Factor space is the adaptive and deepening theory of fuzzy sets

Haitao Liu
Institute of Intelligence Engineering and Mathematics, College of Science
Liaoning Technical University
Fuxin, China
haitao641@163.com

Runjun Wan
Institute of Intelligence Engineering and Mathematics, College of Science
Liaoning Technical University
Fuxin, China
wanrunjun@gmail.com

Shanshan Xue
Institute of Intelligence Engineering and Mathematics, College of Science
Liaoning Technical University
Fuxin, China
1024396818@qq.com

Tiantian Wang
Dept. of Data Science and knowledge Engineering
University of Maastricht
Netherlands
2698264255@qq.com

Sizong Guo
Institute of Intelligence Engineering and Mathematics, College of Science
Liaoning Technical University
Fuxin, China
lnzxtaw32@126.com

Jing He
School of software and electrical engineering
Swinburne University of Technology
Hawthorn, Australia
480245@qq.com

***Abstract* — In recent years, the factor space theory has been promoted gradually as math foundation for mechanistic artificial intelligence theory. The theory was put forward in 1982 by Prof. P Z Wang, he took the fundamental space Ω in probability and the universe U of fuzzy sets both as factor spaces, but put Ω in the sky 2^U=P(U). The established fuzzy shadow theory says that the membership function on the ground is the coverage of random set in the sky, and had proved the Existence and Uniqueness Theorem on the correspondence between earth and heaven. This theory points out that the adaptive platform of intelligent description and subjective measurement is the factor space; and the core transform in between different levels is the power mapping (falling shadow). This is the mathematical secret of artificial intelligence, but also the direction of further improvement of fuzzy sets and systems.**

Keywords—probability, fuzzy sets, power of sets, factor space, falling shadow, random sets

I. Modern Probability Theory Is Built on Factor Space

In 1936, Kolmogorov established the axiomatic system of probability. He defined a random variable as a measurable mapping from basic space to a solid line. With this definition, the classical probabilities in basic space can be transformed into real straight lines through mapping to form various modern probability distributions. Without the definition of random variables, there would be no such probability distributions and modern probability theory. No talk about the measurability of the map, but only talk about the map. This is a rigorous mathematical concept, but how can it be related to random variables? For example, is the number head up 1 or 0 when you are tossing a coin? How can such an accidental variable become a certain mapping? This requires consideration of various factors such as coin shape, initial position, upswing, finger movement, desktop shape, environmental impact, and others. Draw a coordinate frame around these factors with a sufficient number of axes, and the state of each factor is observed, described, and controlled extremely carefully, then each small point representing a situation can get the exact result, either positive or negative. Otherwise, there must be some influential factors that do not have been considered. Once all factors are taken into account which plays a role, in such an ideal space, the direction of the coin can be determinable. The space formed by such factors as the axis is the basic space Ω that Kolmogorov thinks of, and the basic space is factor space. Moreover, each experiment the modern probability theory is built on the basis of factor space. According to Prof. Wang [1-8], modern probability theory is built on factor space.

1.1 Describing probability with factors

If the prediction of the occurrence of an event is taken as the goal, the prediction test is regarded as a factor f, called experiment. It is assumed that the experiment information field $I(f)=\{a_1,\ldots,a_n\}$ contains only a limited number of results, and each experiment must have one and only one result. Randomness is caused by $n>1$. Modern probability theory is developed on measurement theory. Measurement theory can only solve the performance and expansion of probability. The determination of probability comes from the symmetry of the experiment.

Definition 1.1 Let experimental factor be symmetry factor which has symmetry result. Such as, Toss a coin, roll a dice, take a ball out of a bag containing the same shape, etc.

Based on symmetry factor, this paper proposes the symmetry criterion.

Prof. Haitao Liu is an active scholar in factor space and its applications in data mining.

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

1.2 Symmetry criterion

Assume all other factors do not interfere with the symmetry, the symmetry experimental factor guarantees that the results will have the same probability:

$$P(a_1)=\ldots=P(a_n)=1/n. \quad (1)$$

The additional conditions here are fundamental, such as tossing a coin. Although the shape of the coin is symmetrical on both sides, if the movement of the hand (an important factor affecting the result) intentionally leans to one, it destroys the symmetry and the criterion cannot be satisfied. The way to avoid this interference is simple: the initial position and movement of the coin must be designed to be fair.

Definition 1.2 Let $\mathscr{A}$ be the σ-domain generated by $I(f)$. The probability P expanded from equation (1) on $\mathscr{A}$, is called the probability under the symmetric factor f, namely factor probability.

The existing probability theory has no clear definition of conditions in its application. Say that the dam can withstand the once-in-a-century flood, under what conditions is the probability of the once-in-a-century encounter defined? Probability without strictly defined, there is an irreversible risk. Factor probability is clearly defined by conditions, and this is the symmetry of factor f.

Factor probability is essentially a kind of logical probability, which reflects the causal relationship among things. It is the essence and must be supported by frequency experiments.

Now, some scholars mention subjective probability, which is not a pejorative term; it reflects the subjective initiative of the subject and is worth studying. Factor probability is not a subjective probability, but it can provide a logical basis for it.

The application range of factor probability is wide. Most data can be regarded as factor probability, which has the next criterion.

1.3 Data equality criterion

Without special assumptions, all sample points in the sample set are equal and have the same weight. The sample probability distribution determined by a sample of size m in the factor space is that each sample point has the probability $1/m$.

Assume that the sample set is some scores from a group of experts, and the level of evaluation for this group of experts are different. At that situation, the data equality criterion is inapplicable, and it is difficult for us to judge whether a set of sample points is truly equality. It is a detailed indication of the source, author, time, place, and name of the data to remove the problem.

If the data equality criterion works, various sample distributions are formed in the state space. The multifactorial experimental space carries the joint probability distribution of traits, which describes the generalized law of causation between factors.

1.4 Maternal distribution of factor probability

The distribution generated by a sample set is not a true probability distribution, and the probability is also random. Only by increasing the number of samples and reaching the limit properties of the sample distribution can obtain the true probability distribution, that is, the maternal distribution. Furthermore, existing probability distributions can be recertified by this method. In some fields, such as there are many unknown distributions in remote sensing data, Monte Carlo [9] method can be used to obtain new distribution types and expressions according to the idea of data equality.

