ISE-TR-88-74

A METHOD OF HIERARCHICAL CLUSTER ANALYSIS WITH A FUZZY CONSTRAINT AND ITS RELATION TO WISHART'S KTH NEAREST NEIGHBOR METHOD

by

Sadaaki MIYAMOTO

June 14, 1988

INSTITUTE
OF
INFORMATION SCIENCES AND ELECTRONICS

UNIVERSITY OF TSUKUBA

A Method of Hierarchical Cluster Analysis with a Fuzzy Constraint and Its Relation to Wishart's Kth Nearest Neighbor Method

S. Miyamoto

Institute of Information Sciences and Electronics

University of Tsukuba, Ibaraki 305, Japan

ABSTRACT

This paper proposes a new method of hierarchical clustering as follows. Assume a set of objects (v(1),..,v(n)), a proximity measure S(i,j) between v(i) and v(j), and a fuzzy set c(1)/v(1) +...+ c(n)/v(n). The transitive closure of a relation S'(i,j) = min[S(i,j),c(i),c(j)] is considered. This method differs from the nearest neighbor method on the point that the membership c(i) which is called here a fuzzy constraint is taken into account. A well-known technique of the Wishart's Kth nearest neighbor method is proved to be a special case of the method herein. Hence it is evident that the property of the label freedom holds for the Wishart's method and an algorithm for the minimal spanning trees can be applied to the Wishart's method. Moreover, the method herein suggests many versions other than the Wishart's method for improving the nearest neighbor clustering.

1. Introduction

There are three types of major contributions of fuzzy set theory to methods of cluster analysis. First, Tamura and others [1] and Zadeh [2] showed that max-min transitive closure of a reflexive and symmetric fuzzy relation (sometimes called a proximity relation) is a fuzzy equivalence relation, which in turn was shown to be equivalent to a minimal spanning tree (MST) by J. C. Dunn [3]. As MST is equivalent to a well-known method of the nearest neighbor clustering (Anderberg, [4]), the transitive closure of a proximity relation is equivalent to the nearest neighbor method. (See also Ohsumi [5].) Secondly, the k-means method of nonhierarchical clustering was generalized to fuzzy c-means method by Bezdek [6] and active researches are going on in this direction (See, e.g., Hirota and others [7]. Third, measures of relatedness which are called similarities or dissimilarities for hierarchical clustering were generalized using a fuzzy set model, which led to a new algorithm of hierarchical clustering (Miyamoto and Nakayama, [8]).

Various interesting researches will be done in future for all of the above three kinds of considerations, and in this paper we show an application of the idea of the first type. Namely, we apply the idea of the transitive closure to a fuzzy relation with a fuzzy constraint on every vertex. A method of cluster analysis that is equivalent to the transitive closure of the fuzzily constrained relation is considered using a fuzzy graph. This method is proved to include a well-known method of the Wishart's Kth nearest neighbor clustering (Wishart, [9], [10]). As a result, the Wishart's method is proved to be a version of the nearest neighbor method with a modified similarity measure. The last statement does not reduce the value of the Wishart's method. On the contrary, it shows that the Kth nearest neighbor method enjoys theoretical properties of the nearest neighbor method such as the label freedom and applicability of efficient algorithms for the MST.

2. Preliminary results

Let $V=(v_1,v_2,...,v_n)$ be a finite set of objects for clustering. A fuzzy relation R on V x V is assumed to be given. The relation R is reflexive and symmetric: $R(v,v)=1$ and $R(v,w)=R(w,v)$ for all $v,w \in V$. Since the set U is finite, the relation R is idetified with a matrix (r_{ij}), $r_{ij}=R(v_i,v_j)$, $1 \leq i,j \leq n$. In this paper we do not distinguish a fuzzy relation and its matrix representation for simplicity. Note that this abuse of terminology does not induce any confusion. An alpha-cut (alpha level set) of a fuzzy set A is denoted by $C(\alpha)A$ in this paper. Similarly, an alpha-cut of R is represented as $C(\alpha)R$.

