A New Method for Fuzzy Clustering : Vague Clustering

Fang Rui Zhao yi biao
School of Computer and Information Technology, Beijing jiaotong University, Beijing 100044, China

Abstract

Fuzzy clustering techniques have been wildly applied effectively in image processing, pattern recognition, data mining and fuzzy modeling. Based on the Fuzzy set theory, fuzzy clustering techniques has its shortages: fuzzy membership function is a single value which combines the evidence for an object belonging to a cluster and against an object belonging to a cluster, without indicating how much there is of which , and this cannot always effectively get the objective result of clustering. Vague sets is characterized by a truth-membership function and a false-membership function which can solve the problem. In this paper ,the author proposed a new method for clustering Vague clustering method.This new method can characterize both the similarity and the dissimilarity between pairs of objects and can get an objective result for clustering and recognition shown in the example.

Keywords: vague set, fuzzy set, fuzzy clustering , vague clustering, pattern recognition

1. Introduction

The aim of cluster analysis is to group a set of objects into clusters such that objects within the same cluster have a high degree of similarity, while objects belonging to different clusters have a high degree of dissimilarity. Based on the fuzzy set theory, in fuzzy clustering algorithm,objects are represented by fuzzy sets seen in literature [1] [2] [3][4] [5] [6][7][8]and can only consider the similarity or dissimilarity between pairs of objects rather than both of them simultaneously, and this cannot obtain more reasonable description of objects and thus a good result of clustering cannot be acquired. The vague set theory introduced in[9]is a new extended form for fuzzy set: it adopts truth and false membership function to express for and against evidence. It overcomes the disadvantage of membership function in fuzzy set which cannot characterize both the similarity and dissimilarity between pairs of objects. Therefore, it has been conceived as a new efficient tool to deal with ambiguous data and applied successfully in different fields[9, 10, 11, 12].Based on the vague set theory, In this paper the author proposed vague clustering algorithm Based on the vague set theory. Our paper is organized as follows: In section 2, the relationship between fuzzy sets , vague sets intuitionistic fuzzy sets ,and interval-valued fuzzy sets are briefly analyzed . In section 3 , vague clustering algorithm based on vague set is presented and an example is shown in section 4.finally ,section 5 is conclusion.

2.1 Fuzzy Sets

A fuzzy set A of the universe of discourse $U = \{x_1, x_2, \ldots, x_n\}$ can be represented by $A=\mu_A(u_1)/u_1+\mu_A(u_2)/u_2+\ldots+\mu_A(u_n)/u_n$ where μ_A is the membership function of the fuzzy set $\mu_A : U \to [0,1]$ $\mu_A(u_i)$ is a single value and it indicates the grade of membership of u_i in the fuzzy set A.

2.2 Vague Sets

Let U be the universe of discourse, $U = \{x_1, x_2, \ldots x_n\}$ and let, t_v and, f_v be the truth-membership function and false-membership function of the Vague set V. $t_v(x)$ is a lower bound on the grade of membership of x derived from the evidence for x, and $f_v(x)$ is a lower bound on the negation of x derived from the evidence against x. $t_v(x)$ and $f_v(x)$ both associate a real number in the interval [0,1]with each x in U,where $t_v(x)+f_v(x)\le 1$. This approach bounds the grade of membership of x to a subinterval[$t_v(x)$, 1- $f_v(x)$] of [0, 1] A vague set in the universe of discourse U is illustrated in Figure.1

The precision of our knowledge about x is immediately clear, with our uncertainty characterized by the difference $1- t_v(x) - f_v(x)$ If this is small, our knowledge about x is relatively precise;If it is large, we know correspondingly little; if 1- $f_v(x)$ is equal to $t_v(x)$,our knowledge about x is exact, and the theory reverts back to that of fuzzy sets; If 1- $f_v(x)$ and $t_v(x)$

are both equal to 1 or 0,our knowledge about x is very exact and the theory revert back to that of ordinary set.

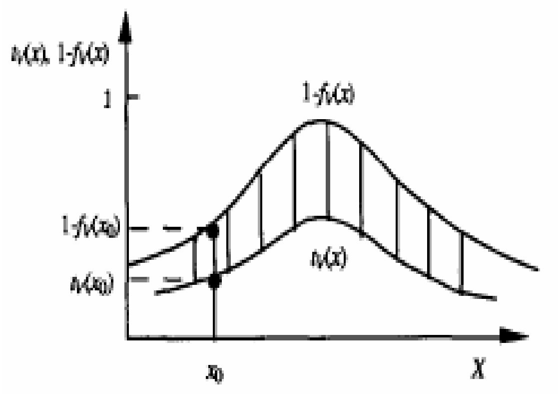

Fig. 1

2.3 Vague Sets, intuitionistic fuzzy sets and interval-valued fuzzy sets

Let U be the universe of discourse, k.Atanassov introduced intuitionistic fuzzy sets[22] :

$A^{*}=\{<u,\mu_A(u),v_A(u)>|u\in U\}$, where $\mu_A : U \rightarrow [0,1]$ $v_A : U \rightarrow [0,1]$, $0\le \mu_A(u)+v_A(u)\le 1$, denote that $A^{*}=(\mu_A,v_A)$.