The experimental factors that can be measured by physical instruments all use a real number interval as the experimental space. If the sample distribution is multimodal in this interval, you can combine the actual needs to divide the property according to the peaks, and change the experimental space into a discrete type for artificial intelligent intension description. (This is only a hard segmentation. For the fuzzy segmentation, it needs the fuzzy set to be further improved). This work is called the calibration of factor space. Without calibration, the coordinate shelves of factor space cannot put together.

The same kind of probability distribution is called a style. Each style has parameters. The two most important parameters are sample mean and sample mean square error. These two parameters serve two purposes: (1) Generalization of data processing. Perform a linear transformation on all sample points so that these two parameters become 0 and 1, respectively, and the homogeneous distribution becomes one. This is the generalization principle in image data processing. (2) Using these two parameters as hidden parameters, solve or iterate in an optimized format, and the result is the key factor sought in classification and learning. Artificial intelligence is to find hidden factors, hidden factors can be set as hidden parameters.

II. The Domain Of Fuzzy Sets Should be Based on Factor Space

In 1978, Prof. Wang began to engage in fuzzy mathematics. L.A. Zadeh[10] defines fuzzy sets as a membership function on the domain U, but nobody cares about U. Differ from other scholars, Prof. Wang insists on treating the domain of fuzzy set as a factor space, and only by clarifying the factors can make fuzziness clear. Using the idea of factor space[8], he has clarified the difference and connection between the two uncertainties of fuzziness and randomness. Randomness is the broken for the law of causation caused by the insufficient grasp of conditional factors. Probability is the general law of causality. Fuzziness is broken for the law of excluded middle due to insufficient cognitive factors. Membership is the general law of the excluded middle. The experimental model of randomness is, "fixed circles, ideas are changing". And the experimental model of fuzzy is, "fixed ideas, circles are changing". Under this duality, Prof. Wang finds that the fuzziness of the domain U (on the ground) can be transformed into the randomness of the power $P(U)$ (in the sky), and the membership can be objectively measured by the fall shadow of the random set (cloud) on the ground. This is the fuzzy set fall shadow theory established by Prof. Wang. This theory not only provides a theoretical basis for set-valued statistics (including interval statistics), but also proves the unity of the existence uniqueness theorem for the four non-additive measures (trust, likelihood, non-likelihood, and non-trust) which are proposed for subjective measures. The fuzzy fall shadow theory deepens the research of the international fuzzy set theory. Based on this theory, in the development of fuzzy inference engines in 1988, there was good news that Dr. Zhang Hongmin and other doctoral students under the guidance by Prof. Wang developed the second international fuzzy inference engine. It improves inference to 15 million per second from 10 for the Fuzzy Computer pioneered by Yamakawa in 1987. Meanwhile, the volume is less than one tenth of Yamakawa's.

Until now, fuzzy mathematics cannot be recognized as type, expression, and table as probability theory. Type is the type of probability distribution, such as binomial distribution, Poisson distribution, normal distribution, etc. The expression refers to the mathematical expression of these distributions, and the table is a table for people to check the confidence limits according to the probability accuracy. None of this work exists in fuzzy mathematics. Therefore, our service to society is inadequate. Factor space is the deepening direction of the fuzzy set theory. This work must use the fuzzy fall shadow theory, which is an essential part of guiding and popularize interval statistics.

The relationship between heaven and earth is the relationship between power and set. There are n elements on the ground, and there are 2^n elements in the sky. It is extremely difficult to verify. The trick to solving the problem is to consider background power. What is background power? Let D=[0,1], we consider all subinterval of D : $P_I([0,1])=\{A\in I|A\subseteq[0,1]\}$ rather than subset of D, where I represents a set of intervals and called the background. Background power limits power in the background. Such a limitation is to limit set-valued statistics to interval statistics in practice. When Nanlun Zhang first surveyed the fuzzy concept of "youth", he used interval statistics, which was reasonable. This limitation cuts down a lot of statistics. Such an example,

D={a,b,c,d,e,f,g}, $P(D)$ contains 128 subsets, but if restricted to ligature, then

$$|P_I(D)|=7+6+5+4+3+2+1=28$$

With empty ligatures, there are only 29 elements. Assume the probability which has total 1 is assigned to the following 5 ligatures:

cde 0.4, bcde 0.2, dcf o.2, bcdef 0.2.

The fuzzy fall shadow of this random ligature on D is

0	0.4	0.8	1	1	0.4	0
a	b	c	d	e	f	g

This is a simple model of random set fall shadow.

With the simplification of the interval as the background power, the membership curve on the real number domain depends only on the changes of the two random variables at the left and right endpoints of the interval. As shown in Fig.1, the left (right) endpoint of the membership curve is the left (right) distribution function of the distribution density for the point on the left (right) side in the random interval.

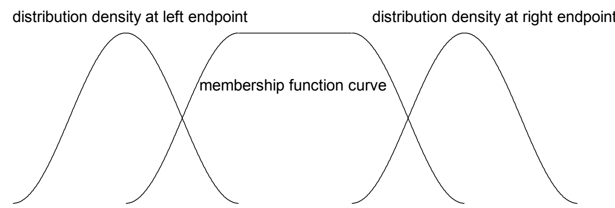

Fig. 1. Left tail, flat top and right tail of membership curve

A one-dimensional membership curve is divided into three segments: left tail, flat top, right tail. If the membership curve is cut off in the middle, only one tail and some flat tops considered, the membership curve equates the distribution function of the random variable. It is not difficult to prove the theorem 2.1.

Theorem 2.1(The conversion between membership curve and probability distribution function theorem) The expression of the left (right) tail of the membership curve is equal to the left (right) distribution function of the distribution density at the left (right) endpoint of the random interval.

According to theorem 2.1, the tail type of the membership curve can be determined by the probability density function. Three main types are described.

