Clustering Based on Fuzzy Tolerance Quotient Spaces*

Shu ZHAO$^{1,2}$, Ling ZHANG$^{1,2}$, Qianqian LIU$^{1,2}$, Yanping ZHANG$^{1,2,*}$

$^1$School of Computer Science and Technology, Anhui University, Hefei 230039, China
$^2$Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230039, China

Abstract

The quotient space theory based on fuzzy tolerance relation is put forward to solve the problem of clustering in this paper. The similarity matrix does not always satisfy ultrametric inequality, theoretically and practically. We give the method to construct the hierarchical quotient space chain if the similarity matrix is only reflexive and symmetric. We consider not only the subset of data (cluster) but also the structure of the different granular clusters. The main contributions in this paper include four parts. (1) We introduce sufficient and necessary condition that normalized distance $d(x, y)$ is an equicrural distance; (2) The relation between equicrural distance and ultrametric inequality is discussed; (3) We propose a method to get the proximate partition of a covering from a tolerance relation; (4) An example for using fuzzy tolerance quotient spaces to clustering is given.

Keywords: Clustering; Quotient Space Theory; Tolerance Relation; Hierarchical Structure

1 Introduction

Clustering techniques are mostly unsupervised methods that can be used to organize data into groups based on similarities among the individual data items. Many clustering algorithms have been introduced in many literatures and applied in various fields [1, 2, 3, 4, 5, 6]. Most of them regard the subsets of the data set as clusters; however, a few consider the structures of the clusters which might guide the analysis.

1967, Johnson developed a useful correspondence between any hierarchical system of such clusters and a particular type of distance measure [7]. He gave hierarchical clustering algorithm that similarity matrix satisfies ultrametric inequality firstly. Then he gave another two methods if the ultrametric inequality is not satisfied: Minimum Method and Maximum Method. Generally, the results of these two methods are not the same if the ultrametric inequality is not satisfied. 1971, Zadeh gave the concept of partition tree based on fuzzy similarity relation [8]. It is a

*This work is partly supported by 973 Program of China under Grant #2007CB311003, supported by National Natural Science Foundation of China under Grant #61073117 and #61175046, supported by Natural Science Foundation of Anhui Province under Grant #11040606M145.

*Corresponding author.

Email address: zhangyp2@gmail.com (Yanping ZHANG).

1553–9105 / Copyright © 2012 Binary Information Press
December 15, 2012
good idea to do clustering and discuss the structure of the clusters. In 2006, Ling Zhang and Bo Zhang investigated the clustering under the concept of granular computing, mainly based on fuzzy equivalence relation and got the structure inherent in the data themselves [9]. But the bases of Zadeh’ and Zhangs’ ideas both require that the matrix is given in advance and supposed to be reflexive, symmetric and transitive. Some other researchers introduced many concepts of transitive and discussed their properties [10, 11]. However, in reality, the similarity matrix is not always transitive. Generally, the distance function is reflexive and symmetric. That is to say, it corresponds to a tolerance relation.

Clustering based on tolerance relation has been discussed [12, 13]. Their main task is to group similar data together. We emphasize the structure among data and the hierarchical structure of different granular clusters.

In the paper, we extend the quotient space theory based on equivalence relation to on tolerance relation and discuss the hierarchical clustering. We firstly introduce the quotient space theory for clustering based on equivalence relation. The method is then put forward for clustering and obtaining the structure of different granular clusters from a given similarity matrix which is tolerant rather than equivalent. And finally the example of this method is discussed.

2 Clustering Based on Equivalence Relation

In this section, we briefly introduce the quotient space theory based on fuzzy equivalence relation and its application on clustering. In quotient space theory, a problem solving space is described by a triplet \((X, f, T)\). \(X\) denotes the problem domain, \(f(\cdot)\) indicates the attributes of domain \(X\), \(T\) is the structure of domain \(X\), i.e. the relationships among elements in \(X\).

