Research on the Application Models of Fuzzy Set-Valued Statistics.

Guo Sizhong and Shi Suying
Mathematics Section, Fuxin Mining Institute,
Fuxin, Liaoning, P.R.China.

Abstract The concept of set-valued statistics was given first by Wang Peizhuang in paper [1]. Also some applications about set-valued statistics were discussed in many other papers. In paper [4], we expended the concepts of random set and its projected function, proposed the method of fuzzy set-valued statistics. In this paper, Some applied models about fuzzy set-valued statistics were discussed and advanced some useful methods for engineering practice.

Kywords: Fuzzy random set, Projected function, Set-valued statistics; Fuzzy clustering, Decision weight, Fuzzy forecast.

1. Introduction

The theoretical basis of classical statistics is probability theory, the method of classical statistics can be called as statistics by points. But, in many engineering practices, for example, when deal with psychological measure, the observational samples are not usually some separate points. Because the concepts formed in the process of mankind understanding natural world are all a division for natural objects. The method of division is based on a general character that exists in some complicacyes. In some complicated situations, the exact diffrence between separate things is very difficult to hold. But the diffrence in several kinds of things is easy to hold. For example, there are some people's photos, take two randomly and compare them. How much they are similar? This is not easy to answer. But they are divided into several groups easily according their more or less same, because the former investigation is isolation, the latter as to consider as a whole. The concept of viewing the situation as a whole is just a important norm in the activities of dividing, choosing ordering or making decision etc., in the natural world. To consider separate points only and not to pay attention to the whole relation of points and points is just one of main reason of unstabilized for that using classical statistics make psychological measurement. The set-valued statistics cast away the method of classical by points statistics, result of each time is a subset on the universe of discourse X. The method of set-valued statistics as an extension of the classical probabilistic statistics and has extended the applicable range of the statistical experment. Furthemore, the system of humanity, such as society, economics has suggested a new numercal method of management.

2. Set-valued statistics method apply to fuzzy clustering

S.Tamura et al. [6] described a hierarchical clustering scheme based on a fuzzy equivalence relation on a patterns set X. Let $X=\{x_1,x_2,\ldots,x_m\}$ be a finite set, this method request to given a real function r on $X\times X$, which satisfies: (1). $r(x,x)=1$, (2). $r(x,y)=r(y,x)$. $\forall x_i,y_j\in X$, denote $r_{ij}\triangleq r(x_i,y_j)$, thus $R=[r_{ij}]$ is called fuzzy similar relation matrix. If it also satisfies the condition (3), $r_{ij}=\bigvee_{k=1}^{m}(r_{ik}\wedge r_{kj})$, $i,j=1,2,\ldots,m$, then R is called a fuzzy equivalence relation matrix. The paper [6] given a method computing transitive closure R^* for a fuzzy similar matrix R, and get system classification with R^*.

In this paper, we prpose some methods of constructing fuzzy similar matrix using set-valued statistics on sample set X.

Let $X=\{x_1,x_2,\ldots,x_m\}$ be sample set , number of peoples who attend test is n.

I. **Hard partition method**: suppose that $\{B_{i1},B_{i2},\ldots,B_{ik_i}\}$ is a hard partition to X by ith people, where $\bigcup_{j=1}^{k_i}B_{ij}=X$ and $B_{ij}\cap B_{il}=\Phi$ when $l\neq j$, $i=1,2,\ldots,n$.

For $x\in X$, denote that

$$\dot{x}\triangleq\{B_{ij}\mid x\in B_{ij},\ j=1,2,\ldots,k_i,\ i=1,2,\ldots,n\} \tag{1}$$

it may be seen observational sample of a random set ξ, projected valued estimate as

$$\mu_x(y)=\frac{1}{n}\sum_{B\in \dot{x}}\chi_B(y). \tag{2}$$

Where χ_B is the characteristic function of set B. $r_{ij}\triangleq\mu_{x_i}(y_j)$ expressed possible degree that sample x_i and x_j are in corporate into same group. By formula (2), we easily see that $r_{ii}=1$. Let $|A|$ be number of elements of set B, then