(1) **Negative power type of membership curve tail**. Assume that the distribution density of the left (right) tail of the membership curve is

$$p(x)=\theta/(x-a)^2\}\ (x>a)$$
$$(p(x)=\theta/(a-x)^2\}\ (x<a)).$$

Then, the left (right) tail expression of membership function is

$$\mu(x)=c/(a-x)\ \ (-c<x<a-1/c\ )$$
$$(\mu(x)=c/(x-a)\ \ (a+1/c<x<c\ )).$$

(2) **Negative exponential type of membership curve tail**. Assume that the distribution density of the left (right) tail of the membership curve is

$$p(x)=e^{-\theta(a-x)}\ (x<a)$$
$$(p(x)=e^{-\theta(x-a)}\ (x>a)).$$

Then, the left (right) tail expression of membership function is

$$\mu(x)=c\theta(e^{-\theta(a-x)}-1)\quad(-1/\varepsilon<x<a-\varepsilon)$$
$$(\mu(x)=c\theta(e^{-\theta(a-x)}-1)\quad(a+\varepsilon<x<1/\varepsilon),$$

where $c=1/\theta(e^{-\theta\varepsilon}-1)$).

(3) **Logarithmic type of membership curve tail**. Assume that the distribution density of the left (right) tail of the membership curve is

$$p(x)=\theta/(a-x)\qquad(x<a)$$
$$(p(x)=\theta/(x-a)\qquad(x>a)).$$

Then, the left (right) tail expression of membership function is

$$\mu(x)=c(\ln\theta(1/\varepsilon)-\ln\theta(a-x))\quad(-1/\varepsilon<x<a-\varepsilon)\ c=-1/2\ln\varepsilon$$
$$(\mu(x)=c(\ln\theta(x-a)-\ln\theta(1/\varepsilon))\quad(a+\varepsilon<x<1/\varepsilon)\ c=1/2\ln\varepsilon).$$

Negative exponential and logarithmic types should become the norm for fuzzy distributions. Logistic regression is a membership curve of negative exponential and logarithmic types.

Let $L=\{x_k=(x_{k1},\ldots,x_{kn};\ y_k)\}_{(k=1,\ldots,K)}$ be a set of equality medical data. Where $x_{k1},\ldots,x_{kn}$ represent the n pathology factor indicators of the kth tester, and $y_k = 1$ (with certain disease) or 0 (without certain disease). Each data point falls in a super matrix of R^n with probability $1/K$. This super matrix can be divided into several grids. Let $f_{i1\ldots in}$ be the ratio (ie. frequency) of diseased sample points to the number of samples fall in the grid with $i_1,\ldots i_n$ as the foot code, and in simplified form f_i. $f_i/(1-f_i)$ is called the likelihood ratio in medicine field. Note that take the logarithm of the likelihood ratio, and let $y_i=\ln f_i/(1-f_i)$, which changes on the n-dimensional points. Fit it with an n-dimensional hyper-plane. Let

$$y=\theta_1x_1+\ldots+\theta_nx_n-a=\theta x-a$$

$$(\theta_1,\ldots,\theta_n;\ a)=Argmin_\theta\Sigma_i(y-y_i)^2. \qquad (2)$$

It is easy to prove that the equation for fitting membership surfaces is

$$\mu(x)= e^{\theta x-a}/(e^{\theta x-a}+1). \qquad (3)$$

Where $\mu(x)$ is the logistic regression membership surface (or logarithmic regression membership surface) in equation (3). The purpose of membership regression is to determine which pathology factor is important based on the size of each component of the hidden parameter $\theta_1=(\theta_1,\ldots,\theta_n)$. Logistic regression is the first but not yet recognized type of membership surface (it is earlier than the type proposed in this paper). The reason why it is not recognized is that most of the proposed and applied people are not the people who studies in fuzzy mathematics. They did not put this fitting surface is classified as a fuzzy membership surface, but they consistently emphasize that this surface should be used when determining the concept of right and wrong, which is actually the feature of the membership surface.

Fuzzy mathematics has its typical distribution, so it is not difficult to establish corresponding tables for people to find the confidence limits of fuzzy reasoning. With the membership curve distribution in this paper, it can be said clearly. Chenguang Lu[11] suggested to call it the Confirmation Measure (CM). Whether or not there is type, expression, table, is a sign that fuzzy mathematics is mature. It will be an important direction for deepening fuzzy mathematics.

III. Promotion of Mathematical Structure

3.1 Power lattice

The set $\mathscr{P}(S)=2^S$ which is all subsets of the set S is the power of S. Each subset of S becomes a point in $\mathscr{P}(S)$. The largest point is S, and the smallest point is the empty set□∅. An order structure $(\mathscr{P}(S), \leq)$ is obtained by describing the containment relationship between sets in power. It is easy to prove that this order structure is a partial order relation. This is because the upper and lower bounds exist. And because it satisfies the four basic law. Then, the order relation is a lattice, denoted as L. Given A, B, C$\in L$. Obviously it satisfies

Reflexivity: $(\forall A\in L)$ A≤A;

Transitivity: If A≤B and B≤C, then B≤C.

Therefore, $(\mathscr{P}(S), \leq)$ is a quasi-ordered set. Obviously, it also satisfies

Antisymmetry: If A≤B and B≤A, then A=B.

Therefore, $(\mathscr{P}(S), \leq)$ is a partially ordered set.

Obviously, for any A,B$\in L$, if the two set are performed unit operation (intersection operation) in S, $A\cup B(A\cap B)$ is obtained, and then placed in $\mathscr{P}(S)$. The true bound is written as A∨B =Sup{A,B}(A∧B =Inf{A,B}). So $(\mathscr{P}(S), \vee)((\mathscr{P}(S), \wedge))$ is an upper (Bottom) semi-lattice. The semi-lattice must satisfy the following laws.

(1) Exchange law: A∨B=B∨A, A∧=B∧A;

(2) Binding law: (A∨B)∨C=A∨(B∨C), (A∧B)∧C=A∧(B∧C);

(3) Idempotent law: A∨A=A, A∧A=A.

Given an upper (lower) semi-lattice $(\mathscr{P}(S), \vee)$ $(\mathscr{P}(S), \wedge)$, the order relation can be defined from the half-lattice operation in turn:

$$A\leq B \Leftrightarrow A\vee B=B\ \ (A\leq B \Leftrightarrow A\wedge B=A).$$

Then $(\mathscr{P}(S), \leq)$ ($(\mathscr{P}(S), \geq)$) must be a partially ordered set. Order ≤ and order ≥ are reciprocal.