Assume that \(X\) is a domain, \(R\) is an equivalence relation on \(X\), \([X]\) is a quotient under \(R\). Regarding \([X]\) as a new domain, we have a new world which is coarser than \(X\). We say that \(X\) is classified with respect to \(R\). So the original problem space \((X, f, T)\) is transformed into a new problem space \(((X), [f], [T])\) with a new abstract level. From these considerations, problem representations between different grain sizes correspond to different equivalence relations \(R\) or different partitions.

Here is one of some basic properties [14].

Falsity preserving property: If there is no solution in a quotient space, there is no solution in its any finer space.

From quotient space theory based on granular computing viewpoint, the objects within a cluster can be regarded as an equivalence class. Then, a clustering of objects in space \(X\) corresponds to an equivalence relationship defined on the space. In quotient space theory, it’s known that it corresponds to constructing a quotient space of \(X\).

The quotient space theory has been limited to the partition so that each element of a domain definitely belongs to one and only one equivalence class. In reality, this is not always the case. Sometimes, the boundaries between two classes are not clear-cut. Two classes may have overlapped elements or their boundaries are fuzzy, i.e. the classification is fuzzy. In order to get a crisp clustering, Zhangs [9] use fuzzy equivalence relation to transform the fuzzy set to cut set and construct a hierarchical clustering. Tang et al [15] get the transitive closure set if the matrix is not transitive. Their main ideas are similar.
Here are some definitions and propositions [14].

**Definition 1** $X$ is a domain. A fuzzy set $A$ on $X$ is defined as: $\forall x \in X$, given $\mu_A \in [0, 1]$, $\mu_A$ is called a membership of $x$ with respect to $A$. Map $\mu_A : x \rightarrow [0, 1]$ is called a membership function. Any function $\mu_A : x \rightarrow [0, 1]$ defines a fuzzy subset on $X$. If $F(x)$ is one of all subsets on $X$, $F(x)$ is a functional space consisting of all functions of $\mu_A : x \rightarrow [0, 1]$.

**Definition 2** $X \times X$ is a product space. $R \in F(X \times X)$ is called a fuzzy equivalence relation, if

(1) Reflexive: $\forall x \in X, R(x, x) = 1$

(2) Symmetric: $\forall x, y \in X, R(x, y) = R(y, x)$

(3) Transitive: $\forall x, y, z \in X, R(x, z) \geq \sup (\min (R(x, y), R(y, z)))$

**Definition 3** $\forall x \in a, y \in b, \forall a, b \in [X], d(a, b) = 1 - R(x, y)$. Then $d(\cdot, \cdot)$ is a distance function on $[X]$ and called a distance function with respect to $R$.

**Proposition 1** $R$ is a fuzzy equivalence relation on $X$. Let $R_\lambda = \{(x, y)|R(x, y) \geq \lambda\}, 0 \leq \lambda \leq 1$. $R_\lambda$ is a common equivalence relation and called as a cut relation of $R$.

**Definition 4** For a normalized distance space $(X, d)$, i.e., if any triangle formed by connecting any three points on $X$ not in a straight line is an equicrural triangle and its crus are the longest side of the triangle, the distance is called as an equicrural distance.

**Proposition 2** If $d(\cdot, \cdot)$ is a normalized distance with respect to fuzzy equivalence relation $R$, then $d(\cdot, \cdot)$ is equicrural distance.

**Theorem 1** Assume that $\{X(\lambda)|0 \leq \lambda \leq 1\}$ is a hierarchical structure on $X$. There exists a fuzzy equivalence relation $R$ on $X$ such that $X(\lambda)$ is a quotient space with respect to $R_\lambda$, where $R_\lambda$ is the cut relation of $R$, $\lambda \in [0, 1]$.

**Definition 5** $R$ denotes all equivalence relations on $X$. Define a relation “$\leq$” on $R$ as $R_1, R_2 \in R, R_2 \leq R_1 \iff x, y \in X, \text{if } xR_1y \text{, then } xR_2y$. The corresponding quotient spaces $X_{R_1}$ and $X_{R_2}$ satisfy $X_{R_1} \geq X_{R_2}$.