A basic fuzzy graph BFG is the pair (V,R). It is a collection of crisp graphs $(V,C(\alpha)R)$ for $0 \leq \alpha \leq 1$. Namely, a pair of vertices $v,w \in V$ has the edge when $\alpha \leq R(v,w)$. Since R is symmetric, BFG is an undirected fuzzy graph. We call BFG as a "basic" fuzzy graph to distinguish it from another version of a fuzzy graph which will be introduced later. We assume that readers are familiar with basic definitions of crisp graph theory such as paths, connected components, the minimal spanning trees (See, e.g., Aho and others, [11]). Note that we consider a maximal spanning tree instead of the minimal spanning tree.

Given two nxn matrices $S=(s_{ij})$ and $T=(t_{ij})$ such that $s_{ij}, t_{ij} \in [0,1]$, $1 \le i,j \le n$, we assume that arithmetic operations on the scalars are defined by maximum for addition and minimum for multiplication, as is usual in fuzzy set theory. Accordingly, we have

$$S + T = (s_{ij} + t_{ij}) = (\max(s_{ij}, t_{ij}))$$
$$S\ T = (\sum_k s_{ik} t_{kj}) = (\max_k \min (s_{ik}, t_{kj}))$$

Moreover, the lattice product (Kandel, [12]) of two matrices are defined by

$$S{*}T = (\min(s_{ij}, t_{ij})) .$$

The following lemma is well-known, therefore we omit the proof.

<u>Lemma</u> <u>1</u> Assume that the relation R is reflexive and symmetric. The sequence $R + R^2 + \ldots + R^k$ is convergent as k goes to infinity. Define $R^{*} = R + R^2 + \ldots.$. Then

$$R^{*} = R + \ldots + R^{n-1}$$

Moreover, R^{*} is reflexive, symmeric, and transitive:

$$R^{*}(v,w) \ge \min[R^{*}(v,u), R^{*}(u,w)] \text{ for any } u \in V.$$

[]

<u>Remark</u> A reflexive, symmetric, and transitive fuzzy relation is sometimes called a fuzzy equivalence relation. []

A partition of V is a finite family of subsets $\{V_1, V_2, \ldots, V_m\}$ such that $\bigcup_i V_i = V$, $V_i \cap V_j = \emptyset$ $(i \ne j)$. A pair of partitions $\{V_1, \ldots, V_p\}$ and $\{W_1, \ldots, W_q\}$ of the same set V is called equivalent if p=q and there is a permutation sigma on

$\{1,2,\ldots,p\}$ such that $V_i = W_{sigma(i)}$, $i=1,\ldots,p$. that is, the partitions are identical except the order of members.

Second lemma on equivalence of four methods is also known and the proof is omitted.

Lemma 2 The four methods of the nearest neighbor clustering, the transitive closure of a proximity relation, connected components of BFG, and the maximal spanning tree is equivalent in the following sense. Given any parameter $\alpha \in [0,1]$, the following four partitions generated by the four methods are all equivalent.

(i) The clusters generated at the level of similarity α using the nearest neighbor method based on the measure $R(v,w)$ of similarity.

(ii) Equivalence classes $\{V_1,\ldots,V_p\}$ generated from $C(\alpha)R^*$: a pair $v_i, v_j \in V$ belongs to the same class, say V_q ($v_i, v_j \in V_q$) if and only if $C(\alpha)R^*(v_i,v_j)=1$. (In other words, $R^*(v_i,v_j) \geq \alpha$).

(iii) Connected components as a subset of vertices of $(V, C(\alpha)R)$ derived from $BFG=(V,R)$.

