Journal of
Mathematics and
Informatics

A Linear Method Based on Retrieval to Estimate Missing Fuzzy-Number in a Fuzzy Set and its Application

Hang Xiao[1] and *Sidong Xian[1,2]*

[1]School of Science, Chongqing University of Posts and Telecommunications Chongqing
Chongqing,400065 P.R. China. e-mail: 970334034@qq.coom
[2]Key Laboratory of Intelligent Analysis and Decision on Complex Systems
Chongqing University of Posts and Telecommunications
Chongqing, 400065, P.R.China. e-mail: xiansd@cqupt.edu.cn

Abstract. This paper studied about the problem of estimation of fuzzy number. Fuzzy numbers can fill the essence of things or procedures more accurately than real numbers. And it has a better efficiency without too much calculation cost. So it is widely applied in industries and many science field. However, data missing and data skew are big trouble caused by some uncontrolled factors. Considering this problem and focusing on how to make up that emergency immediately at less computational costs, we proposed a linear method based on retrieval to figure out this kind of problems. It's a universal method, not only suitable to solve the fuzzy data problem, but also for real number. It is powerful and efficient with both linear system and non-linear system without too much calculation.

Keywords: fuzzy set; estimation of fuzzy number; data missing; linear method; universal method

AMS Mathematics Subject Classification (2010): 94D05

1. Introduction
In the real world, uncertainty is ubiquitous. Much of the information which decision making usually based on is fuzzy. Although human has a significant capacity to deal and make rational and logical decisions based on these uncertain information [1], human's computing power is not a patch on computers in terms of whether imprecise accuracy or processing speed. Since any non-linear system can be transformed into fuzzy set according to fuzzy set theory [2], fuzzy models performs well for uncertainty systems and based on fuzzy rules. In fact, fuzzy models have been used in many areas because of its advantages to extract full information of things or any procedures much better than real number (both linear or nonlinear system, since any nonlinear system can be approximated by several piecewise linear systems [3]), such as operation research [4], decision science [5], physic [6,7,8] and many other branches of social science and hard science.

However, unforeseen emergency like data missing or data skew can be caused by some unpredictable factors, like equipment failure, workers mistake and power failure. Causing a huge loss if we cannot get it right as fast as possible. There are some studies about this, like linear programming (LP), it is often considered to solve this kind of

problems. The concept of fuzzy LP was first proposed by Zimmerman[9]. As the well-known operation research technique frequently used by researchers and practitioners in the fields of science, economics, business, management science and engineering. In classical LP models, a number of parameters is needed to be fixed in the constraint condition equation and objective function (depended on the result that people expected). Interval analysis and type-1 and type-2 fuzzy set theories have been also applied to formulating of LP models. And corresponding methods to capture uncertainty of an investigated real-world problem have been developed [10]. An interval arithmetic-based LP approach is considered in [11,12]. The first research about taking into account real-world soft constraints in the LP problem by using fuzzy sets was suggested in the famous work by Zadeh and Bellman [13]. Some methods base on the Fuzzy Bayes models give a more accurate result from the probability theory. A few Certain methods are very successful in the fields of artificial intelligence and decision making. Such as fuzzy along with the probability method based on Bayesian model, that mainly concerns on the reasoning with uncertain data. Bayes models are most popular for the probability distribution, that usually fits the information of the data in the conditional probability tables, which acts as a subjective probability.

However, all these methods above need to divided problems to the independent part and the connection (or affection) part, then to figure out two parts respectively. This procedures-dividing problems into two parts and figuring them out, would cause some information losing. What's more, we need to analyze the differences of different system as detailed as possible, It's a big work with much computer calculation costs. In the era of big data, little reduce on the quantity of calculation in one step, would bring a great energy efficiency and working efficiency improvement, less time cost on calculation. Consequently, more economic benefits come.