Since equivalence relations and partitions are equivalent, from the concept of partition, it means that all classes must mutually disjoin. This requirement is difficult to come by in reality such as clustering in data mining. It needs to extend the equivalence relation based on granular computing to more general cases. In next section, we will discuss how to use the quotient space theory based on tolerance relation to solve the problem of clustering.

### 3 Clustering Based on Tolerance Relation

#### 3.1 Relation between ultrametric inequality and normalized equicrural distance

Ultrametric inequality is mentioned when many authors discuss the transitivity [7], while in [9], Zhang gave the concept of normalized equicrural distance. In fact, these two concepts are equivalent. We will prove this proposition.
Ultrametric inequality: \( \forall x, y, z \in X, \; d(x, z) \leq \max(d(x, y), d(y, z)) \).

Proposition 3  An normalized equicrural distance is equivalent to ultrametric inequality.

Proof  The statement of ultrametric inequality exactly means that there are at least two longest edges in a triangle. That is to say it is impossible that there is only one longest edge in the triangle. Otherwise, assume \( \exists x, y, z \in X, \; d(x, z) > d(x, y), d(x, z) > d(y, z) \). But ultrametric inequality is that \( \forall x, y, z \in X, d(x, z) \leq \max(d(x, y), d(y, z)) \). It is a contradiction.

So there are at least two longest edges in a triangle. It is exactly an equicrural distance.

3.2 Sufficient and necessary conditions that is an equicrural distance

Proposition 4  The sufficient and necessary condition which \( d(x, y) \) is normalized distance function on \( X \) is that it is a hierarchical quotient chain \( X = X_1 \geq X_2 \ldots \geq X_n \) constructed by quotient spaces which are corresponding cut-set \( R_\lambda \) of fuzzy relation \( R(x, y) = 1 - d(x, y) \), \( 0 \leq \lambda \leq 1 \).

Note: Hierarchical quotient chain means it must satisfy some basic properties, such as falsity preserving property, etc. That is to say, if \( xR_{\lambda_1} y \), then \( xR_{\lambda_2} y \), for \( \lambda_1 \geq \lambda_2 \) Specifically, the necessary condition is that: Given \( d(x, y) \) is a normalized distance function on \( X \), let \( R_\lambda \) is the cut-set of fuzzy relation \( R(x, y) = 1 - d(x, y) \). Define \( R_\lambda = \{ R(x, y) \geq \lambda | x, y \in X \} \), \( 0 \leq \lambda \leq 1 \). Then all the quotient spaces corresponding to different \( R_\lambda \) are a hierarchical quotient space chain, in which quotient space \( X_1 \) corresponding to \( R_1 \) is the finest and \( X_n \) corresponding to \( R_0 \) is the coarsest.

The sufficient condition is that:

Given a hierarchical quotient space chain \( X_1 \geq X_2 \ldots \geq X_n \), take real numbers \( 1 = a_1 > a_2 > \ldots > a_n \geq 0 \), let

\[
R(x, y) = \begin{cases} 
1, & x = y \\
 a_i, & (a, b) \in R_{X_i} \text{ while } (a, b) \notin R_{X_{i-1}} 
\end{cases}
\]

Then \( d(x, y) = 1 - R(x, y) \) must be a normalized distance function on \( X \).

Note: \( R(x, y) \) is a fuzzy equivalent relation.

3.3 Get proximate partition from covering

If \( d(x, y) \) is not an equicrural distance, that is to say, \( R(x, y)(1 - d(x, y)) \) is a tolerance relation rather than an equivalence relation, only a covering rather than a partition is obtained by the cut relation \( R_\lambda \). What we will solve is how to get the approximate partition from the covering and to make the different level partitions become hierarchical quotient space chain.

