INFORMATICA, 1992, Vol.3, No.4, 582–591

MAX-MIN AND MAX-Δ TRANSITIVE RESEMBLANCE MATRICES IN CLASSIFICATION PROBLEMS

Sergey VATLIN

Minsk Radioengineering Institute

220600 Minsk, P.Brovka St. 6, Belorussian Republic

Viktor KRASNOPROSHIN

Belorussian State University

220080 Minsk, Skorina Av. 4, Belorussian Republic

Abstract. The basic properties and methods of developing max-min and max-Δ transitive approximations of resemblance matrices of observed objects are reviewed. A new algorithm of constructing max-Δ transitive closure of such matrices is presented. The conditions of applications of the max-min and max-Δ transitive measures of similarity are considered.

Key words: classification problems, measure of similarity, max-min and max-Δ transitivity, transitive closure.

1. Introduction. Classification is a tool of ordering objects according to the discovery of their similarity and (or) distinction. Before such ordering takes place it is necessary to fix some manner of representing the initial information about observed objects $X_1, X_2, \ldots, X_M$. One of the most common approaches to solving this problem is concerned with application of matrices of paired comparison $R = ||r_{ij}||_{i,j=\overline{1,M}}$. The element r_{ij} of such a matrix reflects the results of the comparison of X_i and X_j objects in the sense of some fixed relation. In particular r_{ij} may reflect a measure of similarity of objects X_i and X_j, $i,j = \overline{1,M}$. In this case, as a rule, the condition

$$r_{ij} = r_{ji}, \quad \forall i,j = \overline{1,M} \tag{1}$$

is satisfied.

Without loss of generality in further considerations we also assume that the similarity measure of the X_i and X_j objects satisfies the conditions:

$$r_{ij} \in [0,1], \quad \forall i,j = \overline{1,M}, \tag{2}$$

$$r_{ii} = 1, \quad \forall i = \overline{1,M}. \tag{3}$$

It follows from (1) – (3) that under fixed $X = \{X_1, X_2, \ldots, X_M\}$ the R matrix is a symmetrical and reflexive fuzzy relation (Kaufmann, 1975).

The square of the fuzzy relation R is such fuzzy relation $R^2 = R \circ R$, that $R^2 = ||r_{ij}^{[2]}||_{i,j=\overline{1,M}}$ and

$$r_{ij}^{[2]} = \max_{1 \leqslant k \leqslant M} \left[\min(r_{ik}, r_{kj})\right], \quad \forall i,j = \overline{1,M}.$$

The arbitrary power of the fuzzy relation R is defined by induction

$$R^3 = R^2 \circ R,\ R^4 = R^3 \circ R,\ \ldots,\ R^M = R^{M-1} \circ R,\ \ldots$$

Various types of resemblance structures fixed in the R matrix may serve as a basis for finding the order in a set of observed objects of X. So, for example, if the condition

$$r_{ij} \geqslant \min(r_{ik}, r_{kj}), \quad \forall i,j,k = \overline{1,M} \tag{4}$$

is satisfied it is said that R is a max-min transitive measure of similarity. It is obvious that the property of max-min transitivity can be written in the form

$$R \circ R = R^2 \subseteq R.$$

(Here and throughout this paper we use the traditional definitions of operations on a set of fuzzy relations of the fixed dimension W (Kaufmann, 1975). That is, if $A, B \in W$, $A = ||a_{ij}||$, $B = ||b_{ij}||$ then $C = A \cup B = ||c_{ij}||$ so that $c_{ij} = \max(a_{ij}, b_{ij})$ and $A \subseteq B \Leftrightarrow a_{ij} \leqslant b_{ij}$, $i,j = \overline{1,M}$).

The result of the classification procedure is partitioning the whole set X into separate subsets (classes) of similar objects. It is important to emphasize that the similarity structure on X obtained in this way depends not only on the similarity fixed in the R matrix but on the type of the respective classification algorithm.