Why not use the information contained in data directly? Our work proposes a new method to making an estimation or prediction about the fuzzy data in a procedure. Generally, Historical data from productive processes or others include the normal situation of every link in a procedure (if we have record it). No matter linear systems or non-linear system, a certain situation of a link is influenced by last one or even far-off ones, and it might influence the next one and after that. Also, the value of every datum is influenced by self-factors of system and other external random factors. So every set of data from a recording reflect the current integer situation. All the possibility situation about the machines, workers, and other random factors is contained in the recorded data. Generally, production runs normally in most time, which is tenable according to our experiences. We don't need to know the certain relationship between any two links of all. The inspiration of this new method based on retrieval is to find the most similar situation in recording while a contingency happened. we put all the data in a matrix, and each event m_{ij} in this matrix contains the connection of f_i and f_j. If we got the data of most f_i, we can find the current situation in recording. Then it will not be a problem to get an estimation value of a losing data f_k through a certain method. While we got all the F = $\{f_j, j = 1 \dots n, j \neq k\}$, and we need to construct a matrix M = $\{e_{ij}, i = 1,2 \dots n, j = 1,2 \dots n\}$ to extract the connections of the links related and the attribution information of each f_i. Also, the final result still need a reliability test by statistical approaches.

A Linear Method Based on Retrieval to Estimate Missing Fuzzy-Number in a Fuzzy Set and its Application

2. Preliminaries

Here, we are going to introduce some basic knowledge which we might need. Some of them come the studies by predecessors, basic conceptions of fuzzy number were defined and developed by Zadeh and other mathematicians.

2.1. Fuzzy numbers

Fuzzy number was firstly proposed by Zadeh, that can fetch more information from the nature than real number and formalizing the ability to make fuzzy decision making by human to some degree. It has been a hit since it was born. There are varies kinds of fuzzy number, they have different features and fit for dissimilar conditions. Since the triangular fuzzy numbers is most widely-used, this paper just introduces the basic definition about it.

Definition 2.1. A triangular fuzzy number f is defined by a triplet, $f = (a, b, c)$, $a, b, c \in \Re$. The left value 'a' and 'c' are the points which minimize the degree of membership. The middle value 'b' is this point that maximizes the degree of membership.

Example 2.1. A triangular fuzzy number $f = (1,5,6)$, the maximal degree of membership is obtained at the middle point whose value is 5 and the degree of membership is 1. Similarly, at the left point and the right one, the degree of membership is minimal to 0. The certain degree of membership of one point is depended on the membership function.

2.2. Approximation operators

After getting the final result, we need to test the reliability of result. This need to transfer fuzzy number to real number. There are some mature methods to get there. we might need the following algorithm.

One of the most widely used methods. For example, in risk analysis [14], is the so-called.

① Center of gravity[15]: $\quad COG(A) = \frac{\int xA(x)dx}{\int A(x)dx}$; $\hspace{3cm}$ (1)

② The Possibility mean by Goetschel and Voxmanand[16] and later studied by Carlsson and Fuller[17]:

$$E(A) = \int_0^1 \alpha(A_\alpha^+ + A_\alpha^-)\,d\alpha \hspace{3cm} (2)$$

③ In the R&D project portfolio selection[18], Amado[19] defined the extremely simple method:

$$AD_\alpha(a) = A_\alpha^+ \hspace{3cm} (3)$$

④ Yager's method [20]: $\quad Y(A) = \frac{1}{2}\int_0^1 (A_\alpha^+ + A_\alpha^-)\,d\alpha \hspace{2cm} (4)$

These methods are not just suitable to approximate triangular fuzzy number, but also other kinds. Science triangular fuzzy number is most widely applied, we just take this kind as instance.

Example 2. For a triangular fuzzy number A=(a_1, a_2, a_3),

$COG(A) = \frac{a_1+a_2+a_3}{3}$; $\quad E(A) = \frac{a_1+4a_2+a_3}{6}$; $\quad AD_{0.5}(A) = \frac{a_2+a_3}{2}$; $\quad Y(A) = \frac{a_1+2a_2+a_3}{4}$

3. Basic concepts and algorithm

This section will introduce numerical concepts which may be involved in this method, according to the fundamental theory of probability theory. Also, the algorithm of this methods will be introduced.

3.1. Basic concepts and theorem

This section would introduce some concepts involved in this method, some of them may notapplied in real practice. But it is necessary to describe and understand the organizing form of sample data. For convenience, we defined a new multiple which is needed in last step.