For the power $\mathscr{P}(S)$, $(\mathscr{P}(S), \vee)$ and $(\mathscr{P}(S), \wedge)$ are both semi-lattices, so the power is a lattice. And obviously, satisfy

(4) Absorption law: (A∧B)∨B=A, (A∨B)∧B=A.

Thus, $(\mathscr{P}(S), \vee, \wedge)$ is a dual lattice.

(5) Distribution law:

(A∨B)∧C=(A∧C)∨(A∧C), ((A∧B)∨C=(A∨C)∧(B∨C)).

Then $(\mathscr{P}(S), \vee, \wedge)$ is a distributive lattice. Any subset A in D has a coset A^c, and the cooperation determines a unary operation ' in the power lattice L, which satisfies the following law.

(6) Duality law: (A')'=A, (∅'=D, D'=∅).

(7) De Morgan's law (reverse convergence):

(A∨B)'=A'∧B' , (A∧B)'=A'∨B'.

$(\mathscr{P}(S), \vee, \wedge,')$ is a Hayting algebra or soft algebra. Obviously, it satisfied the middle row.

(8) The law of the middle row: A∨A'=D, A∧A'=∅.

So $(\mathscr{P}(S), \vee, \wedge,')$ is a Boolean algebra, and it is isomorphic with the map $A \longmapsto$ A with Boolean algebra $(\mathscr{P}(S), \cup,\cap,^c)$. In the following paragraphs, the set algebra $\mathscr{P}(S)$ are all written in power lattice. If the concept of exact bound is generalized, it also satisfies Completeness.

Completeness: ($\forall \mathbf{C} \subseteq L$) If there is an upper (lower) bound, then the upper (lower) bound must be the smallest upper (lower) bound. Let $\vee\{A_t|t\in T\}=\mathrm{Sup}\{A_t|t\in T\}(\wedge\{A_t|t\in T\}=\mathrm{Inf}\{A_t|t\in T\})$.

So $\mathscr{P}(S)$ is a complete partial ordered set, complete upper (lower) semi-lattice, complete lattice and complete Boolean algebra. In S, they have no true subset except for the empty set, so they are called sub-elements in powers. Note that $\mathscr{P}_{\circ}(S)=\{\{x\}|x\in S\}$, which is isomorphic to S. For simplicity, write $\{x\}$ as $\boldsymbol{x}$.

3.2 Order structure promotion on power

Given a complete ordered set $(L, \geq)$, the order $\geq$ induces an order relationship $\geq$ in its power lattice $\mathscr{P}(L)$.

$$A \gg B \Leftrightarrow \forall a\in A \text{ has } b\in B \text{ to make } a\geq b.\ (A,B\subseteq L).$$

It is easy to prove that the relationship $\gg$ satisfies reflexivity, transitivity but does not satisfy anti-symmetry. In order for anti-symmetry to be true, an equivalent relationship must be declared: $A\approx B$ if and only if $A \gg B$ and $B \gg A$. According to this equivalent relationship, the entire power $\mathscr{P}(L)$ is classified to obtain its quotient space $\mathscr{P}'(L)=\mathscr{P}(L)_{/\approx}=\{[A]=\{B|A\approx B\}|A\in\mathscr{P}(L)\}$. The same order relationship is used in the quotient space, so $(\mathscr{P}'(L), \gg)$ constitutes a partial order set called the promotion of $(L, \geq)$. This promotion is obviously important. $B \gg A$ means that from the bigger to the smaller, which means that B has more staying power than A. By using L's inverse order $\leq$\ for promotion, that is, considering the power of $(L, \leq)$, the relation of promotion is recorded as $\ll$. The obtained quotient space $\mathscr{P}''(L)$ cannot be equal to the original quotient space $\mathscr{P}'(L)$. Although $(\mathscr{P}''(L), \ll)$ is also a partial order set, relationship $\ll$ cannot be reversed by relationship $\gg$, that is, $\ll\neq\gg$.

3.3 Representation of topology in power lattice

Topology describes invariance under continuous transformation. The continuity of the function at one point is represented by calculus as $\lim_{x\to a} f(x)=b$, which is described in ε-δ language: for any given $\varepsilon>0$, $\delta>0$ can be found, when $|x-a|<\delta$, that $|f(x)-b|<\varepsilon$. In terms of the distance space, as long as the variable x is close enough to the point a, then the variable $y=f(x)$ is arbitrarily close to the target b. But in the wider non-distance space, two strings of open intervals are used to describe invariance: a string of open intervals $(a-\delta, a+\delta)$ with a center a and a radius δ. The other is an open interval $(b-\varepsilon, b+\varepsilon)$ with a center b and a radius ε. No matter how small ε you give it, as long as x goes into a δ-circle that's small enough from a, you can make sure that y goes into a ε-circle of b. The circle is the neighborhood in topology, which is the first concept of topology.

The limit represents the process of approximation of a thing to a target. In order to show the degree of approximation, a series of the neighborhood should be set around the target. If a neighborhood system contains only a finite number of circles, then the destinations are reachable, so there is no need to talk about limits. The contribution of calculus is that a neighborhood system can contain an infinite number of circles, of which there is no smallest circle, making the goal impossible to reach forever, which is a scenario that mathematicians deliberately set up. ε-δ language is an approximation of the mathematical expression of a causal relationship, and it can be solved by logic judgment, and this is the mystery of the limit theory. The task of topology is to generalize the limit theory of calculus. This section only gives a superficial introduction from the point set topology and only deals with order structure in essence.

The general expression of a sequence is a net, which is a mapping $w:D\to X$. Here D is an index set, and X is an information space, which can be a one-dimensional or higher-dimensional Euclidean space, or a more general factor space. The index set must have a definite direction.

Definition 3.1 Suppose a partial ordered set $(P, \geq)$. $D\subseteq P$ is a directional set, if for any $x,y\in D$, it have $z\in D$ to $x\geq z$ and $y\geq z$.

For the dual lattice $(L, \vee, \wedge)$, the definition of directional set can be changed to: if any $x,y\in D$ has $x\wedge y\in D(x\vee y\in D)$, then D is the lower directional set (upper directional set). A mapping $w:D\to X$ from the directed set D to X is a net of indices D in X.