The application of various classification procedures may lead to different results in spite of the fact that the initial sample doesn't change. This means that such procedures not only disclose the real regularities but fit the experimental data into a preliminarily fixed model.

To agree the above mentioned contradictory tendencies in most classification algorithms special intermediate objects – max-min transitive approximations of the initial R matrices, are used. Further, we briefly survey the basic properties and the construction methods of such approximations.

2. Max-min transitive resemblance matrices in classification problems. In practice, subsets of similar objects are often organized in the form of hierarchical trees or hierarchies.

The hierarchy of S on $X = \{X_1 X_2, \ldots, X_M\}$ is such a system of subsets $\{s/s \subset X\}$ that

1. $X \in S$.
2. $\{X_i\} \in S$, $\forall i = \overline{1, M}$.
3. $s_l \in S$, $s_p \in S$ and $s_p \cap s_l \neq \varnothing \Rightarrow s_p \subset s_l$ or $s_l \subset s_p$.

The remarkable property of hierarchical classifications connected with the consistency of two different similarity structures is given by the following theorem.

Theorem 1. (Johnson, 1967) *The necessary and sufficient condition of constructing a hierarchy on the set X is the existence of the max-min transitivity similarity measure $R: X \times X \to [0, 1]$.*

However, under real conditions, the similarity matrix of observed objects may not satisfy the property (4).

To overcome this restriction in (Hardine, Sibson, 1971) it was proposed to construct a family $W^k = ||w_{ij}^k||_{i,j=\overline{1M},\ k=1,2,\ldots}$ of all these similarity matrices of observed objects that

$$r_{ij} \leqslant w_{ij}^k, \quad \forall i,j=\overline{1,M},\ k=1,2,\ldots,$$
$$w_{ij}^k \geqslant \min(w_{il}^k, w_{lj}^k), \quad \forall i,j,l=\overline{1,M},\ k=1,2,\ldots.$$

It is obvious that $\{W^k\}$ $k=1,2,\ldots$ is not empty and putting

$$\widehat{W} = \|w_{ij}\|, \quad i,j=\overline{1,M},$$
$$w_{ij} = \min_k w_{ij}^k, \quad \forall i,j=\overline{1,M},$$

we will obtain the suboptimal (Hardine, Sibson, 1971) max-min transitivity measure of similarity $\widehat{W}$.

In the theory of fuzzy sets the criterion of suboptimality is formulated by means of the following concept.

A max-min transitive closure of the fuzzy relation R is such fuzzy relation $\widehat{R}$ that $\widehat{R} = R \cup R^2 \cup \ldots \cup R^M \cup \ldots$.

It is easy to show that $\widehat{R}$ is always a transitive fuzzy relation.

In the case when R is an arbitrary fuzzy relation on a set of observed objects X and Card $X = M$ the following condition

$$R \cup R^2 \cup \ldots \cup R^M \cup \ldots = R \cup R^2 \cup \ldots \cup R^M$$

takes place (Kaufmann, 1975), and therefore

$$\widehat{R} = R \cup R^2 \cup \ldots \cup R^M. \tag{5}$$

Besides, if R is a similarity measure on a set of elements from X, then $R \subseteq R^2 \subseteq \ldots \subseteq R^{M-1} = R^M$ and therefore

$$\widehat{R} = R^{M-1}. \tag{6}$$

The relation (6) provides a formal method of constructing a suboptimal max-min transitive similarity matrix and by means of it defines the hierarchy structure on a set of elements of X.

A similar but more effective in a computational way method of constructing matrix $\widehat{R}$ is given in (Low, Tong, 1981).

3. Max-Δ transitive resemblance matrices in classification problems. Hierarchies are a useful tool in solving many problems of pattern recognition, forecasting and data analysis. The characteristic features of hierarchical classification algorithms are high speed of data processing and a relative simplicity of their programme implementation. A possibility of graphical interpretation of the obtained results is also an essential advantage of these algorithms.