Definition 3.1. Every element $\frac{\partial}{\partial f_j} f_i$ represent the relationship of f_i and f_j. It's just a visual form of the relationships between these data, we don't need to do any calculation. m is the times of observation frequency, so the number of abstract reference matrix is k. Each $f_i = (a_i, b_i, c_i)$. So this matrix is like this:

$$
M_k = \begin{vmatrix}
\frac{\partial}{\partial f_1} f_1 & \frac{\partial}{\partial f_2} f_1 & \cdots & \frac{\partial}{\partial f_n} f_1 \\
\frac{\partial}{\partial f_1} f_2 & \frac{\partial}{\partial f_2} f_2 & \cdots & \frac{\partial}{\partial f_n} f_2 \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial}{\partial f_1} f_n & \frac{\partial}{\partial f_2} f_n & \cdots & \frac{\partial}{\partial f_n} f_n
\end{vmatrix}, k = 1,2 \ldots m
$$

For convenience, denoted e_{ijk} is the event e_{ij} in the k_{th} matrix M_k, f_{ik} is the value of k_{th} observation of f_{ik}. $f_{ik} = (a_{ik}, b_{ik}, c_{ik})$, a_{ik} is the value of a_i observed in the k_{th} observation. So, the matrix M_k can be expressed by:

$$
M_k = \left\{ \frac{\partial}{\partial f_{jk}} f_{ik} \right\} = \{e_{ijk}\}
$$

According to the definition, all the element in k_{th} row of the matrix includes the relationship between f^k and all other indicators. If we know all the data about others. We can calculate the discreet value. A convenience is that data can be easily organized in computer and dealt with a simple data structure, which means a lot of work saved.

Then we shall to making estimation, for a certain event in the matrix,

$$
e_{ijk} = \left(\frac{f_{i1}}{f_{j1}}, \frac{f_{i2}}{f_{j2}} \ldots \frac{f_{ik}}{f_{jk}} \ldots \frac{f_{im}}{f_{jm}} \right), \quad f_{j1} \leq \cdots \leq f_{jk} \leq \cdots \leq f_{jm}).
$$

Definition 3.2. To define a new multiplication denoted by *. Giving a real number x from the sample population about f^j, which is $N(u_j, \sigma_j^2)$. And $f^i \sim N(u_i, \sigma_i^2)$

$$
x * e_{ijk} = \begin{cases}
f_{ik}, & if \ x = f_{jk} \\
f_{ik} + (x - f_{jk}) * \frac{(f_{i(k+1)} - f_{ik})}{f_{j(k+1)} - f_{jk}}, & if \ f_{jk} < x < f_{j(k+1)}
\end{cases} \tag{8}
$$

Examples 3.1. If $x_1 = 5$, p $= (\frac{5}{2}, \frac{9}{3}, \frac{12}{4}, \frac{16}{5}, \frac{20}{6}, \frac{25}{7})$, then x $* p = 16$

If $x_1 = 5.5$, p $= (\frac{5}{2}, \frac{9}{3}, \frac{12}{4}, \frac{16}{5}, \frac{20}{6}, \frac{25}{7})$, then x $* p = 16 + (20 - 16) * \frac{(5.5 - 5)}{6 - 5} = 18$

A Linear Method Based on Retrieval to Estimate Missing Fuzzy-Number in a Fuzzy Set
and its Application

Also, for a real number vector $X = (x_1, x_2, \dots, x_n)$,
$$X * e_{ijk} = \sum_{l=1}^{n} x_l * e_{ijk} \tag{9}$$
Obviously, this is a linear operation.

Theorem 1. The result is still obeying the same kind distribution, if the sample population obeying normal distribution.