In practice, two index sets should be fixed: the large order structure is fixed in real space R, and the directional subset D^0 is sorted by non-negative integers from small to large, which can be used as the index set for any sequence. The directional subset D^1 is real numbers in the interval $(0,1]$ which are sorted from large to small, which can be used as the index set for any one-dimensional continuous variable. For example, $w\to+\infty$ can be represented as $w=1/d\ (d\to 0)$. For two-dimensional sequences, a new type of through-flow net could be defined based on D^0 or D^1.

Definition 3.2 In two dimensions real space, suppose $(T,\geq)$ is a directional subset on the X-axis, and for each $t\in T$, we have a directional subset $(D_t, \geq)$ on the Y-axis. Denote $D_{\Pi}=\{(t, \beta)|t\in T, \beta\in D_t\}$, define the dictionary order:

$$(t_1, \beta_1)|\geq (t_2, \beta_2) \Leftrightarrow t_1>t_2 \text{ or when } t_1\neq t_2, \beta_1>\beta_2.$$

Here $t_1>t_2$ stands for $t_1\geq t_2$ and $t_1\neq t_2$. It is not difficult to prove that $(D_{\Pi}, \geq)$ is also a directional set. The net whose direction of passage is an index set is called a through-current net.

Subsequences also need to be generalized to subnets:

Definition 3.3 Suppose $(D, \geq)$ is a directional set, $(D', \geq)$ is called its sub-directional set if $D\subseteq D'$ and D' with $\geq$ is also a directional set. Let w and w' be two nets with D and D' as index sets respectively, w' is called the subnet of w, if D' is a sub-directional set of D and $d\in D'$, then $w(d)=w'(d)$.

Now, taking S as the real Euclidean space $X=R^n$, the topological space in the real function can be reconstructed in the order structure language. First of all, translate the ε-δ language into the power $\mathscr{P}(X)$.

Definition 3.4 Define the net $w=\{w(d)|d\in D\}$ in the given $\mathscr{P}(X)$ and the point $\boldsymbol{x}=\{x\}$ in $\mathscr{P}_\circ(X)$. If there is $d\in D$ for any neighborhood $\boldsymbol{y}$ of any $\boldsymbol{x}$, then there is $\boldsymbol{y}\geq w(d')$ when $d'\geq d$, then the net w converges to $\boldsymbol{x}$ and is denoted as $w(d)\to\boldsymbol{x}$.

The sequence convergence in a real variable function should satisfy a few basic rules. 1. If $x_n\equiv a$, then $x_n\to a$. 2. If $x_n\to a$, then any subsequence $x_n'\to a$. Conversely, the subsequence converges to a point, but the original sequence does not necessarily converge to that point. But there is a rule as followed. 3. If every subsequence has a subsequence that converges to a, then the original sequence must converge to a. And the last rule follows. 4. If for any n, we have $x_{nk}\to a_n(k\to\infty)$ and $x_{nk}\to a$ $(k,n\to\infty)$ when $a_n\to a$.

Moor-Smith once proved that the convergence relation satisfying the four rules can uniquely determine a topological space in the real field. These rules can be described in order structure language as below.

Definition 3.5(Moor-Smith convergence criterion) Let w be the set of all nets on $\mathscr{P}(X)$.Then Moor-Smith criterion can be written as:

W1 If $w(d)\equiv a$, then $w\to a$. $(a\in\mathscr{P}_\circ(X))$

W2 If $w\to a$ and w' is a subnet of w, then $w'\to a$.

W3 If any subnet of w has a subnet convergent to a, then $w\to a$.

W4 Suppose a set of directional indexes $(T,\geq)$. If there is $w_t\to a_t$ and $a_t\to a$ for any $t\in T$, then it goes through the stream-net $w_\Pi\to a$.

According to these four criteria to check the neighborhood system, the topological space can be constructed based on the neighborhood system.

Definition 3.6(Topological space constructed by neighborhood system) For each $x\in X$, i.e., $\boldsymbol{x}=\{x\}\in\mathscr{P}_\circ(X)$, there is $\boldsymbol{y}\subseteq X$, i.e., $x\in\boldsymbol{y}\in\mathscr{P}(X)$, called the neighborhood of point x, denoted as $N(\boldsymbol{x})$, called the neighborhood system of point x or single point set $\boldsymbol{x}=\{x\}$. $\mathscr{N}=\{N(\boldsymbol{x})|x\in X\}$ is called a topological space on X. If it satisfies

N1 (Directed) : $(N(\boldsymbol{x}),\geq)$ is the lower directed set $(x\in X)$;

N2 (Full) : $\boldsymbol{y}\in N(\boldsymbol{x})$ and $\boldsymbol{z}\geq\boldsymbol{y}\Rightarrow\boldsymbol{z}\in N(\boldsymbol{x})$ $(x\in X)$;

N3 (Consistency): $\boldsymbol{y}\in N(\boldsymbol{x})\Rightarrow(\exists\boldsymbol{z})(\boldsymbol{y}\geq\boldsymbol{z}\geq\boldsymbol{x})(((\forall\boldsymbol{x}')(\boldsymbol{z}\geq\boldsymbol{x}'))\Rightarrow\boldsymbol{z}\in N(\boldsymbol{x}'))$ $(x\in X)$;

This definition is a direct translation of the real function on the lattice language of the definition of the neighborhood system, which is not novel, so there is no need to prove why this definition defines a topological space, which guarantees the four rules of convergence of Moor-Smith.

N3 means that any neighborhood must contain a consistent neighborhood, which is $\boldsymbol{z}\geq\boldsymbol{x}\Rightarrow\boldsymbol{z}\in N(\boldsymbol{x})$, where $\boldsymbol{x}$ is a single point set, $\boldsymbol{z}$ is a subset of X, and for any single point it contains, it is the neighborhood of that point. The neighborhood is the first concept of topology, but it is so broad and numerous that any set containing a neighborhood of a point is a neighborhood of that point, in the sense of N2. In order to show the infinite process of the full performance, the target only can appear in the interior of the neighborhood where it cannot run to the border. Any point in the open interval is its interior point. So it can be a neighborhood of any of them. Non-open intervals do not have this property. Therefore, N3 emphasizes that consistency is to distinguish the open neighborhood from the general neighborhood, and remove the division of point system so that there is only one open set system in the whole topological space.