(iv) Consider a network that is a complete graph whose set of vertices is V and the weight $R(v,w)$ is given on the edge (v,w), $v,w \in V$. Apply an algorithm of the maximal spanning tree to the network. From the resulting tree T, delete those edges (v,w) such that $R(v,w) < \alpha$. Then we have a forest $\{T_1,\ldots,T_s\}$. Let V_j be the subset of vertices of T_j, $j=1,\ldots,s$. Then $V_1,\ldots,V_s$ forms a partition of V. This partition is equivalent to the partitions obtained in (i), (ii), and (iii). []

Remark Although the author believes the above result of the equivalence among the four kinds of partitions is already known as a fact, he does not know any publication on which the result is proved in a perfect manner. Since the purpose of the present is not to exhibit the above result, we omit the proof. Readers who are interested in the lemma 2 may try its proof. A guideline to its proof is to show equivalence between the nearst neighbor method and the Kruskal's algorithm for the MST given in [11], then to see the Kruskal's algorithm generates the connected components of BFS. Any algorithm of the MST generates the same partition according to the way in (iv) should also be shown. Finally, it is easily seen that the connected components are equivalent to R^*. The author is preparing a complete proof in an tutorial work. []

3. A fuzzy graph and cluster analysis with a fuzzy constraint

Another version of a fuzzy graph FG is defined to be a triplet FG=(V,R,A), where V and R are the same as those in BFG in section 2. The third element A is a fuzzy set of V. Thus, the set of vertices is not V but the fuzzy set A. Although BFG is a collection of $(V,C(\alpha)R)$ for $0\leq\alpha\leq 1$, another pair $(C(\alpha)A,C(\alpha)R)$ may not define a proper graph. Thereore we introduce a restriction of R on A. First, A restriction $R|_K$ of R onto a crisp set K of V is defined as $R|_K(v,w) = R(v,w)$ for $v,w\in K$. Then the restriction $R|_A$ of R onto a fuzzy set A is defined to be a collection of $R|_{C(\alpha)A}$ for $0\leq\alpha\leq 1$. In other words, for $\alpha\in[0,1]$, $C(\alpha)R|_A=(C(\alpha)R)|_{C(\alpha)A}$. Now, for any $\alpha\in[0,1]$, we define alpha-cut of FG by $C(\alpha)FG=(C(\alpha)A,C(\alpha)R|_A)$. Namely, FG is a collection of crisp graphs $\{(C(\alpha)A,C(\alpha)R|_A)\}$, $0\leq\alpha\leq 1$. Accordingly, connected components are defined on $C(\alpha)FG=(C(\alpha)A,C(\alpha)R|_A)$.

Our purpose here is to consider relation of FG to a transitive closure and a clustering algorithm. Assume that

$$A = a_1/v_1 + a_2/v_2 \cdots + a_n/v_n.$$

Now, we have the following proposition.

Prop. 1 Let

$$L = [\ R*(a\ a^T)\ ]^* \quad .$$

Assume that for an arbitrary fixed $\alpha\in[0,1]$, equivalence classes $V_1,\ldots,V_p$ generated from $C(\alpha)L$ means that a pair $v,w\in V$ belongs to the same class, say V_k, $(v,w\in V_k)$ if and only if $C(\alpha)L(v,w)=1$. Then, for any $\alpha\in[0,1]$, the partition $\{V_1,\ldots,V_p\}$ is equivalent to the connected components of $C(\alpha)FG=(C(\alpha)A, C(\alpha)R|_A)$

(Proof) Consider FG. It is clear that for arbitrarily fixed $\alpha\in[0,1]$, two vertices v_i and v_j are connected if and only if there exists a sequence of vertices $v_i,v_k,\ldots,v_q,v_j$ such that $\min[a_i,a_k,\ldots,a_p,a_j]\geq\alpha$ and $\min[R(v_i,v_k),\ldots,R(v_q,v_j)]\geq\alpha$. If we define $R'(v_i,v_j)=\min[a_i,a_j,R(v_i,v_j)]$ then the last property is equivalent to

$$\min[\ R'(v_i,v_k),\ldots,R'(v_q,v_j)\ ]\geq\alpha. \qquad (1)$$

Note that the (i,j) element of $R*(aa^T)$ is equal to $R'(v_i,v_j)$. The last relation is equivalent to

$$L(v_i,v_j) = (R')^*(v_i,v_j)\ \geq\alpha,$$

since there exists a sequence v_i, $v_k,\ldots,$ v_q, v_j that satisfies the relation (1), which means the equivalence between $\{V_1,\ldots,V_p\}$ and connected components of $C(\alpha)FG$. []

