our decisions when selecting among alternative methods of classification.

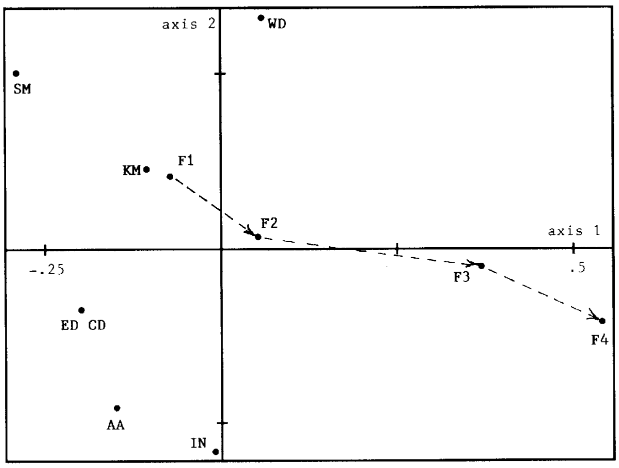

Fig. 2. Nonmetric multidimensional scaling of 11 partitions of 80 sample plots from the Sashegy Nature Reserve, Hungary. For abbreviations, see text. Arrows indicate increase of fuzziness.

Example 2: fuzzy consensus of artificial hard partitions

Although developed for fuzzy classifications, the use of consensus generating methods suggested in this paper will be first demonstrated based on a set of three artificial hard partitions of six objects into three clusters. These are:

$F_1 = \{1, 2\}, \{3, 4\}, \{5, 6\}$

$F_2 = \{1, 2, 3\}, \{4\}, \{5, 6\}$

$F_3 = \{1, 2\}, \{3, 4, 5\}, \{6\}$

The matrices of cluster membership weights are not presented here. Using such a simple example, one can easily compare the results with his own expectations. In addition, the example will show that a fuzzy partition may be used efficiently as a consensus of hard partitions. It also demonstrates that by applying the majority rule (Day 1988) to a fuzzy consensus partition, one can easily generate a hard consensus partition.

Both consensus generating methods produced the same result: the best match of clusters happened to be the same arrangement that specifies the partitions above. The weights for the consensus partition are given by

$$U_c = \begin{pmatrix} 1.0 & 0 & 0 \\ 1.0 & 0 & 0 \\ .33 & .66 & 0 \\ 0 & 1.0 & 0 \\ 0 & .33 & .66 \\ 0 & 0 & 1.0 \end{pmatrix}$$

Only objects 3 and 5 have a really fuzzy membership in the consensus partition. They contribute 50% each to the sum of squares ($\sigma^2=2.66$). The contribution of F_1 to σ^2 is lower than that of the other two classifications ($x_1=16.66\%$, $x_2=x_3=41.66\%$). It is plausible, because the hard partition generated by considering the highest weights for each object in the consensus partition, an operation analogous to the application of the majority rule for hard partitions, is identical to F_1. It is also straightforward that consensus cluster 2 has higher contribution to σ^2 (50%) than the other two (25% each).

Example 3: consensus community classification

For generating the consensus classification of the 11 partitions described in Example 1, only the centroid clustering approach was used. Examination of all possible matches of clusters would have involved 6^{10} computational steps. The rank order of percentage contributions to the error sum of squares is as follows: F2 (0.8%), F1 (4.7%), KM (5.7%), SM and CD (6.8% each), AA (8.5%), IN and F3 (11.4% each), SM (12%), WD (12.8%) and F4 (19%). These values confirm that the NMDS ordination depicts the interrelationship of results better than the first two PCoA axes. Removal of the two fuzziest partitions (F3 and F4) did not change this rank order, suggesting that in this study F2 represents most closely the other results. An important conclusion is that in this way we may select an 'optimum value' of the coefficient of fuzziness; in this case it proved to be about 1.2. The percentage contributions of objects to the sum of squares fall between 0.4% and 3.6%. Objects with values larger than 3% are responsible for most of the differences among the classifications. In particular, sampling units #56-59 and 80 should be mentioned here. Their assignment to clusters is the most ambiguous suggesting that they are the least classifiable objects in the present classification system. The percentage contributions of consensus clusters to the sum of squares are 44.7%, 34% and 21.3%, respectively. Cluster 1 corresponds to the community type representing a transition between the other two (cf. Podani 1989b), and it explains the more uncertain group membership of the sample plots included. Finally, it is worth mentioning that the hard consensus partition obtained by selecting the highest weight for each object in the consensus is almost identical to the 3-cluster level hard consensus classification obtained by the MINGFC method for CD, ED, AA, SM, IN and WD (Podani 1989b). Thus, the fuzzy consensus approach serves as an alternative to the other consensus generating methods applicable exclusively to hard partitions. The question remains to be answered whether a fuzzy consensus partition can be considered as a more faithful representation of a profile of hard partitions than a hard, and perhaps not unique, consensus classification.

Acknowledgements. This research was supported by an ICTP grant to me during my stay in Trieste. I have benefitted much from discussions with Prof. E. Feoli on the necessity of fuzzy classifications.

REFERENCES

BANYKWA, F., E. FEOLI and V. ZUCCARELLO. 1990. Fuzzy set system ordination of Serengeti short grasslands, Tanzania. J. Vegetation Sci. 1: 97-104.

BEZDEK, J.C. 1987. Some non-standard clustering algorithms. In: P. Legendre and L. Legendre (eds), Developments in Numerical Ecology. pp. 225-287. Springer, New York.

BRAY, J.R. and J.T. CURTIS. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27: 325-349.

DAY, W.H.E. 1981. The complexity of computing metric distances between partitions. Math. Soc. Sci. 1: 269-287.

DAY, W.H.E. 1988. Consensus methods as tools for data analysis. In: H.H. Bock (ed.), Classification and Related Methods of Data Analysis. pp. 317-324. North-Holland, Amsterdam.

FEOLI, E. and V. ZUCCARELLO. 1986. Ordination based on classification: yet another solution? Abstracta Botanica 10: 203-219.

FEOLI, E. and V. ZUCCARELLO. 1988. Syntaxonomy: a source of useful fuzzy sets for environmental analysis? Coenoses 3: 141-147.

MARSILI-LIBELLI, S. 1989. Fuzzy clustering of ecological data. Coenoses 4: 95-106.

PODANI, J. 1985. Syntaxonomic congruence in a small-scale vegetation survey. Abstracta Botanica 9: 99-128.

PODANI, J. 1986. Comparison of partitions in vegetation studies. Abstracta Botanica 10: 235-290.

PODANI, J. 1988. SYN-TAX III. A package of programs for data analysis in community ecology and systematics. Coenoses 3: 111-119.

PODANI, J. 1989a. Comparison of classifications and ordinations of vegetation data. Vegetatio 83: 111-128.

PODANI, J. 1989b. A method for generating consensus partitions and its application to community classification. Coenoses 4: 1-10.

ROBERTS, D.W. 1986. Ordination on the basis of fuzzy set theory. Vegetatio 66: 123-131.

ZADEH, L.A. 1965. Fuzzy sets. Inf. Control 8: 338-353.