Proof: For example, supposing f^i obeying normal distribution $N(u_i, \sigma_i^2)$. According to the calculation formula about $x * e_{ijk}$. The result is nothing else than these two kinds, f_{ik} or $f_{ik} + (x - f_{jk}) * \frac{(f_{i(k+1)} - f_{ik})}{f_{j(k+1)} - f_{jk}}$. The former one, obviously it belongs to sample population about f^i. For the latter one, since

$$\begin{cases} 0 < \dfrac{(x - f_{jk})}{f_{j(k+1)} - f_{jk}} = k < 1 \\ (f_{i(k+1)} - f_{ik}) \sim N(0, 2\sigma_i^2) \end{cases}$$

So the result $x * e_{ijk} \sim N(u_i, \sigma_i^2 + 2k\sigma_i^2)$, the average value is still unchanged. But just the variance have a little expanding. We can still get to know the new variance: $\sigma_{i\_new}^2 = \sigma_i^2 + 2k\sigma_i^2$. So $x * e_{ijk} \sim N(u_i, \sigma_{i\_new}^2)$

Theorem 2. To make a transformation to standard normal distribution: $x' = \dfrac{x - u_i}{\sqrt{3\sigma_i^2}}$, since standard normal distribution is $x = \dfrac{x - u_i}{\sigma_i}$, according to the $''3\sigma''$ principle, while $|x| \leq 2.78$, the confidence level $1 - \alpha \geq 99.73\%$, α is significance level. Then, the x is acceptable, Meanwhile,
$$|x'| = |\sqrt{3}x| \leq \sqrt{3} * 2.78 = 4.8151$$
Also, if x' satisfies this condition, we should accept it as a reliable value.

3.2. Algorithm of this method

This part would introduce the steps of algorithm in appliance, according to the property1.

Step 1. As we need to use the information from historical data. So the firststep is to get primary data, denoted by $F = [f_1, f_2, \dots, f_n]$, f_i is a link in this procedure, and its value is $f_i = [a_i, b_i, c_i]$, $i = 1, 2 \dots n$ If each f_i has been observed m times. Denoted that $f_{ik} = [a_{ik}, b_{ik}, c_{ik}]$, $k = 1, 2, \dots m$.

Step 2. The data we have get is 'original'. They have many skew data from some extremity. So we should make some preprocess. Filtering some bad original data, according to conventional statistical theory. If a fuzzy sample $f_{ik} = [a_{ik}, b_{ik}, c_{ik}]$, satisfies any following requirement, we will delete it.

 A. $a_{ik} \notin [Q_{i1} - 1.5R_{i1}, Q_{i3} + 1.5R_{i1}]$, $R_{i1} = Q_{i3} - Q_{i1}$ (10)
    Q_3 is the upper quarterback point of $\{a_{ik}\}$, $k = 1, 2 \dots m$, Q_{i1} is the lower quarterback point of $\{a_{ik}\}$.
 B. $b_{ik} \notin [Q_{i1} - 1.5R_{i1}, Q_{i3} + 1.5R_{i1}]$, $R_{i1} = Q_{i3} - Q_{i1}$ (11)

Q_3 is the upper quarterback point of $\{b_{ik}\}, k = 1,2 \dots$ m , Q_{i1} is the lower quarterback point of $\{b_{ik}\}$.

C. $c_{ik} \notin [Q_{i1} - 1.5R_{i1}, Q_{i3} + 1.5R_{i1}], \quad R_{i1} = Q_{i3} - Q_{i1}$ (12)

Q_3 is the upper quarterback point of $\{c_{ik}\}, k = 1,2 \dots$ m , Q_{i1} is the lower quarterback point of $\{c_{ik}\}$.

f_{ik} is an outlier, if it satisfies any one condition, which will be deleted. After this step, if the number of sample data is not the same, we should delete the redundant data of a indicator to make sure the same number of data sample.

Example 3.2.1. If the numbers of data of indicators are x, y, z, and $x > y > z$. Then, the amount of redundant data for indicator f^x is $(x - z)$, we should delete $(x - z)$ data from f^x randomly. And so on, delete $(y - z)$ data from f^y.