The above proposition shows a new method of hierarchical clustering. A set of objects for clustering is regarded as the set of vertices. A measure of relatedness between a pair of objects is defined in some way and the measure is regarded as the fuzzy relation. (Sometimes, a transformation from the original measure to a fuzzy relation is needed, which is dealt with in many literature in cluster analysis. See, e.g., Anderberg, [4].) In addition to the fuzzy relation, a constraint a_i is defined on each object v_i. In hierarchical clustering, clusters are merged dynamically according to a threshold parameter applied to merge levels of the measure of relatedness. Here the threshold is the alpha-cut that starts from unity and gradually decreases its value. Each object as a vertex $v_i \in V$ is not qualified as a candidate for clustering until when the parameter alpha becomes smaller than or equal to a_i. Therefore we name this method as a hierarchical clustering with a fuzzy constraint. This method is similar to the nearest neighbor method in the sense that the clusters are defined as connected components of fuzzy graphs; it differs from the nearest neighbor method in the sense that a constraint is considered on each object.

4. Wishart's Kth nearest neighbor method

A well-known technique of hierarchical agglomerative clustering, that is, the Wishart's Kth nearest neighbor method (KNN) is shown to be an instance of the above method. For showing this, we review an algorithm for the KNN clustering.

In general, a method of hierarchical agglomerative clustering consists of two stages:

I) computation of values of a measure of relatedness between all pairs of elements in the set of objects,

II) successive merges of pairs of clusters based on the measure.

In stage I), one of two kinds of the measures that have different names is used. One is called a similarity measure and the other is called a dissimilarity measure. When a pair of objects becomes more similar, then the value of a similarity measure becomes larger, while a dissimilarity measure becomes smaller. A typical example of a similarity comes from counting common features between a pair of objects, whereas a typical dissimilarity is a distance of a metric space. In general, a dissimilarity is taken for explaining the nearest neighbor method and the Wishart's KNN method. Nevertheless, we consider here a similarity to adapt the discussion to the concept of fuzzy relations and fuzzy graphs. By the same reason we also assume that the value of the similarity is in the unit interval. Namely, a similarity measure here is denoted by $S(v,w)$, $v,w \in V$ which satisfies $0 \leq S(v,w) \leq 1$, for all $v,w \in V$. Accordingly, some terminologies in clustering should be interpreted in terms of similarity. The nearest neighbor to an object v means the element w that has the maximum value of the similarity to v: $S(v,w) = \max S(v,v')$ for all $v' \in V$. K nearest neighbors to v means K different objects w_1, $w_2, \ldots,$ w_K of V that is determined as follows. Sort similarity values $S(v,w)$, for all $w \in V$ in the decreasing order of S. Then $w_1, \ldots, w_K$ are on the first K records of the sorted sequence. Clearly, w_1 is the nearest neighbor. On the other hand, w_K is named as Kth nearest neighbor to v. It is the nearest neighbor to v except $w_1, \ldots, w_{K-1}$.

We do not discuss in detail how the measure is defined, since the definition of a measure for clustering needs a different kind of consideration. Therefore we assume that the measure S is given somehow.

The following algorithm called here as W is a modified version of the Wishart's Kth nearest neighbor algorithm shown in [9], [10]. The modifications do not affect the Wishart's method in an essential way. The algorithm W simply specifies some parts that are not specified in the original algorithm. There are three kinds of the modifications. First, merge levels that are not specified in the original algorithm is shown with underlines. Second, order of merge steps is changed. That is, in the original version the step (c1) in the algorithm W is after the step (c3). Moreover the step (d) in the algorithm W is missing in the original version. Readers who refer to the original algorithm in [9], [10] will find that all these modifications are reasonable.