In a lot of cases, no objective procedure is available to select the crisp membership degree of every element in a fuzzy set, so in [23],zadeh suggests the concept of interval-valued fuzzy sets, where the degree of membership of an element to a set is represented by a closed subinterval of [0,1]:

$\overline{A}(u)=[A^{-}(u),A^{+}(u)]$,

where $A^{-}:U\rightarrow[0,1]$, $A^{+}:U\rightarrow[0,1]$ and $A^{-}(u)\le A^{+}(u)$, $\forall u\in U$,denote that $\overline{A}=[A^{-},A^{+}]$.

Intuitionistic fuzzy sets and interval-valued are two intuitively straightforward extensions of fuzzy sets.We find that intuitionistic fuzzy sets are the equivalent form of interval-valued fuzzy sets (let $A^{-}=\mu_A, A^{+}=1-v_A$, $A^{*}=(\mu_A,v_A)$ can induce $\overline{A}(u)=[A^{-}(u),A^{+}(u)]$) . H.Bustince and P.Burillo pointed out in [10]that vague sets are intuitionistic fuzzy sets.

From the discussion above, we know clearly the relation between different set theory.we can see that the grade of membership of an element $u_i\in U$ in a fuzzy set is represented a by single value which combines the evidence for $u_i\in U$ and the evidence against $u_i\in U$, and without indicating how much there is of each. Furthermore, it tells us nothing about its accuracy. Thus, in[9],Gau and Buehrer presented the concepts of vague sets that can solve the problem effectively. vague set contains the information of similarity and dissimilarity, thus , the author proposed a new method for clustering based on vague set vague clustering.

3 Vague clustering based on Vague set

In this section, the author discussed in details the proposed new method of vague clustering. Let $\{X_1,X_2,...X_n\}$ be the set of N object to be classified , and each object has N features and is represented by a p-dimensional feature vector $X_k=\{u_{1k},u_{2k},...u_{Nk}\}^T$ Thus , a set of N features in the object set can be represented as a $p\times N$ feature matrix:

$$X=\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1N} \\ u_{21} & u_{22} & \cdots & u_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ u_{p1} & u_{p2} & \cdots & u_{pN} \end{bmatrix}$$

Each object can be represented by a vague set as follows:

$$X_i=[t_{xi}(u_1),1-f_{xi}(u_1)]/u_1+\ldots$$
$$+[t_{xi}(u_p),1-f_{xi}(u_p)]/u_p$$
$$X_j=[t_{xj}(u_1),1-f_{xj}(u_1)]/u_1+\ldots$$
$$+[t_{xj}(u_p),1-f_{xj}(u_p)]/u_p$$

vague set contains the information of similarity and dissimilarity. Let' s Assume that $V_{xi}(u_k)=[t_{xi}(u_k),1-f_{xi}(u_k)]$ stand for the membership degree of u_k in X_i $V_{xj}(u_k)=[t_{xj}(u_k),1-f_{xj}(u_k)]$ stand for the membership degree of u_k in X_j , so the scores of X_i, and X_j are

$$S(V_{xi}(u_k))=t_{xi}(u_k)-f_{xi}(u_k)$$
$$S(V_{xj}(u_k))=t_{xj}(u_k)-f_{xj}(u_k)$$

Where k=1,2,3,…n. Assume that the weight of feature u_p in an object is ω_k $\omega_k\in(0,1]$,the weighted similarity measure between the two vague objects X_i and X_j can be evaluated by the function[15][16]

$$T'(X_i,X_j)=\sum_{k=1}^{p}\omega_k M'(V_{Xi}(u_k),V_{Xj}(u_k))/\sum_{i=1}^{n}\omega_i=$$
$$\sum_{k=1}^{p}\omega_k|1-[S(V_{Xi}(u_k)-S(V_{Xj}(u_k))]/4-$$
$$[|t_{xi}(u_k)-t_{xj}(u_k)|+|f_{xi}(u_k)-f_{xj}(u_k)|]/4|/\sum_{k=1}^{p}\omega_k$$

we can obtain the similarity degree between pairs of objects and represent them by vague similarity