Step 3. Supposing we still have m times observation value after last step.

$$a_{ij} = (\frac{a_{i1}}{a_{j1}}, \frac{a_{i2}}{a_{j2}} \dots \frac{a_{im}}{a_{jm}})$$

$$b_{ij} = (\frac{b_{i1}}{b_{j1}}, \frac{b_{i2}}{b_{j2}} \dots \frac{b_{im}}{b_{jm}})$$

$$c_{ij} = (\frac{c_{i1}}{c_{j1}}, \frac{c_{i2}}{c_{j2}} \dots \frac{c_{im}}{c_{jm}})$$

We can construct such a reference matrix $M = \{e_{ij} | e_{ij} = (a'_{ij}, b'_{ij}, c'_{ij})\}$.Then we reorder the a_{ij} according the value of a_{jk}, so we get:

$$a'_{ij} = \left(\frac{a_{i1}}{a_{j1}}, \frac{a_{i2}}{a_{j2}} \dots \frac{a_{im}}{a_{jm}}\right) a_{j1} < a_{j2} < \cdots < a_{jm}$$

$$b'_{ij} = \left(\frac{b_{i1}}{b_{j1}}, \frac{b_{i2}}{b_{j2}} \dots \frac{b_{im}}{b_{jm}}\right) b_{j1} < b_{j2} < \cdots < b_{jm}$$

$$c'_{ij} = \left(\frac{c_{i1}}{c_{j1}}, \frac{c_{i2}}{c_{j2}} \dots \frac{c_{im}}{c_{jm}}\right) c_{j1} < c_{j2} < \cdots < c_{jm}$$

Step 4. the matrix includes all the relationship between all indicators, while the data of ith is missing, we just need these event, that related with f^i. according to the definition 2, we just need the data in ithrow. So constructing a n-order matrix

$$E_i = \begin{vmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 1 & \dots & 0 \end{vmatrix}$$

just the events on column i are 1, and all other events is 0. i is the sequence number of the lost indicator in F. A default vector is $F = [f_1, f_2 \dots, \overline{f_l}, \dots f_n]$, the event $\overline{f_l}$ is unknown, $\overline{f_l} = (0,0,0)$.

$$E_i * M = (\frac{\partial}{\partial f_1} f_i, \frac{\partial}{\partial f_2} f_i, \dots, \frac{\partial}{\partial f_n} f_i)$$ (13)

We call this a reference vector, denote $V = E_i * M$

Step 5. Using the data we have known in F, $F = [f_1, f_2 \dots, \overline{f_l}, \dots f_n]$ to calculate.

$$\overline{f_l} = \frac{1}{n-1} F * V^T = \frac{1}{n-1} [\sum_{j=1, j\neq i}^{n} f_i * V^T] \tag{14}$$

The $*$ is the multiplication defined in definition 4. According to the last discussion, we can know that $f_i * V^T \sim N(u_i, 3\sigma_i^2)$. So $\frac{1}{n-1} \sum_{j=1, j\neq i}^{n} f_i * V^T \sim N(u_i, 3\sigma_i^2)$. So this result would have the same average value, with three times expansion about variance unfortunately.

4. Application

This section gives an application about this proposed method. Since this method is proposed for this condition in which a fuzzy number is lost in a successive procedure. We will use recorded historical data to make an immediately estimation. So this method involves data obtaining, data preprocessing, abstracting proper data we needed, and calculation. The last step is making a test of the final result.

4.1. Obtaining data sample

For a 3-indicator procedure $F = [f^1, f^2, f^3]$. since every data from a link in a procedure. So we must give those data certain meaning. We generated 100 sample data of each indicator randomly by computer software, according to an average array and a covariance matrix. So we set:

$$obj1 = [1\ 4.5\ 8]^T obj2 = [0\ 3.5\ 8]^T obj3 = [1\ 5\ 9]^T$$

$$sigma1 = \begin{bmatrix} 5 & 1 & 3 \\ 1 & 5 & 2 \\ 3 & 2 & 8 \end{bmatrix} sigma1 = \begin{bmatrix} 4 & 1 & 3 \\ 1 & 7 & 2 \\ 3 & 2 & 7 \end{bmatrix} sigma1 = \begin{bmatrix} 3 & 1 & 3 \\ 1 & 2 & 2 \\ 3 & 2 & 6 \end{bmatrix}$$