Definition 3.7(The topological space defined by the open set) $(\mathscr{P}(X),\mathscr{O})$ is a topological space on X if it satisfies

K1 $\emptyset, X\in\mathscr{O}$;

K2 if $\boldsymbol{x},\boldsymbol{y}\in\mathscr{O}$, then $\boldsymbol{x}\wedge\boldsymbol{y}\in\mathscr{O}$;

K3 if $\boldsymbol{x}_t\in\mathscr{O}(t\in T)$, then $\vee_{(t\in T)}\boldsymbol{x}_t\in\mathscr{O}$

Notice that the element $\boldsymbol{x}$ in $\mathscr{P}(X)$, if it isn't restricted to being a member of the chassis $\mathscr{P}_\circ(X)$, then it's just any subset of X. They're in the power $\mathscr{P}(X)$, and the operation $\vee,\wedge$ is the operation $\cup,\cap$ in X.

From the concept of the neighborhood to the concept of an open set is a sublimation. It is much easier to describe topological Spaces with open sets. An open set is a set that is closed to a finite union and an arbitrary set and contains an empty set and a complete set. With an open set, it is easy to determine the neighborhood $N(\boldsymbol{x})$ of each point, which is filled with the power $\mathscr{P}(X)$ in reverse by the densification of N2 of all the open sets containing the point x. Compared with the open set, the concept of the neighborhood can be further simplified to filter form:

Definition 3.8 Let $(P,\geq)$ be a complete partial ordered set. The so-called Filter in P refers to a set F, which is a directional set and satisfies its fullness: if $\boldsymbol{y}\geq\boldsymbol{x}$ and $\boldsymbol{x}\in F$ then $\boldsymbol{y}\in F$.

From the definition of the neighborhood, N1 and N2 indicate that $N(\boldsymbol{x})$ is a filter.

Let $\boldsymbol{x}\uparrow=\{\boldsymbol{y}\in F|\ \boldsymbol{y}\geq x\}$, called top bullet of x. Easy to see, if $\boldsymbol{y}\in\boldsymbol{x}\uparrow$ and $\boldsymbol{y}\in\boldsymbol{x}\uparrow$, then $\boldsymbol{y}\wedge\boldsymbol{z}\in\boldsymbol{x}\uparrow$, so $\boldsymbol{x}\uparrow$ is the down directional set. Easy to see, if $\boldsymbol{y}\in\boldsymbol{x}\uparrow$ and $\boldsymbol{z}\geq\boldsymbol{y}$, then $\boldsymbol{z}\in\boldsymbol{x}\uparrow$, so $\boldsymbol{x}\uparrow$ is full set, then $\boldsymbol{x}\uparrow$ is filter.

Notice in the definition of a neighborhood, if $\boldsymbol{y}$ is a neighborhood of x, then $x\in\boldsymbol{y}$. This means that the neighborhood $\boldsymbol{y}$ of x must contain the point x, so the neighborhood filter appears to be $\boldsymbol{x}\uparrow$, the only difference being that $x\in\boldsymbol{y}$ means that it is not a single point set that $\boldsymbol{y}$ cannot be equal to x, and therefore $\boldsymbol{x}$ cannot belong to $\boldsymbol{x}\uparrow$. The conclusion is: $N(\boldsymbol{x})=\boldsymbol{x}\uparrow\backslash\{\boldsymbol{x}\}$.In other words, the neighborhood filter is the lower top bounce off the bounce vertex. $\boldsymbol{x}\uparrow\backslash\{\boldsymbol{x}\}$ is called the empty top bullet at point $\boldsymbol{x}$. The definition and symbol of empty lower ejection are simplified to $\boldsymbol{x}^{o}\uparrow=\{\boldsymbol{y} \mid \boldsymbol{y}>\boldsymbol{x}\}$.Here $\boldsymbol{y}>\boldsymbol{x}$ means $\boldsymbol{y}\geq\boldsymbol{x}$ and $\boldsymbol{y}\neq\boldsymbol{x}$.

The open set simplifies the structure of the topological space, and the topological space can also be simplified to the topological base.

Definition 3.9 $\mathscr{B}$ as a basis for topology C. if $\mathscr{B}$ is closed to the operation (i.e., $(\boldsymbol{y}, \boldsymbol{z}\in\mathscr{B}\Rightarrow(\boldsymbol{y}\wedge\boldsymbol{z}\in\mathscr{B})$ and can generate $\mathscr{C}$ (i.e., Any elements in $\mathscr{C}$ can be represented as the union($\vee$)of elements in $\mathscr{B}$.) $\mathscr{C}$ is called a quasi-basis for topology $\mathscr{C}$, if $\mathscr{B}$ can generate $\mathscr{C}$.

For example, all the open intervals in the real line R are the basis for the ordinary real open set. All open circles in the real plane R^2 are quasi-bases for ordinary real open sets. Given the base or quasi-base, topological spaces can be generated.

3.4 Lattice topology

The last section was a review of the topology of the real variable functions. However, the difference is that the narrative is completely in the language of order and lattice, although there is no inherent difficulty in this, the significance is remarkable. In 1985, prof. Wang proposed the concept of lattice topology for the first time, which made the theory of hyper-topology very simple (fuzzy sets and random set fall shadow). This section describes what a lattice topology is.

Lattice topology is to extend the content of the previous section from the power $\mathscr{P}(X)$ of real Euclidean space X to the power $\mathscr{P}(S)$ of any set S. Then extend the neighborhood system of point to the neighborhood system of set. This breaks the boundary between the chassis $\mathscr{P}_c(X)$ and the non-chassis in the power $\mathscr{P}(X)$. Previously, the element $\mathscr{P}_c(X)$ in the chassis was a single point set in X, and the single point set $\boldsymbol{x}=\{x\}$ and single point x were both separated by weight by the same letter. Now, remove all parenthesis restrictions in the topological space definition for neighborhood system construction.