$$\mu_x(y)=\frac{1}{n}\sum_{B\in \dot{x}}\chi_B(y)=\frac{1}{n}\sum_{B\in \dot{x}\cap\dot{y}}\chi_B(y).$$

$$=\frac{1}{n}|\dot{x}\cap\dot{y}|=\frac{1}{n}|\dot{y}\cap\dot{x}|=\mu_y(x) \tag{3}$$

it show that $r_{ij}=r_{ji}$, thus $R\triangleq[r_{ij}]\triangleq[\mu_{x_i}(y_j)]$ is a fuzzy similar relation matrix on X.

II. Confidence method: On the base of hard partition method, every people gives a degree of confidence $\lambda_i\in[0,1]$ for their given partition $\{B_{i1},B_{i2},\ldots,B_{ik_i}\}$. Denote

$$\mu_x(y)\triangleq\frac{1}{\sum_{i=1}^{n}\lambda_i}\sum_{B_{ij}\in\dot{x}}\lambda_i\,\chi_{B_{ij}}(y) \tag{4}$$

thus the matrix $R=[r_{ij}]\triangleq[\mu_{x_i}(y_j)]$ is a fuzzy similar relation matrix on X.

III. Composite method: This method is the combination above two methods. Suppose that $\{B_{i1},B_{i2},\ldots,B_{ik_i}\}$ is a given partition by ith people, for each B_{ij} $(j=1,2,\ldots,k_i)$ endow with a degree of confidence $\lambda_{ij}\in[0,1]$. For $x\in X$, denote that

$$\dot{x}\triangleq\{\lambda_{ij}|x\in B_{ij},\ j=1,2,\ldots,k_i;\ i=1,2,\ldots,n\} \tag{5}$$

$$\mu_x(y) \triangleq \frac{1}{\sum\limits_{\lambda_{ij}\in \underset{\cdot}{x}\cup \underset{\cdot}{y}} \lambda_{ij}} \sum_{B_{ij}\in \dot{x}} \lambda_{ij} \chi_{B_{ij}}(y), \quad \forall x,y \in X. \tag{6}$$

then $R=[r_{ij}] \triangleq [\mu_{x_i}(y_j)]$ is a fuzzy similar relation matrix on X.

Ⅳ. Soft partition method: In this method, every people give a soft partition $\{B_{i1}, B_{i2}, \ldots, B_{ik_i}\}$ on X, where $\bigcup_{j=1}^{k_i} B_{ij}=X$, when $l\neq j$, subset $B_{ij}\cap B_{il}$ may is no-empty, $\lambda_{ij}\in[0,1]$ is a degree of confidence of B_{ij}, $i=1,2,\ldots,n$; $j=1,2,\ldots,k_i$. Using methods (1)、(5) and (6) may obtain fuzzy similar relation matrix R on X.

Now we only prove formula (6).

1°. For $x\in X$,

$$\mu_x(x) = \frac{1}{\sum\limits_{\lambda_{ij}\in \underset{\cdot}{x}} \lambda_{ij}} \sum_{B_{ij}\in \dot{x}} \lambda_{ij} \chi_{B_{ij}}(x)$$

$$= \frac{1}{\sum\limits_{\lambda ij\in \underset{\cdot}{x}} \lambda_{ij}} \sum_{B_{ij}\in \dot{x}} \lambda_{ij} \quad = \frac{1}{\sum\limits_{\lambda_{ij}\in \underset{\cdot}{x}} \lambda_{ij}} \sum_{\lambda_{ij}\in \underset{\cdot}{x}} \lambda_{ij} = 1$$

thus we have $r_{ii} \triangleq \mu_{x_i}(x_i)=1$, $i=1,2,\ldots,n$.