Algorithm W (Wishart's Kth nearest neighbor method)

a) Given an integer $K > 1$, calculate similarity measures c_i for every $v_i \in V$ to its Kth nearest neighbor.

b) Sort $\{c_i\}$ into the decreasing order. The resulting sequence is denoted by $\{c_{j_1}, c_{j_2}, \ldots, c_{j_n}\}$, $c_{j_{k-1}} \geq c_{j_k}$, $k=2,\ldots,n$. c_{j_k} corresponds to v_{j_k}.

c) Select thresholds pmin from successive c_{j_k} values. At each cycle, introduce a new "dense point" v_{j_k} and test the following (c1)-(c3). Repeat (c) until all points become dense. (Remark: Wishart called those objects that are qualified as candidates for clusters as dense points. In the terminology herein, the dense points may be called as objects that satisfy the constraint.)

c1) if there is a pair of clusters G_p and G_q such that

$$S(G_p, G_q) = \max_{\substack{v \in G_p \\ w \in G_q}} S(v,w) \geq pmin \qquad (2)$$

then merge G_p and G_q such as $G_p \cup G_q$

<u>at the level $s(G_p, G_q)$.</u>

Repeat the merges until there is no pair of clusters that satisfies (2).

c2) If there is no cluster G_i such that $\max_{w \in G_i} S(v_{j_k}, w) \geq pmin$,

then v_{j_k} generates a new cluster $\{v_{j_k}\}$ that consists of v_{j_k} alone.

c3) If there are clusters, say, $G_1, \ldots, G_s$, such that

there exists w_i which satisfies $S(v_{j_k}, w_i) \geq pmin$, $i=1,\ldots,s$,

then the clusters concerned are merged into a new cluster

$$\{v_{j_k}\} \cup G_1 \cup \ldots \cup G_s$$

<u>at the level pmin.</u>

d) When all points become dense and if there are two or more clusters, then repeat the merge process according to the nearest neighbor method.

End-of-algorithm W.

<u>Remark</u> In general, output of a hierarchical agglomerative clustering is a graphical representation of a tree called dendrogram that shows process of successive merges of the clusters. Since dendrogram itself is not necessary for the present consideration, we omit how informations of the merges are saved for the output of a dendrogram in the algorithm W. []

His concept of the dense points can be compared to the fuzzy constraint herein. Now, the Wishart's method defined by the algorithm W is shown to be a special case of the above method with a fuzzy constraint.

<u>Prop. 2</u> For arbitrarily fixed $\alpha \in [0,1]$, the clusters generated by the algorithm W is equivalent to the equivalence classes generated by

$$C(\alpha)[\ S*(c\ c^T)\ ]^*$$

where S is a matrix whose (i,j) element is $S(v_i, v_j)$ and $c=(c_1, c_2, \ldots, c_n)^T$ is defined by the quantities c_i's calculated in the step (a) of the algorithm W.

(Proof) We call the fuzzy graph FG and the algorithm W simply as FG and W, respectively, in the following. Each step in W is based on introduction of a new dense point in the decreasing order of $\{c_i\}$. This corresponds to downward change of the threshold α of alpha-cut $C(\alpha)$FG at the levels $\alpha = c_i$ beginning from $\max c_i$. Therefore we will show that at the levels c_{j_k}, k=1,...,n, clusters by W and connected components of $C(c_{j_k})$FG are equivalent. We use induction on k that proceeds with the level c_{j_k}.

First, let k=1 and consider W. First point v_j becomes dense and no clusters that contains more than one element are formed. Consider $C(c_{j_1})$FG. Then we see only one vertex v_{j_1} in $C(c_{j_1})$A: clearly, no components are observed in $C(c_{j_1})$FG.