Definition 3.10 (Topological space constructed by neighborhood system) Let $\boldsymbol{x}\in\mathscr{P}(S)$,then $N(\boldsymbol{x})\subseteq\mathscr{P}(S)$, called neighborhood system of point x or single point set $\boldsymbol{x}=\{x\}$.Note $\mathscr{N}=\{N(\boldsymbol{x})|x\in S\}$, called a topological space on S, when

N0 $\boldsymbol{y}\in N(\boldsymbol{x})\Rightarrow\boldsymbol{y}\geq\boldsymbol{x}$;

N1 (Directed) : $(N(\boldsymbol{x}), \geq)$ is the lower directed set;

N2 (Full) : $\boldsymbol{y}\in N(\boldsymbol{x})$ and $\boldsymbol{z}\geq\boldsymbol{y}\Rightarrow\boldsymbol{z}\in N(\boldsymbol{x})$;

N3(Consistency): $\boldsymbol{y}\in N(\boldsymbol{x}) \Rightarrow (\exists\boldsymbol{z})(\boldsymbol{y}\geq\boldsymbol{z}\geq\boldsymbol{x})(((\forall\boldsymbol{x}')(\boldsymbol{z}\geq\boldsymbol{x}')) \Rightarrow \boldsymbol{z}\in N(\boldsymbol{x}'))$.

The N0 added to the definition is a simplification of the chassis settings in the original definition. Whether $\boldsymbol{x}$ is a single point set or not, it has a neighborhood $N(\boldsymbol{x})$, on condition that the neighborhood is $\boldsymbol{y}\geq\boldsymbol{x}$. The other three properties satisfy all $\boldsymbol{x}\in\mathscr{P}(S)$ ($\boldsymbol{x}$ is any subset of S, whether or not it is a single point set). With these changes, all topology definitions in the previous section remain the same, it proves that topology theory over the real field can be transplanted to the general set power $\mathscr{P}(S)$.

When S is a finite set of n elements, the power $\mathscr{P}(S)$ contains 2^n elements. Although the limit process of infinity cannot be described, there are still important applications in mathematical logic. Any kind of group of elements can be chosen to produce a quasi-base $\mathscr{B}$, from which topological $\mathscr{C}$ is generated. All the combinations of elements can also be defined as open sets, i.e. $\mathscr{C}=\mathscr{P}(S)$. Now, the remainder of all of S open sets are still open sets, and the remainder of an open set is called a closed set, so an open set is a closed set, and a closed set is an open set. Both nonintersecting subsets in S have two nonintersecting open cover sets (that is, themselves) that separate them, so S becomes a Hausdorff space.

3.5 Hyper-topology

Lattice topology expresses ordinary topology in the language of a lattice. Change point x in set S into single point set $\boldsymbol{x}=\{x\}$ and change the neighborhood $\boldsymbol{y}$ of point x into neighborhood of single point set $\boldsymbol{x}$, extending the neighborhood of a single point set to the neighborhood $\boldsymbol{y}$ of any set $\boldsymbol{z}$. The result of this aims to change the two-layer structure into a single-layer structure. The so-called two-layer structure means that the point and the neighborhood are elements in two different levels. The point x is a member of the set S, the neighborhood Y of x is a member of the power, the set S is the bottom, the power $\mathscr{P}(S)$ is the upper layer. The point x and its neighborhood $\boldsymbol{y}$ are the elements of the upper and lower layers, so a common topology is considered a two-layer structure because it involves two layers. The lattice topology is considered as a monolayer structure because the set $\boldsymbol{x}$ and its neighborhood $\boldsymbol{y}$ are both elements in power. The advantage of lattice topology is to change the double layer to the single layer. This brings great convenience to hyper-topology. A hyper-topology is a topology of powers. Off-the-shelf hyper-topologies are a tedious task. With a lattice topology, a hyper-topology is a common topology on a lattice topology.

When S is an infinite set, $\boldsymbol{x}\uparrow$ cannot be defined as an open set, because $\{\boldsymbol{x}\}$ cannot be an inner point of $\boldsymbol{x}\uparrow$ in the sense of order $\geq$. But if you redefine the intersection operation of the set in $\mathscr{P}(S)$: the intersection of the two empty lower ejector bullets $\boldsymbol{x}\uparrow\backslash\{\boldsymbol{x}\}$ and $\boldsymbol{y}\uparrow\backslash\{\boldsymbol{y}\}$, with x and y as the lower ejector, is a real lower ejector $(\boldsymbol{x}\vee\boldsymbol{y})\uparrow$, now change it to empty lower ejector $(\boldsymbol{x}\vee\boldsymbol{y})\uparrow\backslash\{\boldsymbol{x}\vee\boldsymbol{y}\}$. But you can define all empty bullets $\boldsymbol{x}^{o}\uparrow$ as open sets. Because $\boldsymbol{x}^{o}\uparrow$ is a directed set, for any point $\boldsymbol{y}$, there is $\boldsymbol{z}\in\boldsymbol{x}^{o}\uparrow$ so there is $\boldsymbol{y}\geq\boldsymbol{z}$, so there must be $\boldsymbol{y}\geq\boldsymbol{z}\geq\boldsymbol{x}$ and $\boldsymbol{z}\neq\boldsymbol{x}$, which means $\boldsymbol{y}\in\boldsymbol{z}^{o}\uparrow\subseteq\boldsymbol{x}^{o}\uparrow$. So $\boldsymbol{y}$ is the inner point of $\boldsymbol{x}^{o}\uparrow$. So, so a blank bullet can be defined as an open set.

Note that the order relation in $\mathscr{P}(S)$ is $\boldsymbol{x}\geq\boldsymbol{y}$ if and only if their set relation in S is $\boldsymbol{x}\supseteq\boldsymbol{y}$. Denote

$$\mathscr{P}(S)\uparrow=\{\boldsymbol{x}\uparrow|\boldsymbol{x}\in\mathscr{P}(S)\},\ \mathscr{P}(S)^{o}\uparrow=\{\boldsymbol{x}^{o}\uparrow|\boldsymbol{x}\in\mathscr{P}(S)\}.$$

Consider $\mathscr{P}^2(S) = \mathscr{P}[\mathscr{P}(S)]=\{\{A\}|A\subseteq S\}$. In the definition of ascending to the power of the upper segment structure, $A \gg B$ in $\mathscr{P}^2(S)$ if and only for any $\boldsymbol{x}\in A$ must have $\boldsymbol{y}\in B$ to $\boldsymbol{x}\geq\boldsymbol{y}$.