At the same time such algorithms have a rather serious disadvantage. The necessary and sufficient condition of developing hierarchies postulates very "hard" restrictions on the internal structure of similarity represented in the matrix. The formal aspect of this "hardness" is illustrated by the following theorem.

Theorem 2. (Kaufmann, 1975). *Let $R = ||r_{ij}||_{i,j=\overline{1,M}}$ be a max-min transitive similarity matrix on a set of observed objects of X. Let also X_1, X_2, X_3 be three arbitrary elements from X. Denote $a = r_{12}$, $b = r_{13}$, $c = r_{23}$. Then*

$$a \geqslant b = c \quad \text{or} \quad b \geqslant c = a \quad \text{or} \quad c \geqslant a = b. \tag{7}$$

The statement of this theorem means that in the space $\langle X,\ D\rangle$ where $D = I - R(D = ||d_{ij}||, d_{ij} = 1 - r_{ij})$ is a dual with respect to R measure of distinction on elements from X, every triangle is either equilateral or isosceles.

The following example is also an illustration of the constraints which are imposed on the structure of the experimental data by the condition (4). Suppose, as a result of an experiment, we obtain the sequence of measurements $\{a_i\}_{i=1}^{N}$, $a_i \in \mathbf{R}^1$, $i = \overline{1,N}$, that

$$a_1 \leqslant a_2 \leqslant \ldots \leqslant a_N.$$

The requirement of max-min transitivity of this sequence immediately implies the following relation (Bezdek, Harris, 1978):

$$a_1 = a_2 = \ldots = a_{N-1} \leqslant a_N. \tag{8}$$

It is evident that the constraints of the type (7) - (8) are very strong and for practical purposes they can hardly be found valid.

In order to solve the above problem in (Bezdek, Harris, 1978) the concept of a max-Δ transitive similarity measure on a set of observed objects of X was introduced. First of all, we will note that operation Δ for two real numbers is defined as follows:

$$a \Delta b = \max(a + b - 1, 0).$$

The similarity matrix $R = ||r_{ij}||_{i,j=\overline{1,M}}$ will be called max-Δ transitivity iff

$$r_{ij} \geqslant r_{ik} \Delta r_{kj}, \quad \forall i, j, k = \overline{1, M}. \tag{9}$$

The characteristic of such matrices is that measures of distinction $D = I - R$ dual with respect to them are distances. In this case the space $\langle X, D \rangle$ is a metric space which is the most useful for description of the majority of physical and mathematical systems (Bezdek, Harris, 1978).

Though condition (9) is much weaker than condition (4), in real situations it also may be unsatisfied. Therefore for practice it is very important to have regular procedures allowing us to develop "good" max-Δ transitive approximations to the initial data.

The rest of the paper is devoted to the development of a formal ground on the basis of which the suboptimal strategy of constructing such approximations can be realized.

The max-Δ square of fuzzy relation $R = ||r_{ij}||_{i,j=\overline{1,M}}$ is such fuzzy relation $R^2_\Delta = R \Delta R$, that $R^2_\Delta = ||r^*_{ij}||_{i,j=\overline{1,M}}$ and

$$r^*_{ij} = \max_{1 \leqslant k \leqslant M} [r_{ik} \Delta r_{kj}], \quad \forall i, j = \overline{1, M}.$$

The arbitrary max-Δ power of the fuzzy relation R is defined by induction

$$R^3_\Delta = R^2_\Delta \Delta R, \ R^4_\Delta = R^3_\Delta \Delta R, \ \ldots, \ R^M_\Delta = R^{M-1}_\Delta \Delta R, \ \ldots.$$

The max-Δ transitive closure of fuzzy relation R is such fuzzy relation $\widehat{R}_\Delta$, that

$$\widehat{R}_\Delta = R \cup R^2_\Delta \cup R^3_\Delta \cup \ldots \cup R^M_\Delta \cup \ldots.$$

It is easy to check that $\hat{R}_\Delta$ will always satisfy the condition of max-Δ transitivity (9).