The obj1~obj3 is the mean value of $f^1 \sim f^3$ respectively, which gives 'meaning' to the data, just like real data from a procedure. and the sigma1~sigma3 is the variance of matrix of $f^1 \sim f^3$ respectively, representing some uncontrolled random factors in real life. Here is just a part of data (We don't have enough space show them all):

	f^1.a	f^1.b	f^1.c		f^2.a	f^2.b	f^2.c		$f^3.a$	f^3.b	f^3.c
1	-0.045	2.887	5.966		4.048	5.605	8.853		0.478	4.913	7.759
2	0.720	8.407	7.933		-4.719	3.361	5.304		-0.210	4.355	7.479
3	4.307	2.795	14.115		-1.019	3.492	4.788		2.442	5.857	9.602
4	-0.924	4.551	9.261		-2.643	4.129	4.213		-0.203	5.874	9.076
5	2.754	1.518	7.041		-1.272	5.993	6.176		0.199	5.240	12.168
	……	……	……		……	……	……		……	……	……
96	-1.501	2.834	2.243		-1.185	-2.520	9.759		0.338	5.576	9.117
97	-0.400	6.350	8.181		-0.939	1.780	4.553		1.742	6.038	9.574
98	1.557	3.467	8.906		1.772	4.499	14.486		0.481	5.200	9.261
99	-1.220	4.442	6.234		-2.770	5.255	9.619		-0.558	4.990	7.803
100	3.180	7.062	7.163		-3.913	2.765	5.468		2.099	4.242	8.860

4.2. Filtering data

Since all the data is normally distributed, there are just few samples don't meet the specification in step2. For example, the upper quarterback point of each component of f^1 is $(2.323, 5.984, 9.668)$, The lower quarterback point of each component of f^1 is $(-0.543, 3.176, 5.842)$. So according to equation 10~12, the distance between upper quarterback and lower quarterback of each component is

$$\begin{cases} 2.323 - (-0.543) = 2.866 \\ 5.984 - 3.176 = 2.808 \\ 9.668 - 5.842 = 3.826 \end{cases}.$$

the upper critical value is: $\begin{cases} 2.323 + 1.5 * 2.866 = 6.622 \\ 5.984 + 1.5 * 2.808 = 10.196 \\ 9.668 + 1.5 * 3.826 = 15.407 \end{cases}$

also the lower critical value is $\begin{cases} -0.543 - 1.5 * 2.866 = -4.842 \\ 3.176 - 1.5 * 2.808 = -1.036 \\ 5.842 - 1.5 * 3.826 = 0.103 \end{cases}$

if any number in $f^1.a$ is bigger than 6.622 or smaller than -4.84161, then delete this sample, there is only one sample satisfying, so we delete 1 sample. Do the same to $f^1.b$ and $f^1.c$. So we get:

	f^1.a	f^1.b	f^1.c		f^2.a	f^2.b	f^2.c		$f^3.a$	f^3.b	f^3.c
1	-0.045	2.887	5.966		4.048	5.605	8.853		0.478	4.913	7.759
2	0.720	8.407	7.933		-4.719	3.361	5.304		-0.210	4.355	7.479
3	4.307	2.795	14.115		-1.019	3.492	4.788		2.442	5.857	9.602
4	-0.924	4.551	9.261		-2.643	4.129	4.213		-0.203	5.874	9.076
5	2.754	1.518	7.041		-1.272	5.993	6.176		0.199	5.240	12.16
	……	……	……		……	……	……		……	……	……
95	-1.501	2.834	2.243		1.772	4.499	14.48		0.338	5.576	9.117
96	-0.400	6.350	8.181		-2.770	5.255	9.619		1.742	6.038	9.574
97	1.557	3.467	8.906		-3.913	2.765	5.468		0.481	5.200	9.261
98	-1.220	4.442	6.234						-0.558	4.990	7.803
99	3.180	7.062	7.163						2.099	4.242	8.860

So the final number of each component in f^1 is(99,97,99). To guarantee all the indicators have Same number of samples, we should delete 2 samples in $f^1.b$, and 2 samples in $f^1.c$ according to the Step2. By the way, another simulation experiment with 1000 data suggests that there are about 2% normal random data generate from computer meet the filtering rules in Step2. Then we get:

1	-0.045	2.888	5.967		4.049	5.605	8.853		0.479	4.913	7.760
2	0.720	8.408	7.933		-4.720	3.361	5.304		-0.210	4.355	7.480
3	4.307	2.796	14.116		-1.020	3.4921	4.789		2.442	5.857	9.603
4	-0.924	4.551	9.261		-2.643	4.130	4.214		-0.203	5.874	9.076
5	2.754	1.519	7.041		-1.272	5.994	6.177		0.200	5.241	12.169
…	……	……	……		……	……	……		……	……	……
93	-1.501	2.835	2.243		-1.185	-2.521	9.760		0.339	5.576	9.118
94	-0.401	6.351	8.181		-0.940	1.781	4.55		1.743	6.039	9.578

95	1.558	3.468	8.907		1.773	4.500	14.487		0.482	5.201	9.261
96	-1.220	4.443	6.234		-2.770	5.256	9.620		-0.559	4.991	7.804
97	3.180	7.063	7.163		-3.914	2.765	5.469		2.100	4.243	8.860

4.3. Constructing an abstracting matrix

Supposing we lost some data, there is just
$F = [f^1, 0, f^3]$.we can get, and$f^1 = [1 \; 4.5 \; 8]^T$, f^2 is lost and $f^3 = [1 \; 5 \; 9]^T$. the number l is 2, according the step 4, we can construct an extracting matrix $E_l =$

$$\begin{vmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{vmatrix}.$$

4.4. Calculation

According to the equation 14，

$$f^2 = \frac{1}{3-1} * [f^1 * (EM)^T + f^3 * (EM)^T] = (0.982, 4.009, 6.657)$$

This is the final estimation value of f^2.

4.5. Making a hypothesis test

The last step is to check up on the reliability of this result. For convenience, denoting the sample population of f^2 by S, the final result by R. Since this comparing is between a vector and a sample population, and the data are normal random. we consider making this check with one-sample t test method [21].

4.5.1. Making test hypothesis:

$$\begin{cases} \text{The Null Hypothesis: } H_0: \theta = 0 \\ \text{Alternative Hypothesis: } H_1: \theta \neq 0 \end{cases}$$

And we should convert the fuzzy number in sample of f^2 to a real number.

	$f^2.a$	$f^2.b$	$f^2.c$	$f^2.R$		$f^2.a$	$f^2.b$	$f^2.c$	$f^2.R$
1	4.049	5.605	8.853	6.028	⋮	⋮	⋮	⋮	⋮
2	-4.720	3.361	5.304	1.827	90	0.241	6.541	6.537	4.965
3	-1.020	3.492	4.789	2.688	91	-1.980	4.558	6.116	3.313
4	-2.643	4.130	4.214	2.457	92	2.396	0.770	10.810	3.686
5	-1.272	5.994	6.177	4.223	93	-1.185	-2.521	9.760	0.883
6	0.636	6.180	6.330	4.832	94	-0.940	1.781	4.554	1.794
7	0.276	2.092	7.358	2.954	95	1.773	4.500	14.486	6.314
8	-1.421	5.250	5.411	3.622	96	-2.770	5.256	9.620	4.340
⋮	⋮	⋮	⋮	⋮	97	-3.914	2.765	5.468	1.771

And do the same to R, $\bar{R} = \frac{(0.982 + 2*4.009 + 6.657)}{4} = 3.914$, and subtract $\bar{R}$ with $f^2.R$. denote a vector $v = f^2.R - \bar{R} = \{v_i, i = 1,2 \dots 97\}$.

4.5.2. Calculating average value and variance of each component in f^2, according to equation 6~7:

$$\bar{x} = \frac{1}{97}\left(\sum_{i=1}^{97} v_i\right) = -0.478.$$

$$s^2 = \frac{1}{96}\sum_{i=1}^{97}(v_i - \bar{x})^2 = 2.202$$

And the test statistic $\text{t} = \sqrt{97}\frac{\bar{x}}{s} \approx -3.173$

This is a reliable result, since $|\text{t}| = 3.173 \leq 4.815$ according to theory 1.