Notice that both $\mathscr{P}(S)\uparrow$ and $\mathscr{P}(S)^{o}\uparrow$ are subsets of $\mathscr{P}^2(S)$. The sequence relationship $\gg$ is clearly defined in both $\mathscr{P}(S)\uparrow$ and $\mathscr{P}(S)^{o}\uparrow$. In this order, it is not difficult to prove that in $(\mathscr{P}(S)\uparrow,\gg)$, the deterministic lattice operation $\boldsymbol{x}\uparrow\wedge\ \boldsymbol{y}\uparrow=(\boldsymbol{x}\wedge\boldsymbol{y})\uparrow$. The lattice operation $\boldsymbol{x}^{o}\uparrow\wedge\ \boldsymbol{y}^{o}\uparrow=(\boldsymbol{x}\wedge\boldsymbol{y})^{o}\uparrow$ is determined in $(\mathscr{P}(S)^{o}\uparrow,\gg)$. Thus, $\mathscr{P}(S)^{o}\uparrow$ forms an open set basis in $\mathscr{P}^2(S)$, from which a hyper-topology is determined.

3.6 Existence and uniqueness for correspondence between heaven and earth theorem

From the aforementioned hyper-topology, a variety of hyper-measurable structures are generated, which form a variety of random sets in the sky, and then fall to the ground in different ways to form a variety of subjective non-additive measures. There are 4 popular methods: Belief, Plausibility, Anti-belief and Anti-plausibility. The definition of them are complicated, but it is very simple to define them with the random set fall shadow theory:

Suppose H is a σ-field defined on 2^D .For any subset A of D, let

$$A_{o}=\{B|B\subseteq A\},\ A^{o}=\{B|A\subseteq B\},\ A_{o}^{c}=\{B|B\notin A_{o}\},\ A^{oc}=\{B|B\notin A^{o}\}$$

Where A_{o} and A^{o} are called the ideal and filter of A respectively.

Definition 3.11 Suppose p is the probability of H, let

$$\mu_{BL}(A)=p(A^{o}),\ \mu_{PL}(A)=p(A_{o}^{c}),$$

$$\mu_{APL}(A)=p(A_{o}),\ \mu_{ABL}(A)=p(A^{oc}),\quad (A\in 2^{U}) \quad (4)$$

They respectively are called trust, likelihood, anti-likelihood, and anti-trust measures on 2^U.

It is not difficult to verify that they are no longer additive with probability. They are called nonadditive measures and are the specialty of subjective measures. The fuzzy measure is also included as an anti-likelihood measure.

Fuzzy fall shadow theory not only gives a concise definition of four non-additive measures, but also proves the existence of a unique probability distribution in the sky for each measure to realize the realization of such measures.

Theorem 3.1(Existence and uniqueness theorem for correspondence between heaven and earth) For any nonadditive measure μ (belonging to BL, PL, ABL or APL) given a complex definition, there must be only one probability p on H, so that the fall relation (4) can be established.

Without this existence and uniqueness theorem, both the fuzzy set theory and Dempster-Shafer's evidence theory lose their solid foundation in practical application. It is tough to prove this theorem. The starting point of extension in the measure extension theorem should be shifted from semi-ring to π-system.

IV. Conclusion

Factor space is an adaptive theory which has had an important contribution to fuzzy set and system. Firstly, this paper describes the relationship between modern probability theory and factor space. When factors are used to describe probability, the basic space in modern probability theory is the factor space, and the symmetry criterion, data parity criterion, and the maternal distribution of factor probability are given.

Then, it deeply explores the relationship between fuzzy sets and factor space, points out the underdeveloped part of fuzzy mathematics system, and gives the conversion between membership curve and probability distribution function theorem.

Finally, the mathematical structure is promoted to power set and hyper-topology. Define the lattice on the power set and give its satisfied properties. Introduce the performance of lattice order structure in power set and the topological structure on the power lattice. The explanation of the hyper-topology by the lattice topology theory are further given. Meanwhile, given the existence and uniqueness for correspondence between heaven and earth theorem by fuzzy set fall shadow theory, the four subjective non-additive measures are unified.

Acknowledgment

Special thanks and respects to Professor P. Z. Wang for his guidance and suggestions for this paper. And many thanks to Prof. H. X. Li and Prof. Y. Shi for their supports. This work was supported by the grant (Grant Nos. 61350003) from the Natural science Foundation of China, and the grants (Grant Nos. L2014133, LJ2019JL019) from the department of education of Liaoning Province.

References

[1] P Z Wang, Sugeno M. Factor field and the background structure of fuzzy sets[J], Fuzzy Mathematics, 1982 (2): 45-54

[2] P Z Wang, Y C Chen, B T Low, Measure extension from meet-system and falling measure representation, in Analysis and Management of Fuzzy Data, New York,2002,145-159

[3] P Z Wang, Fuzzy Sets and Falling Shadow of Random Sets, Beijing, Beijing Normal University Press, 1985.

[4] P.Z Wang., Rules detecting and rules-data mutual enhancement based on factors space theory. Inter J Information Technology & Decision Making, Vol.1,No.1(2002)73-90.

[5] P.Z Wang. A factor space approach to knowledge representation. Fuzzy Sets and Systems, 36(1990)113-124.

[6] A Kendel, X.T Peng., Z.Q Cao., P.Z Wang., Representation of concepts by factor spaces. Cybernet. Systems, 21(1)(1990)43-37.

[7] P Z Wang, H X Li, The Mthematica Theory of Knowledge Representation, Tianjin: Tianjin Scince and Technology Press, 1994.

[8] P Z Wang, H X Li, Fuzzy System Theory and Fuzzy Computers, Beijing: Science Press, 1995.

[9] Manuel Arellano and Stephen Bond. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations[J]. Review of Economic Studies, 58(2):277-297.

[10] Zadeh L A . Fuzzy sets as a basis for a theory of possibility[J]. Fuzzy Sets and Systems, 1978, 1(1):3-28.

[11] Lu C G. Channels’ Matching Algorithm for Mixture Models[C]// International Conference on Intelligence Science. Springer, Cham, 2017.