Given a tolerance relation \( R(x, y) \) on \( X \) and get different cut sets \( R_\lambda, 0 \leq \lambda \leq 1 \). Suppose each level covering is \( C_1, C_2, \ldots, C_n \), respectively, corresponding to different \( R_\lambda, 0 \leq \lambda \leq 1 \). Our aim is to obtain the partitions \( X_1, X_2, \ldots, X_n \), in which \( X_1 \) is the finest and \( X_n \) is the coarsest, i.e. \( X_1 > X_2 > \ldots > X_n \).

Main frame of the algorithm:
Step 1: Get the most approximate partition $X_i$ corresponding to the covering $C_i$ (initially, $i = 1$), the specific algorithm is as following.

Step 2: If $i + 1 > n$ or $X_i = X$, stop; or, let $max\{C_{i+1}, X_i\} \rightarrow C_{i+1}$, $i + 1 \rightarrow i$, return Step 1.

Note: $max\{C_{i+1}, X_i\} \rightarrow C_{i+1}$ is to ensure the falsity preserving property. Because our final goal is to construct a hierarchical quotient space chain, we must ensure that the elements belong to the same equivalent class in a coarser level if they are equivalent in a finer level.

We will discuss the details of Step 1. According to the changing value, we decompose or combine the intersected coverings until the covering becomes a partition. The specific steps are as follows.

Suppose the $i^{th}$ level covering is $C_i = C^1_i, C^2_i, \ldots, C^k_i$.

If $\forall p, q \in 1, 2, \ldots, k, C^p_i \cap C^q_i = \emptyset$, stop. Now the partition is obtained.

If $\exists p, q \in 1, 2, \ldots, k$, make $C^p_i \cap C^q_i = \{b_1, b_2, \ldots, b_t\}, b_j \in X, j \in \{1, 2, \ldots, t\}$, $1 \leq t \leq min(|C^p_i|, |C^q_i|)$, let $m = |C^p_i/(C^p_i \cap C^q_i)|, n = |C^q_i/(C^p_i \cap C^q_i)|$.

(1) If $t \geq m$ and $t \geq n$, combine $C^p_i$ and $C^q_i$, i.e. $C^p_i \cup C^q_i$.

(2) If $t < m$ or $t < n$, decompose the minimum covering of $C^p_i$ and $C^q_i$. Suppose $m \leq n$, then decompose the $C^p_i$ into $C^p_i/(C^p_i \cap C^q_i)$ and $C^q_i \cap C^q_i$.

4 Examples

4.1 Using fuzzy tolerance quotient spaces to clustering

Take IRIS sample from UCI as an example. In order to explain the whole process and compare with other methods, we choose six samples, only first two of each class, from IRIS.

Calculate their normalized Euclid distances and get a reflective, symmetric matrix.

The corresponding tolerance relation matrix is as Table 1.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>1</td>
<td>0.9461</td>
<td>0.5996</td>
<td>0.6383</td>
<td>0.4715</td>
<td>0.5792</td>
</tr>
<tr>
<td>(2)</td>
<td>0.9461</td>
<td>1</td>
<td>0.5996</td>
<td>0.6383</td>
<td>0.4715</td>
<td>0.5792</td>
</tr>
<tr>
<td>(3)</td>
<td>0.9461</td>
<td>0.5996</td>
<td>1</td>
<td>0.6383</td>
<td>0.4715</td>
<td>0.5792</td>
</tr>
<tr>
<td>(4)</td>
<td>0.6383</td>
<td>0.6314</td>
<td>0.9360</td>
<td>1</td>
<td>0.8192</td>
<td>0.8937</td>
</tr>
<tr>
<td>(5)</td>
<td>0.4715</td>
<td>0.4661</td>
<td>0.8156</td>
<td>0.8192</td>
<td>1</td>
<td>0.8666</td>
</tr>
<tr>
<td>(6)</td>
<td>0.5792</td>
<td>0.5819</td>
<td>0.8551</td>
<td>0.8937</td>
<td>0.8666</td>
<td>1</td>
</tr>
</tbody>
</table>

Take $\lambda = 1, 0.9461, 0.9360, 0.8937, 0.8666, 0.8551, 0.8192, 0.8156, 0.6383, 0.6314, 0.5996, 0.5904, 0.5819, 0.5792, 0.4742, 0.4661, i = 1, 2, \ldots, 16$, respectively and get the corresponding cut relation, some of the cut relation matrices are listed as Figure 1.