4.5.3. According to the critical value table, since:

$$|\text{t}| = 3.173 > t_{1-\frac{\partial}{2}}(97) \approx 2.63, \quad \partial = 0.01$$

So we cannot reject the null hypothesis, while the significant level ∂ is 0.01. we can accept this estimation as a reliable result.

5. Conclusion

This method is largely depended on the retrieval with little calculation, just once deal to preprocess original samples and construct the relationship by an abstract matrix. Then, all the operation depends on retrievals and compare, it would not be a big problem to find position of a certain value while we put all data in order. And little calculation is needed in last step. Which can save a lot of computational and time costs. "Invest once, lifelong benefit" is the most prominent property. Generally, this method is more efficient and accurate while dealing with more criteria procedure.

REFERENCES

1. L.A.Zadeh, A Note on Z-numbers, *Information Sciences*, 181 (2011) 2923-2932.
2. L.A.Zadeh, Toward a generalized theory of uncertainty (GTU)-an outline, *Information Sciences,* 172 (2005) 1-40.
3. T.Takagi and M.Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man and Cybernetics,* 15(1) (1985) 116-132.
4. A.Kaufmann, On the relevance of fuzzy sets for operations research, *Eur. J. Oper. Res,* 25(3) (1986) 330-335.
5. D.Dubois, The role of fuzzy sets in decisions in science: old techniques and new directions, *Fuzzy Sets Syst,* 184(1) (2011) 3-28.
6. Venkatesh A, Manikandan R, Fuzzy Mathematical Modeling for The Effect of Oxytocin using ENH Distribution, *Aryabhatta Journal of Mathematics and Informatics,* 8 (2016) 89-96.
7. A.Venkatesh, S.Elango, A mathematical model for the treatment of thyrotropin releasing hormone in humans using fuzzy approach, *Journal of Mathematics and Informatics*, 8 (2016) 75-80.
8. A.Venkatesh, K.P.Senthil, Fuzzy threshold shock model for the effect of corticosterone, *Journal of Mathematics and Informatics*, 8 (2016) 23-30.
9. H.J.Zimmermann, Fuzzy programming and linear programming with several objective functions, *Fuzzy Set Syst*, 1 (1978) 45-55.
10. T.Jiafu, W.Dingwei, R.Y.K.Fung and K.L.Yung, Understanding of fuzzy optimization: theories and methods, *J SystSci Complex,* 17 (2004) 117-136.
11. M.Allahdali, H.Mishmast Nehi, The optimal solution set of the interval linear programming problems, *Optim Lett.*, 7 (2013) 1893-1911.

12. M.Fiedler, J.Nedoma, J.Ramik and K.Zimmermann, Linear optimization problems with inexact data, New York: Springer Science Business Media; 2006.
13. R.E.Bellman and L.A.Zadeh, Decision making in a fuzzy environment, *Manage Sci,* 17 (1970) 141-164.
14. C.Huang, An application of calculated fuzzy risk, *Inf. Sci.,* 142(1) (2002) 37-56.
15. B.Jakovljević Ivan, D.Mirosavljević Petar, M.Mijailović Radomir, Optimization of aircraft center of gravity, *Tehnika,* 72(1) (2017) 411-414
16. R.Goetschel and W.Voxman, Elementary fuzzy calculus, *Fuzzy Sets Syst.,* 18(1) (1986) 31–43.
17. R.Carlsson, Fullér, Onpossibilistic mean value and variance of fuzzy numbers, *Fuzzy Sets Syst*, 122 (2) (2001) 315–326.
18. C.Carlsson, R Fullér, M.Heikkila and P. Majlender, A fuzzy approaches to R&D project portfolio selection, *Int. J. Approx. Reason*, 44(2) (2007) 93-105.
19. J.Adamo, Fuzzy decision trees, *Fuzzy Sets Styst.,* 4(3) (1980) 207-219.
20. R.R.Yager, A procedure for ordering fuzzy subsets of the unit interval, *Inf. Sci.,* 24(2) (1981) 207-215
21. G.Cicchitelli, On the robustness of the one sample *t* test, *Journal of Statistical Computation and Simulation,* 32 (1989) 249-258.