As a result of reasonings analogous to the situation with a max-min transitive closure it can be shown that if R is a similarity matrix on a set of elements from X and Card $X = M$, the following equality holds

$$\hat{R}_\Delta = R_\Delta^{M-1}. \tag{10}$$

The relationship (10) gives a practical algorithm of building-up suboptimal max-Δ transitive similarity matrix on a set of elements from X. Thus, for example, if $X = \{X_1, X_2, X_3, X_4\}$ and

$$R = \begin{pmatrix} 1 & 0.20 & 0.45 & 0.85 \\ 0.20 & 1 & 0.90 & 0.15 \\ 0.45 & 0.90 & 1 & 0.05 \\ 0.85 & 0.15 & 0.05 & 1 \end{pmatrix},$$

then

$$R_\Delta^2 = R \Delta R = \begin{pmatrix} 1 & 0.35 & 0.45 & 0.85 \\ 0.35 & 1 & 0.90 & 0.15 \\ 0.45 & 0.90 & 1 & 0.30 \\ 0.85 & 0.15 & 0.30 & 1 \end{pmatrix},$$

$$R_\Delta^3 = \hat{R}_\Delta^2 \Delta R = \begin{pmatrix} 1 & 0.35 & 0.45 & 0.85 \\ 0.35 & 1 & 0.90 & 0.20 \\ 0.45 & 0.90 & 1 & 0.30 \\ 0.85 & 0.20 & 0.30 & 1 \end{pmatrix}.$$

In general case, when Card $X = M$ for obtaining each of $(M-1)\cdot M/2 + M$ elements of the top triangle of matrix R_Δ^i, $i = \overline{2, M-1}$ it is necessary to perform M operations Δ. Hence, the algorithm under consideration has computational complexity of the order of $O(M^3)$.

Forming $M-1$ $(M \geqslant 2)$ power of an arbitrary R similarity matrix requires storing all elements of its top triangle. Therefore, the requirement of the above mentioned algorithm in fixed memory amounts to $(M-1)\cdot M/2 + M$.

4. New algorithm for constructing max-Δ transitive closure of similarity matrix. This TC (Transitive Closure) algorithm involves the following sequence of steps:

Step 1. Set $\hat{r}_{ij}^{\circ} = r_{ij}$, $\forall i, j = \overline{1, M}$.

Step 2. Put $k = 1$.

Step 3. Compute $\hat{r}_{ij}^{k} = \max[\hat{r}_{ij}^{k-1},\ \hat{r}_{ik}^{k-1} \Delta \hat{r}_{kj}^{k-1}]$, $\forall i, j = \overline{1, M}$.

Step 4. If $k < M$, increment k by one and go to Step 3.
Else

Step 5. Set $\hat{r}_{ij} = \hat{r}_{ij}^{k}$, $\forall i, j = \overline{1, M}$ and STOP.

It can be proved that the matrix $\hat{R} = ||\hat{r}_{ij}||_{i,j=\overline{1,M}}$ formed at the output of this procedure will satisfy the property of max-Δ transitivity. The method of proving is based on the traditional scheme of constructing the algorithms of max-min transitive closure of fuzzy relations (Kaufmann, 1975) and that is why it is not given here.