From the matrices, we get the hierarchical clustering step by step.

(1) $C_1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$, $C_1$ is a partition. So $X_1 = C_1$.

(2) $C_2 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$, $max\{X_1, C_2\} = C_2$, $C_2$ is a partition. So $X_2 = C_2$.

(3) $C_3 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$, $max\{X_2, C_3\} = C_3$, $C_3$ is a partition. So $X_3 = C_3$.

(4) $C_4 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$, $max\{X_3, C_4\} = C_4$, $\{3\} \cap \{4\} = \{4\}$, $t = m = n = 1$, so merge $\{3\}$ and $\{4\}$, and get $X_4 = \{\{1\}, \{2\}, \{3\}, \{4\}\}$.
(5) $C_5 = \{\{1,2\}, \{3,4\}, \{4,6\}, \{5,6\}\}$, $\max\{X_4, C_5\} = \{\{1,2\}, \{3,4,6\}, \{5,6\}\}$, $\{3,4,6\} \cap \{5,6\} = \{6\}$, $t = 1, m = 2, n = 1, t < m$, so decompose $\{5,6\}$ into $\{5\}$ and $\{6\}$ and get $X_5 = \{\{1,2\}, \{3,4,6\}, \{5\}\}$.

According to the algorithm, in the same way, we get $X_6 = \{\{1,2\}, \{3,4,6\}, \{5\}\}$, $X_7 = \{\{1,2\}, \{3,4,5,6\}\}$, $X_8 = \{\{1,2\}, \{3,4,5,6\}\}$, $X_9 = \{\{1,2\}, \{3,4,5,6\}\}$, $X_{10} = \{\{1,2\}, \{3,4,5,6\}\}$, $X_{11} = \{\{1,2,3,4,5,6\}\}$, respectively. The whole process is completed. The quotient space chain is that $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\} > \{\{1,2\}, \{3\}, \{4\}, \{5\}, \{6\}\} > \{\{1,2\}, \{3,4\}, \{5\}, \{6\}\} > \{\{1,2\}, \{3,4,6\}, \{5\}\} > \{\{1,2\}, \{3,4,5,6\}\} > \{\{1,2\}, \{3,4,5,6\}\}$.

4.2 Other methods

Take single-link and complete-link methods, respectively. In complete-link (or maximum method in [5]) hierarchical clustering, we merge in each step the two clusters whose merger has the smallest diameter (or the two clusters with the smallest maximum pairwise distance). In single-link (or minimum method in [5]) hierarchical clustering, we merge in each step the two clusters whose two closest members have the smallest distance (or the two clusters with the smallest minimum pairwise distance).

1) Complete-link method

Distance matrix after the first clustering is as Table 2.

Table 2: Distance matrix after first clustering in Complete-link method

<table>
<thead>
<tr>
<th></th>
<th>(1,2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2)</td>
<td>0</td>
<td>0.4096</td>
<td>0.3686</td>
<td>0.5339</td>
<td>0.4208</td>
</tr>
<tr>
<td>(3)</td>
<td>0</td>
<td>0.4096</td>
<td>0.3686</td>
<td>0.5339</td>
<td>0.4208</td>
</tr>
<tr>
<td>(4)</td>
<td>0.3686</td>
<td>0.0640</td>
<td>0</td>
<td>0.1808</td>
<td>0.1063</td>
</tr>
<tr>
<td>(5)</td>
<td>0.5339</td>
<td>0.1844</td>
<td>0.1808</td>
<td>0</td>
<td>0.1334</td>
</tr>
<tr>
<td>(6)</td>
<td>0.4208</td>
<td>0.1449</td>
<td>0.1063</td>
<td>0.1334</td>
<td>0</td>
</tr>
</tbody>
</table>

In the same way, we get $\{\{1,2\}, \{3,4\}, \{5\}, \{6\}\}$, $\{\{1,2\}, \{3,4\}, \{5,6\}\}$, $\{\{1,2\}, \{3,4,5,6\}\}$, $\{\{1,2,3,4,5,6\}\}$, step by step.