2°. For all $x,y\in X$,

$$\mu_x(y) = \frac{1}{\sum\limits_{\lambda_{ij}\in \underset{\cdot}{x}\cup \underset{\cdot}{y}} \lambda_{ij}} \sum_{B_{ij}\in \dot{x}} \lambda_{ij} \chi_{B_{ij}}(y) = \frac{1}{\sum\limits_{\lambda_{ij}\in \underset{\cdot}{x}\cup \underset{\cdot}{y}} \lambda_{ij}} \sum_{B_{ij}\in \dot{x}\cap\dot{y}} \lambda_{ij} = \mu_y(x)$$

then have $r_{ij}=r_{ji}$, for any $i=1,2,\ldots,m$; $j=1,2,\ldots,m$.

3. Set-valued statistics method apply to establish decision weight

The set-valued statistics method can be used to determine weight vector in multiply object decison (MADM). At fact, the estimate method by specialists can also be regarded as a set-valued statistics method, namely, the weight vector $W^{(i)}(\omega)=(\omega_1^{(i)},\omega_2^{(i)},\dots,\omega_n^{(i)})$ estimated by every specilist may be seen as a fuzzy subset on factor set X. Then the estimate of projected function $W=\frac{1}{m}\sum_{i=1}^{m}W^{(i)}(\omega)$ may be taken as decison weight, where m is number of specilists.

In a lot of practice, it has been found that there are many weaknesses in this method. Where main weaknesses are the following: ① If the factors of multiply object decison are too many, then it is very difficult to hold the tiny difference between these factors. ②. In many decision problems, the weight is not unique, etc.. In this section, we using the set-valued statistics method give a way to determine the decision weight vector.

Let $\mathscr{A}=\{A_1,A_2,\dots,A_n\}$ be the factors set, $X_k\subseteq[0,1]$ the states set of the factor A_k, and suppose that X_k is a finite set for all k=1,2,...,n. Denote that $X_k=\{x_{k1},x_{k2},\dots,x_{kt}\}$ and $x_{k1}<x_{k2}<\dots<x_{kt}$. x_{ki} is called the ith level of factor A_k. For convenience take t levels for every factor. call $X=\prod_{k=1}^{n}X_k$ as the state space of $\mathscr{A}$. Any element $x\in X$ is called a tack. Obviously, X has t^n tacks. Denote $V=\{v_1,v_2,\dots,v_l\}$ as set of tack groups.

In the weight estimate method by specialists, when factors are more and the difference between these tacks are little, specialist's estimate is instable usually. Beacuse, when the specialist looks at a conerete tack or programme, there is a viewpoint of viewing the situation as a whole in his mind. The last decision made by him is not simply a weighted arithmetic mean. The analytical process used the set-valued statistics provide with the concept of viewing the situation as a whole.

We select a part of typical tacks from X, and show clearly every tacks belong to which group V_i. In order to guarantee that tacks $x\in X$ have typicality, may use the method of orthogonal experiment, design a table of orthogonal experiment $L_T(l^n)$, where n is the number of factors, l the number of levels and T the number of typical tacks. Denote $M(\subset X)$ as the tacks set selected.

Suppose that m persons (specilists) attend this test, every person makes a partition to M. Let $\{V_1^{(k)},V_2^{(k)},\dots,V_l^{(k)}\}$ be a partition obtained by kth person, and $\bigcup_{i=1}^{l} v_i^{(k)}=M$, $v_i^{(k)}\cap v_j^{(k)}=\Phi$ when $i\neq j$.

Suppose that $x_1^{(i)},x_2^{(i)},\ldots,x_p^{(i)}$ are tack samples puted into the group v_i in the result of partition by m persons. Denote that $x_j^{(i)}=(x_{1j}^{(i)},x_{2j}^{(i)},\ldots,x_{nj}^{(i)})$, $j=1,2,\ldots,p$. Every tacks may be thinked a fuzzy set on discussed field $\{X_1,X_2,\ldots,X_n\}$, the level of factor as the degree of membership. We denote that

$$\bar{u}_{V_i}(x_k)=\frac{1}{p}\sum_{j=1}^{p} x_{kj}^{(i)} \tag{7}$$

$$\bar{D}_{V_i}(x_k)=\frac{4}{p-1}\sum_{j=1}^{p}[x_{kj}^{(i)}-\bar{u}_{V_i}(x_k)]^2 \tag{8}$$