Assume that for $k \leq m-1$, clusters by W and connected components of $C(c_{j_k})$FG are equivalent. Consider W when the m-th dense point is introduced at the level $pmin = c_{j_m}$. Consider the following three cases.

i) If there are clusters that satisfy the condition (2) in (c1), then the step (c1) is applied. Note that the clustes are merged according to the way of the nearest neighbor clustering, since the relation $S(G_p, G_q) = \max S(v,w)$, $v \in G_p$, $w \in G_q$, means the nearest neighbor linkage. Note also that these clusters are equivalent to connected components formed on $C(c_{j_{m-1}})FG$. Since each point, say v_s, in the clusters concerned satisfies $c_s \geq c_{j_{m-1}}$, a merge of G_p and G_q in (c1) agrees with connection of the same components G_p and G_q in $C(c_{j_{m-1}})FG$ that has an edge $S(v,w) \geq pmin$, $v \in G_p$, $w \in G_q$. Conversely, if there is a pair of connected components G_p and G_q in $C(c_{j_{m-1}})FG$ that should be connected in $C(c_{j_m})FG$, then an edge (v,w), $v \in G_p$, $w \in G_q$ satisfies $S(v,w) \geq c_{j_m}$. Therefore the clusters are merged in W, since the condition (2) holds. Thus, the step (c1) generates newly connected components according to the nearest neighbor linkage.

ii) For the point v_{j_m}, suppose that there is no point v_s such that $c_s \geq c_{j_m}$ and $S(v_{j_m}, v_s) \geq pmin = c_{j_m}$, then step (c2) in W is applied and a cluster with the single element $\{v_{j_m}\}$ is formed. Consider FG. Then the last condition means that v_{j_m} is not connected to any other vertices in $C(c_{j_m})FG$, and vice versa.

iii) Suppose that points $v_{s_1}, \ldots, v_{s_t}$ satisfy $c_{s_i} \geq c_{j_m}$ and $S(v_{j_m}, v_{s_i}) \geq pmin = c_{j_m}$, $i=1,\ldots,t$, and there are no other points which satisfy the last condition. Assume that v_{s_i} belongs to a cluster G'_i, $i=1,\ldots,t$. Then, the step (c3) in W is applied and we have a new cluster $\{v_{j_m}\} \cup G'_1 \cup \ldots \cup G'_t$. For FG, this condition means that in $C(c_{j_m})FG$ the vertex v_{j_m} is connected to $v_{s_1}, \ldots, v_{s_t}$. By the inductive hypothesis, v_{s_i} is in the component G'_i, $i=1,\ldots,t$. Therefore, we have a component $\{v_{j_m}\} \cup G'_1 \cup \ldots \cup G'_t$. It is easy to see the converse is also true.

Note that if these is a tie $c_{j_m}=c_{j_{m+1}}=\ldots=c_{j_{m+t}}$, we may take $c_{j_{m+t}}$ instead of c_{j_m} and introduce dense points $v_{j_m},\ldots,v_{j_{m+t}}$ at a time. Then the above argument (i), (ii), and (iii) are directly applied with little modification.

Finally, after all points become dense at the level c_{j_n}, the algorithm W proceeds in the same way as the nearest neighbor method. When $\alpha\le c_{j_n}$, $C(\alpha)FG=(C(\alpha)A,\ C(\alpha)R|_A)=(V,C(\alpha)R)$. Therefore by the equivalence between the nearest neighbor clustering and connected components of BFG, the inductive hypothesis implies the equivalence between the KNN and connected components of FG when $\alpha\le c_{j_n}$. []

Let us compare two transitive closures:

$$[\ S*(\ c\ c^T)\ ]^* \tag{3}$$

and

$$[\ I\ +\ S*(\ c\ c^T)\ ]^* \tag{4}$$

The former is equivalent to the Wishart's method, whereas the latter is equivalent to the nearest neighbor method based on the similarity

$$S'(v_i,v_j) = \min[\ c_i,\ c_j,\ S(v_i,v_j)\ ] \tag{5}$$

Although fuzzy graphs that correspond to $S*(cc^T)$ and $I+S*(cc^T)$ are different, clusters that have more than one elements are the same for the both transitive closures, since the only difference between the two is on values of the diagonal elements. Therefore we have