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

where $r_{ij}=T'(X_i,X_j)$ Vague similarity matrix contains the information of similarity and dissimilarity between pairs of objects. But R here is not transferable, it cannot be used for clustering, it is similar to the case in fuzzy relation matrix that there must be a natural number k, and $R'=R\circ R\circ\cdots\circ R$ is not changed with k. Then we indicate $R^*=R^K$ as the vague equivalent matrix,and can be used for clustering. after getting R* , select different $\lambda(0\le\lambda\le1)$ as the level of cutting distance and can obtain different classes of clustering.

$$T_{ij}=\begin{cases}1 & t_{ij}\ge\lambda \\ 0 & t_{ij}<\lambda\end{cases}$$ Where tij is the element in the matrix

R* There are some optimized method for transitive closure proposed in literature [18][19][20][21] and here we choose the traditional method to discuss the Problem for briefly.

4 Example

In this section, the proposed method is applied to classify different objects.

Let $A=\{A_1,A_2,...A_5\}$ be the set of vague objects.

Let $u=\{u_1,u_2,...u_5\}$ be the set of features of every pattern, these five patterns can be expressed as follows:

$A_1=[0.4,0.4]/u_1+[0.3,0.7]/u_2+[0.5,0.7]/u_3+[0.7,0.9]/u_4+[0.8,1.0]/u_5$

$A_2=[0.3,0.5]/u_1+[0.4,0.6]/u_2+[0.4,0.8]/u_3+[0.7,0.9]/u_4+[0.9,0.9]/u_5$

$A_3=[0.6,0.8]/u_1+[0.5,0.6]/u_2+[0.7,0.8]/u_3+[0.3,0.4]/u_4+[0.8,0.9]/u_5$

$A_4=[0.5,0.7]/u_1+[0.7,0.8]/u_2+[0.1,0.4]/u_3+[0.2,0.3]/u_4+[0.6,0.7]/u_5$

$A_5=[0.7,0.7]/u_1+[0.2,0.6]/u_2+[0.8,0.9]/u_3+[0.4,0.6]/u_4+[0.5,0.8]/u_5$

Their vague relation matrix is represented as follows:

$$\begin{bmatrix} [0.4,0.4] & [0.3,0.7] & [0.5,0.7] & [0.7,0.9] & [0.8,1.0] \\ [0.3,0.5] & [0.4,0.6] & [0.4,0.8] & [0.7,0.9] & [0.9,0.9] \\ [0.6,0.8] & [0.5,0.6] & [0.7,0.8] & [0.3,0.4] & [0.8,0.9] \\ [0.5,0.7] & [0.7,0.8] & [0.1,0.4] & [0.2,0.3] & [0.6,0.7] \\ [0.7,0.7] & [0.2,0.6] & [0.8,0.9] & [0.4,0.6] & [0.5,0.8] \end{bmatrix}$$

the weights of $u_1,u_2,...u_5$ are chosen as: {0.2,0.4,0.8,0.6,1.0} according to the formula of similarity measure for two vague sets[15, 16]we can get :

$T'(A_1,A_2)=0.96$ $T'(A_1,A_3)=0.82$

$T'(A_1,A_4)=0.67$ $T'(A_1,A_5)=0.68$

$T'(A_2,A_3)=0.82$ $T'(A_2,A_4)=0.76$

$T'(A_2,A_5)=0.75$ $T'(A_3,A_4)=0.75$

$T'(A_3,A_5)=0.85$ $T'(A_4,A_5)=0.72$

their vague similarity relation matrix is represented by following matrix:

$$R=\begin{bmatrix} 1 & 0.96 & 0.82 & 0.67 & 0.68 \\ 0.96 & 1 & 0.82 & 0.76 & 0.75 \\ 0.82 & 0.82 & 1 & 0.75 & 0.85 \\ 0.67 & 0.76 & 0.75 & 1 & 0.72 \\ 0.68 & 0.75 & 0.85 & 0.72 & 1 \end{bmatrix}$$

According to the discussion above, we calculate the value of k satisfying $R^*=R=R^3$ and acquire vague equivalent relation matrix:

$$R^*=\begin{bmatrix} 1 & 0.96 & 0.82 & 0.76 & 0.82 \\ 0.96 & 1 & 0.82 & 0.76 & 0.82 \\ 0.82 & 0.82 & 1 & 0.76 & 0.85 \\ 0.76 & 0.76 & 0.76 & 1 & 0.76 \\ 0.82 & 0.82 & 0.85 & 0.76 & 1 \end{bmatrix}$$

by selecting different $\lambda(0\le\lambda\le1)$ as the level of cutting distance we can obtain different classes:

let $\lambda=0.96$ we get: $\{A_1, A_2\},\{A_3\}, \{A_4\},\{A_5\}$

Iet $\lambda=0.82$ we get: $\{A_1, A_2, A_3, A_5\}, \{A_4\}$

let $\lambda=0.76$ we get: $\{A_1, A_2, A_3, A_4, A_5\}$

The method of vague clustering contains complete information of every object and hence is more flexible for classification and pattern recognition.