$i=1,2,\ldots,l$. Where $\bar{u}_{V_i}(x_k)$ is the level mean of samples of the factor A_k for group v_i, $\bar{D}_{V_i}(x_k)$ the degree of stability of levels. On the factor A_k, the separate degree of v_i and v_j is defined the following

$$\rho(x_k;v_i,v_j)=\{|\bar{u}_{V_i}(x_k)-\bar{u}_{V_j}(x_k)|+[1-\min(\bar{D}_{V_i}(x_k),\bar{D}_{V_j}(x_k))]\}.$$

By

$$\bar{\rho}(x_k)\triangleq\frac{2}{l(l-1)}\sum_{i=1}^{l}\sum_{i<j}^{l}\rho(x_k;v_i,v_j) \tag{9}$$

express the mean separate degree of a factor A_k. Obviously, the larger $\rho(x_k)$ is, the larger the action that factor A_k is used to distinguish class V, thus the more importand this factor is. Let

$$W=(\bar{\rho}(x_1)/\alpha,\ \bar{\rho}(x_2)/\alpha,\ldots,\bar{\rho}(x_n)/\alpha) \tag{10}$$

where $\alpha=\sum_{i=1}^{n}\bar{\rho}(x_i)$, then W may be used as a weight vector of MADM.

4. Set-valued statistical method apply to forecast problems.

Let $\mathscr{A}=\{A_1,A_2,\ldots,A_n\}$ be a factors set, X the forecast field. f is a set-valued mapping from $\mathscr{A}$ to X, $\forall A_i\in\mathscr{A}$, $f(A_i)=\sigma_i\in\mathcal{P}(X)$, $\chi_{\sigma_i}(x)$ the characteristic function of set σ_i. There is a probability distribution under the condition of mapping f on $\mathscr{A}$,

$$p(A_i|f)\triangleq p_i,\quad i=1,2,\ldots,n \tag{11}$$

The projected function

$$u(x)\triangleq\sum_{i=1}^{n}p(A_i|f)\chi_{\sigma_i}(x) \tag{12}$$

is called the conditional fuzzy set under $\mathscr{A}$ and condition of mapping f on X.

The concept of conditional fuzzy set can be applied to make some forecast problems.

Example (Decide ignited location in stope space)

This example belongs to the station forecast problem about complicated environment, it is a typical example which using the informations reflected on every factors infer the station that event will happen.

In mining production, the self-ignited in coal face is a great natural calamity, it will menace and influence seriously the safety in the coal mining production, and will bring a large number of economic loss to mine. Because the condition in stope space is complicated and bad, and the station of source of fire is not to be closed vith, all these vill make that work to put out a fire has heary blindness. So we only can infer the possible station of the ignition according to some informations relative to ignition and other factors which can be measured directly.

Consider factor set $\mathscr{A}=\{A_1,A_2,\ldots,A_6\}$, where

A_1: Exist the remnant coal of easy self-ignited;

A_2: Have the condition of oxygen supplied;

A_3: Have the environment of store heat;

A_4: The oxidizing time;

A_5: The move of gas (CO);

A_6: The source of temperature.

The mining stope space is expressed by set D on plane, (s,m) is a point in D. Let

$$H_i \triangleq H_i(s,m) \triangleq \{(s,m) \mid (s,m) \in D \text{ and satisfies factor } A_i\}$$

$$i=1,2,\ldots,6. \tag{13}$$

Obviously, H_1 is a subset of D. But in the prectical problems, whether the point (s,m) satisfies condition A_1 can not be answered certainly. For example, H_1 is dicided by the distribution state of remnant coal inside stope space. And it depend on the structural state of coal bed, the degree of carburization of coal, the forms of mine coal, etc.. So the range of H_1 can only be given an approximate estimation by person basis on the subjective experience and comprehensive analysis for many factors. This range may be given by many specilists according to their speciality experience using set-valued statistics method and these differient ranges (some subsets on X). This range is a fuzzy field on D, denoted as $\underset{\sim}{H}_i$, i=1,2,...,6. The station of ignition is determined as the following

$$\underset{\sim}{H} = \bigcap_{i=1}^{6} \underset{\sim}{H}_i, \quad \underset{\sim}{H}(s,m) = \bigwedge_{i=1}^{6} \underset{\sim}{H}_i(s,m) \tag{14}$$