Let, as before, $X = \{X_1, X_2, X_3, X_4\}$ and

$$R = \begin{pmatrix} 1 & 0.20 & 0.45 & 0.85 \\ 0.20 & 1 & 0.90 & 0.15 \\ 0.45 & 0.90 & 1 & 0.05 \\ 0.85 & 0.15 & 0.05 & 1 \end{pmatrix}.$$

Then according to the TC algorithm we will have

$$\hat{R}^{o} = R = \begin{pmatrix} 1 & 0.20 & 0.45 & 0.85 \\ 0.20 & 1 & 0.90 & 0.15 \\ 0.45 & 0.90 & 1 & 0.05 \\ 0.85 & 0.15 & 0.05 & 1 \end{pmatrix},$$

$$\hat{R}^{1} = \begin{pmatrix} 1 & 0.20 & 0.45 & 0.85 \\ 0.20 & 1 & 0.90 & 0.15 \\ 0.45 & 0.90 & 1 & 0.30 \\ 0.85 & 0.15 & 0.30 & 1 \end{pmatrix},$$

$$\hat{R}^{2} = \begin{pmatrix} 1 & 0.20 & 0.45 & 0.85 \\ 0.20 & 1 & 0.90 & 0.15 \\ 0.45 & 0.90 & 1 & 0.30 \\ 0.85 & 0.15 & 0.30 & 1 \end{pmatrix},$$

$$\hat{R}^{3} = \hat{R}^{4} = \begin{pmatrix} 1 & 0.35 & 0.45 & 0.85 \\ 0.35 & 1 & 0.90 & 0.20 \\ 0.45 & 0.90 & 1 & 0.30 \\ 0.85 & 0.20 & 0.30 & 1 \end{pmatrix}.$$

It is easily seen, that the given algorithm has computational complexity of an order of $O(M^3)$ and its requirement in fixed memory is also equal to $P = (M-1) \cdot M/2 + M$.

From the computational point of view the main effect of the TC-algorithm is reduced to the optimization of the needed core memory capacity. Really, when applying the relationship (10) it is necessary to save 2K elements of intermediate calculations. In the case of the application of the TC-algorithm it is sufficient to save only K such elements.

5. Conclusion. Thus, we have reviewed the basic properties and methods of generating max-min and max-Δ transitive approximations of resemblance matrices of observed objects. A new algorithm of constructing max-Δ transitive closure of such matrices has been presented.

The $\widehat{R}$ and $\widehat{R}_{\Delta}$ matrices serve as a basis for forming (systems) classifications on a set of experimental data. Classifications developed on the basis of these matrices will be different, since max-min and max-Δ transitivities define various structures of similarity in X.

In the case of max-min transitivity the condition of consistency of the internal structure of similarity and the structure of similarity, defined by the classification algorithm, is given by Theorem 1. The development and analysis of an analogous condition for situations with max-Δ transitive structures of internal similarity is the object of further investigations.

Acknowledgements. The authors would like to thank an anonymous referee for his useful comments on an earlier version of this paper.

REFERENCES

Bezdek, J.C., J.D.Harris (1978). Fuzzy partitions and relations: An axiomatic Basis for clustering. *Fuzzy Sets and Systems,* 1(1), 111–117.

Hardine, N., R.Sibson (1971). *Mathematical Taxonomy.* J.Willey Sons, London. 231 pp.

Johnson, S.C. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**(3), 241–254.

Kaufmann, A. (1975). *Introduction to the Theory of Fuzzy Subsets*, Vol.1.*Fundamental Theoretical Elements*, Academic Press, New York. 325 pp.

Low, S.P., Z.X.Tong (1981). An Algorithm for finding the transitive closure of fuzzy matrix. *Journal of Shanghai Railway Institute*, **1**(1), 37–42.

Received June 1992

S. Vatlin received the Degree of Candidate of Technical Science from the Minsk Radio Engineering Institute (MREI) in 1990. He is an associate professor at the Computer Science Department, MREI. His interests include mathematical theory of classification, concept learning and pattern recognition.

V. Krasnoproshin is a Head of the Department of mathematical modelling and automatic control systems of the Byelorussian State University. He received the Degree of Candidate of Physical and Mathematical Science from the Institute of Mathematics, Minsk, Acad. Sci. Rep. of Byelorus. Author and co-author of 3 books and more than 90 papers. His research interests include artificial intelligence, pattern recognition and their application to medical diagnostic and ecological monitoring.