<u>Cor. 1</u> For arbitrary fixed $\alpha\in[0,1]$, the set of clusters that have more than one elements formed at the level α by the Wishart's

method based on the similarity $S(v_i, v_j)$ and the other set of clusters of more than one element by the nearest neighbor method based on $S'(v_i, v_j)$ given by (5) are equivalent. []

Thus, we have proved that the Wishart's method is in a sense a version of the nearest neighbor method with the modified similarity given by (5).

The last statement does not reduce the value of the Wishart's method in any sense. On the contrary, it implies that a number of theoretical properties that is valid for the nearest neighbor method holds also for the Wishart's method. One of the most important properties is the label freedom.

A technical difficulty in most of hierarchical agglomerative clustering is that the resulting dendrogram depends on the order of numbering on the objects, since merges are always carried out pairwise. A method of hierarchical agglomerative clustering is said to have the property of label freedom if the result by this method is independent of the order of the numbering, that is, if the dendrograms have the same structure when the ordering on the objects is changed. It is well-known that the nearest neighbor method has the label freedom: it is free from the dependence on the ordering. By the above corollary it follows that

<u>Property 1</u> The Wishart's KNN method has the property of the label freedom. []

<u>Remark</u> It is easily seen directly from the discussion herein that the nearest neighbor method and the Wishart's method has the label freedom, since it is evident that connected components of BFG and FG are independent of the numbering on the vertices. []

Another good property of the nearest neighbor method is that efficient algorithms for the maximal spanning trees can be used for obtaing clusters. For example, we can see that the Kruskal's algorithm [11] (in other words, the greedy algorithm) generates connected components successively by introducing edges one by one. Therefore, Corollary 1 implies the following.

Property 2 The Wishart's KNN method, or more precisely, a version of the nearest method with the modified similarity equivalent to the KNN method, is carried out by the following procedure in three steps.

A) Calculate measures c_i for all $v_i \in V$ according to the step (a) in the algorithm W.

B) For all v_i, $v_j \in V$, $i>j$, calculate

$$S'(v_i, v_j) = \min[\, c_i,\ c_j,\ S(v_i, v_j)\,]$$

C) Apply an algorithm for the maximum spanning tree. (For example, the Kruskal's algorithm to the network with the modified weight $S'(v,w)$ on the edge (v,w).) []

If the number of edges with nonzero weight $S'(v,w)$ and the number of vertices are denoted by $|E|$ and $|V|$, respectively, then the Kruskal's algorithm generates the maximal spanning tree by the amount of computation of Order($|E| \log |V|$). Moreover, since the Kruskal's algorithm uses sorting of a sequential file, it is unnecessary to keep the values of the similarity measure as an array in a random access memory. Therefore a large number of objects can be handled by the above procedure of the Wishart's KNN method, as in the case of the nearest neighbor method.

5. Conclusions

Connected components of a fuzzy graph is a fundamental concept in understanding the nearest neighbor clustering. Here we defined a new type of a fuzzy graph FG on which existence of vertices themselves is fuzzy. The fuzzy graph is equivalent to $[R*(aa^T)]^*$ which is called here a cluster analysis with a fuzzy constraint includes the well-known Wishart's KNN clustering as a special case.

In the Wishart's KNN method the term c of the fuzzy constraint is calculated as the value of similarity for the Kth nearest neighbor to each object. The above consideration shows that the concept of the fuzzy constraint is independent of the Kth nearest neighbor. Therefore, considering different ways of calculating the fuzzy constraints, we are led to different versions of the method developed herein. For example, the step (a) of the algorithm W may be replaced by another procedure:

c_i:={average value of similarity of a number K of the nearest neighbors to v_i} .