In order to decide $\underset{\sim}{H}_5$ (the field of harmful gas CO gushed out), we advance a concept of CO relation set.

After occur self-ignited in a part of stope space, the temperature will increase at this point, and release harmful gas CO etc.. With the flowing of gas in stope space, gas CO will go on convection and diffusion, such that CO content of air in some points $(s,m) \in D$ will take place changer. Suppose that $Q=\{q_1,q_2,\ldots,q_n\}$ is n points observed. If we have good reason for believing the CO gas measured at point qi is released from point $q \in D$, then point q is called a CO-relation point of q_i. $\underset{\sim}{F}_{\bullet}^{(i)}(s,m) \in [0,1]$ is the probable degree that point (s,m) is a CO-relation point of q_i. Thus we obtain a fuzzy subset $\underset{\sim}{F}_{\bullet}^{(i)}$, its membership function is $F_c^{(i)}(s,m)$, $F_c^{(i)}$ is called as CO-relation fuzzy set of q_i.

Suppose that c_i is CO content value measured at point q_i at a certain time. That c_i value is big or small may depict a probable degree which CO-relation fuzzy set of q_i is the release field of harmful gas CO, take the probability distribution on Q as

$$P_i=\frac{C_i}{\sum_{j=1}^{n} C_j}, \quad i=1,2,\ldots,n.$$

Hence, the membership function of fuzzy set H_s may defind with formular

$$H_s(s,m)=\sum_{i=1}^{n} p_i F_c^{(i)}(s,m). \tag{15}$$

5. Summary

The basic of set-valued statistics is theory of random set and it's projection. For $\forall x \in X$, by random set $R(\Omega,\mathcal{A};\mathcal{F}(X),\mathcal{B})$ may induce a random variable on $I_x=[0,1]$, i.e. for $\forall x \in X$, it follows a random variable that takes value at $[0,1]$. In that case, it is different from a family of random variables with parameter x (as random process), beecause, for a given family of random variables with parameter x, the corresponding random set is not unique. However, the most different is the form of samples obtained in statistics test. The samples of set-valued statistics test are some subsets, these sets are often a fiducial field given by person who is asked to do the test, through a whole synthetical analysis in a larger psychological fiducial degree. It can mirror very well the whole of psychological measure to the object. It may avoid statistical unstability of psychology when isolat consider problems.

It only has a few years histry that the theory of set-valued statistics is advanced, there are a lot of work to need in the theory. However, many applied methods about the set-valued statistics have been advanced or are being advanced continually. They will be apply to more and more of engineering practical problems.

References

[1].Wang Peizhuang and Liou Xihui, Set-valued Statistics, J. of Engineering Math., No1,(1984).

[2].Wang Peizhuang, Fuzzy Set and Projection of random Set, Beijing Normal University Publishers, (1985).

[3].Wang Peizhuang and E.Sanchez, Treating a Fuzzy Subset as a Projectable Random Subset, Fuzzy Information and Decision Processes, M.M.Gupta and E.Sanchez (eds.) North-Holland, (1982).

[4].Guo Sizhong, Fuzzy Set-Valued Statistics, J.of Fuxin Mining Institute, No4,(1986).

[5].Guo Sizhong and Peng Xiantu, Method of Set-Valued Statistics Apply to Fuzzy Clustering,J. of Engineering Math., No3,(1988).

[6].S.Tamura, S. Higuchi and K. Tanaka, Patten Classification Basedsion Fuzzy Relations, IEEE Trans, Syst., Man Syberml, P61-66, (1971).