5. Conclusion

This paper presented a new method of clustering using the vague set theory. The proposed new method consider the aspect of both the similarity and dissimilarity between pairs of objects and thus can characterize both the favoring and the opposing evidence for every object belonging to a cluster and can get an objective result for pattern recognition and classification.

In section 2.3, we analyze the relationship between vague sets, intuitionistic fuzzy set and interval-valued set. if the true and false membership degree of an element in a set are given as the closed subinterval[0,1],the proposed algorithm of vague clustering can easily be developed into vague interval-valued clustering algorithm.

6. References

[1] J.Bezdek, pattern Recognition With Fuzzy Objective Function Algrithms. New York: Plenum, 1981.

[2] L.T.Koczy and K.Hirota, Ordering. distance and closeness of fuzzy sets, Fuzzy Sets and Systems. 2002,10(6) 713-727.

[3] Dave R N, Sen S. Robust fuzzy clustering of relational data. IEEE Trans on Fuzzy Systems. 2002,10(6) 713-727.

[4] Corsini P; Lazzerini B; Marcelloni F. A Fuzzy Relational ClusteringAlgorithm Based on a Dissimiarity Measure Extracted From Data. IEEE Transactions on Systems, Man, and Cybernetics PartB,2004,(3) 41775-781.

[5] Sato M;Sato Y. Fuzzy Clustering Model for Fuzzy Data. Proceedings of 1995 IEEE International Conference on Fuzzy Systems. 1995(4) 2123-2128.

[6] E.Diday, The Symbolic Approach in Clustering, Classificatio n and Related Methods of Data Analysis, H.H.Bock, Ed.Amsterdam, The Netherlands :Elsevier,1988.

[7] K.C.Gowda and E.Diday, "Symbolic clustering using a new dissimilarity" Pattern Recognition., vol.24,no.6, pp. 567-578,1991.

[8] K.C.Gowda and E.Diday, "Symbolic clustering using a new similarity measure, Trans. Syst .Man,Cybern. vol.22, pp. 368-378Feb.1992

[9] Gau W L, Buehrer D J.Vague sets[j].IEEE Transactions on Systems,Man,and Cybernetics, 1993,23(2) 610-614.

[10] Bustince H, Burillo P. Vague sets are intuitionistic fuzzy sets. Fuzzy Sets and Systems 1996, 79(l) 403-405.

[11] Hong D H,Choi C H. Multicriteria fuzzy decision making problems based on Vague set theory.Fuzzy Set and Systm, 2000,114:103-113.

[12] Chen,S.M. Fuzzy system reliability analysis based on vague set theory.1997IEEE International Conference on Computational Cybernetics and Simulation. 1997,2-1650-1655.

[13] Zadeh ,L .A.(1965) Fuzzy Sets, Inf. Control. 8, 338-353.

[14] He ZhongXiong. Weight Analysis System. [J]journal of Systems Engineering and Electronics 1992(4) 41-44.(in Chinese).

[15] Chen, S. M. Similarity Measures Between Vague Sets and Between Elements. IEEE Trans. Systems Man Cybernetics-Part B, 1997, 27: 153-158.

[16] Li Fan, Xu Zhang-yan. Measures of Similarity between Vague Sets [j]. journal of Software, 2001,12(6) 921 -927. (in Chinese)

[17] Li Fan.Fuzzy Information Processing Systems. Beijing: Beijing University Press, 1998 (in Chinese).

[18] Fu Guo yao. Optimization Methods for Fuzzy Clustering. Fuzzy Sets and System 1998(93) 301-309.

[19] Lee Hsuan-Shih. An Optimal Algorithm for Computing the Max-min Transitive Closure ofa Fuzzy Similarity Matrix. Fuzzy Sets and Systems , 2001(123) 129-136

[20] Chepoi V, Dumitrescu D. Fuzzy Clustering with Structural Constains. Fuzzy Sets and System, 1999(105) 91-97

[21] Josien K, Liao T W. Simultaneous Grouping of Parts and Machines with an Integrated Fuzzy Clustering Method Fuzzy Sets and Systems, 2002 (126), 1-21.

[22] K.Atanassov. Intuitionistic Fuzzy set[J].Fuzzy Set and Systems,1986,20(2);87-96

[23] L.A.Zadeh, The concept of a linguistic variable and its application to approximate reasoning-I.Inform.Sci. (1975)199-249.