2) Single-link method

Distance matrix after the first clustering is as Table 3.

In the same way, we get $\{\{1,2\}, \{3,4\}, \{5\}, \{6\}\}$, $\{\{1,2\}, \{3,4,6\}, \{5\}\}$, $\{\{1,2\}, \{3,4,5,6\}\}$, $\{\{1,2,3,4,5,6\}\}$, step by step.

Because the similarity matrix does not satisfy ultrametric inequality, generally the results of
Table 3: Distance matrix after first clustering in Single-link method

<table>
<thead>
<tr>
<th></th>
<th>(1,2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,2)</td>
<td>0</td>
<td>0.4004</td>
<td>0.3617</td>
<td>0.5285</td>
<td>0.4181</td>
</tr>
<tr>
<td>(3)</td>
<td>0.4004</td>
<td>0</td>
<td>0.0640</td>
<td>0.1844</td>
<td>0.1449</td>
</tr>
<tr>
<td>(4)</td>
<td>0.3617</td>
<td>0.0640</td>
<td>0</td>
<td>0.1808</td>
<td>0.1063</td>
</tr>
<tr>
<td>(5)</td>
<td>0.5285</td>
<td>0.1844</td>
<td>0.1808</td>
<td>0</td>
<td>0.1334</td>
</tr>
<tr>
<td>(6)</td>
<td>0.4181</td>
<td>0.1449</td>
<td>0.1063</td>
<td>0.1334</td>
<td>0</td>
</tr>
</tbody>
</table>

4.3 Discussion

In [8], Zadeh got the partition based on the matrix which is transitive, i.e., it satisfies the ultrametric inequality. In [7], the author clarified if the similarity matrix satisfies the ultrametric inequality, he could get an unambiguous manner with hierarchical clustering method in which the pair of points or clusters is merged each step. If the matrix dissatisfies the ultrametric inequality, they had two methods: Minimum Method (refers to single-link method here) and Maximum Method (refers to complete-link method here), but he pointed out that “It is tacitly assumed in the discussion of the methods that the distances in the original matrix are all distinct except for 0. It is not important in the Minimum Method, but difficulties do arise when applying the Maximum Method to matrices with large numbers of identical entries”. There is no this restriction for our method. Merging and Decomposition are the main ideas in any hierarchical clustering methods. They merge only two points (clusters) into one cluster or decompose one cluster into two clusters (points) each step. Finally, a hierarchical structure is obtained. The hierarchical structure, i.e., hierarchical quotient space chain, is also our goal. In our method, we get the binary matrix from the distance matrix or the relation matrix based on the value of cut relation. If the distance matrix satisfies equicrural distance, i.e., ultrametric inequality, we get the partition from the binary matrix directly; otherwise, we get the approximate partition of the covering. At each step, more points or clusters are merged together rather only two points or clusters are merged if they have the same distances in the matrix.

We give the method to get the approximate partition of the covering in the present paper. However, it is a local solution rather than a global one. The base of our method is calculating the changing numbers of the overlapping sets. We do merging if the changing number is smaller by merging than decomposition, vice versa.

Possible extension: we would take the value of cut set into consideration when we make decisions of merging or decomposition.

5 Conclusion

The requirement of transitivity for the similarity matrix is not always satisfied. We discuss quotient space theory based on tolerance relation to clustering in present paper. Our method not only puts similar data together but also gets the structure of different granular clusters. For any given distance matrix, transform it to a relation matrix firstly. If it is an equivalence relation, get the partition directly and construct hierarchical quotient space chain by defining the cut relation. If it is a tolerance relation, use our method to get the approximate partition from the covering and then construct hierarchical clustering. We can obtain the approximate partition for each level.
covering corresponding to the tolerance relation. Finally, we get not only different level partitions but also the structure of the original data.

References