In this way, changing the step (a) by another procedure for defining c_i, we can modify the algorithm W (and the procedure in Property 2) to a more general version of the method of fuzzy constraint. Thus, we introduced here a family of new techniques that improves the nearest neighbor method by considering a simple term $(a\ a^T)$ of the fuzzy constraint.

References

1. S. Tamura, S. Higuchi, and K. Tanaka, Pattern classification based on fuzzy relations, IEEE Trans., Syst., Man, and Cybern., Vol.SMC-1, No.1, 61-66, 1971.

2. L. A. Zadeh, Similarity relations and fuzzy orderings, Information Sciences, 3, 177-200, 1971.

3. J. C. Dunn, A graph theoretic analysis of pattern classification via Tamura's fuzzy relation, IEEE Trans., Syst., Man, and Cybern., Vol.SMC-4, 310-313, 1974.

4. M. R. Anderberg, Cluster Analysis for Applications, New York, Academic Press, 1973.

5. N. Ohsumi, Fuzzy clustering and its applications, The Institute of Statistical Mathematics Tokyo, Japan, 1979.

6. J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, New York, Plenum Press, 1981.

7. K. Hirota, K. Iwama, W. Pedrycz, FCM-AD (Fuzzy Cluster Means with Additional Data) and its application to area division of aerial images, Preprint, Second IFSA Congress, Tokyo, July, 1987, 729-732.

8. S. Miyamoto and K. Nakayama, Similarity measures based on a fuzzy set model and application to hierarchical clustering, IEEE Trans., Syst., Man, and Cybern., Vol.SMC-16, No.3, 479-482, 1986.

9. D. Wishart, Mode analysis: a generalization of nearest neighbor which reduces chaining effects, in A. J. Cole, ed., Numerical Taxonomy, Proc. Colloq. in Numerical Taxonomy, Univ., St. Andrews, Sept, 1968, 283-311.

10. D. Wishart, Numerical classification method for deriving natural classes, Nature, Vol.221, Jan. 4, 97-98, 1969.

11. A. V. Aho, J. E. Hopcroft, and J. D. Ullman, The Design and Analysis of Computer Algorithms, Reading, Massachusetts, Addison Wesley, 1974.

12. A. Kandel, Fuzzy Mathematical Techniques with Applications, Reading, Massachusetts, Addison Wesley, 1986.

INSTITUTE OF INFORMATION SCIENCES AND ELECTRONICS
UNIVERSITY OF TSUKUBA
SAKURA-MURA, NIIHARI-GUN, IBARAKI 305 JAPAN

REPORT DOCUMENTATION PAGE	REPORT NUMBER ISE-TR-88-74

TITLE

A method of hierarchical cluster analysis with a fuzzy constraint and its relation to Wishart's Kth nearest neighbor method

AUTHOR(S)

Sadaaki Miyamoto

REPORT DATE	NUMBER OF PAGES
June 14, 1988	18
MAIN CATEGORY	CR CATEGORIES
Information Systems	H.3.3, G.2.2

KEY WORDS

fuzzy graphs, maximal spanning trees, hierarchical agglomerative clustering

ABSTRACT

This paper proposes a new method of hierarchical clustering as follows. Assume a set of objects (v(1),..,v(n)), a proximity measure S(i,j) between v(i) and v(j), and a fuzzy set c(1)/v(1) +...+ c(n)/v(n). The transitive closure of a relation S'(i,j) = min[S(i,j),c(i),c(j)] is considered. This method differs from the nearest neighbor method on the point that the membership c(i) which is called here a fuzzy constraint is taken into account. A well-known technique of the Wishart's Kth nearest neighbor method is proved to be a special case of the method herein. Hence it is evident that the property of the label freedom holds for the Wishart's method and an algorithm for the minimal spanning trees can be applied to the Wishart's method. Moreover, the method herein suggests many versions other than the Wishart's method for improving the nearest neighbor clustering.

SUPPLEMENTARY NOTES