Selecting Visualization Alternative Based on Uncertain Theory

Ping Huang1,*, Zhiying Lv1,2 and Liwei Zheng3

1College of Management, Chengdu University of Information Technology, Chengdu 610103,China
2College of Mathematics, University of Electronic Science and Technology of China, Chengdu 610103,China
3College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610103,China
Corresponding Email: 229059014@qq.com

Abstract. Multiple attribute decision making (MADM) is an efficient way to solve complex systems, and has wide application. This research develops MCDM model based on uncertain theory, used for selecting a suitable visualization alternative for tourism. First, in order to achieve desirable decision making, a new concept is proposed, which is called the best and the worst reference uncertain linguistic variable as a datum uncertain linguistic variable. At the same time, a new method for ranking uncertain linguistic variable is also presented. Second, based on the preference order relation of attributes given by the experts, a new score function is introduced to get the weight vector of attributes. Finally, the evaluation system of tourism big data visualization alternatives is constructed and the order of those alternatives is acquired by the decision method.

Keywords. MADM, score function, reference uncertain linguistic variable, tourism big data visualization.

1 Introduction

Big data [1] has been risen as national strategic resources. In recent years, after many developed countries such as Europe, the United States, Japan and South Korea took large data as a national level strategy, China have brought the construction of large data into the national strategic choice. Large datas was written in the government work report in 2014. In January 2015, the National Tourism Administration in the "Guidance on the promotion of the development of intellectual tourism," said: "By 2020, large data mining and intelligent marketing capabilities will be improved significantly, mobile e-commerce, tourism data system analysis, artificial intelligence technology will be applied in the tourism industry more widely. And the main task of building tourism data has been clear many times in the views.

Tourism big data visualization [2] decision-making is an emerging field of the internet age. In the internet context, tourism big data has a large scale and has a high complexity. It is a huge challenge to search, analysis and understand the large number of unstructured data and semi structured data. The value of the rich information behind the hidden is reflected only through the collection, analysis, interpretation and expression. And visualization is the most effective way for people to understand easily the value of complex information data. Therefore, it is significant to study the decision making of tourism big data visualization. It can not only promote the fusion and innovation of data mining, analysis techniques and methods, computer graphics technology and decision theory and methods, but also have a

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).
transformational influence on thinking and methods of government departments, tourism enterprises and tourists. Therefore, it can provide a more rapid, effective and scientific decision-making protection.

In the process of decision, there are some difficulties for experts to express their preference degrees with crisp numerical values. So, it is another possible way to use linguistic labels [3], which represents qualitative aspects values. We will consider a finite and totally ordered label set \( S = \{s_1, s_2, \ldots, s_n\} \) in the usual sense and with odd cardinality, where each label \( s_i \) represents a possible value for a linguistic real variable. The number of linguistic terms in the set \( S \) is called the cardinality of \( S \). In the symbolic computation process, the discrete linguistic set \( S \) is extended to a continuous interval \( \tilde{S} = \{s_\alpha : \alpha \in [0,1]\} \). \( \tilde{S} \) is called an extended linguistic term set associated with \( S \). Let \( \tilde{s} \in \tilde{S} \), \( I(s) \) is denoted as the position index of \( s \) and called the gradation of \( s \) in \( \tilde{S} \). For example, \( I(s_\alpha) = \alpha \). The 2-tuple fuzzy linguistic representation model represents the linguistic information by means of a 2-tuple \( (s, \alpha) \), where \( s \) is a linguistic label and the numerical value and \( \alpha \) represents the value of the symbolic translation [4, 5].

In this paper, the focus is on the extension of discrete support model for MADM [6, 7] in which the experts express their opinions by means of uncertain linguistic setting instead of precise numerical values. In order to get the weight vector, a new method by the definition of score function of attribute and the preference order of attribute presented by different experts is presented. Moreover, a new ranking method is presented to rank uncertain linguistic variable, which is based on the best and the worst reference uncertain linguistic variable.

## 2 Preliminaries

In this section, the basic concepts and their extensions of the this paper are briefly introduced.

**Definition 1** Let \( \tilde{s} = [s^-_a, s^+_a] \), where \( s^-_a, s^+_a \in S \), and \( s^-_a \) and \( s^+_a \) are lower and upper limits, respectively. Then \( \tilde{s} \) is called an uncertain linguistic variable.

**Definition 2** Let \( \tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n \) be a set of uncertain linguistic variables, where \( \tilde{s}_n = [s^-_a, s^+_a] \in \tilde{S} \). The associated weight vector is \( \omega = (\omega_1, \omega_2, \ldots, \omega_n) \), where \( \omega_i \geq 0 \) and \( \sum_{i=1}^{n} \omega_i = 1 \). The uncertain linguistic weighted average (ULWA) operator comes from the extended linguistic representation model is defined as:

\[
ULWA(\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n) = \tilde{s}_u = [s^-_a, s^+_a]
\]

where

\[
a = [a^-, a^+]
\]

\[
= [\sum_{i=1}^{n} \omega_i I(s^-_a), \sum_{i=1}^{n} \omega_i I(s^+_a)]
\]

\[
= [\sum_{i=1}^{n} \omega_i a^-_i, \sum_{i=1}^{n} \omega_i a^+_i]
\]

The deviation measures have been discussed between two uncertain linguistic terms. The following deviation measure for two uncertain linguistic variables is first introduced.

**Definition 3** Let \( \tilde{s}_1 = [s^-_a, s^+_a] \) and \( \tilde{s}_2 = [s^-_b, s^+_b] \) be two uncertain linguistic variables. The deviation measure between \( \tilde{s}_1 \) and \( \tilde{s}_2 \) is defined as follows:
\[ d(\tilde{s}_1, \tilde{s}_2) = \frac{1}{2U} \left( |I(s_{a_1}) - I(s_{a_2})|^2 + |I(s_{a_1}) - I(s_{a_2})|^2 \right) \]
\[ = \frac{1}{2U} \left( |a_1^k - a_1^l|^2 + |a_2^k - a_2^l|^2 \right) \]

where \( U \) is the number of linguistic terms in the set \( S \).

Clearly, \( 0 \leq d(\tilde{s}_1, \tilde{s}_2) \leq 1 \), \( d(\tilde{s}_1, \tilde{s}_2) = 0 \) if and only if \( \tilde{s}_1 = \tilde{s}_2 \).

**Definition 4** Suppose \( s_i = [s_{a_i}, s_{a_i}'], i \in I \) is a group of linguistic interval valuables, the best and the worst reference uncertain linguistic valuable as a datum uncertain linguistic variables are \( s_p = [s_p', s_p'], i \in I \) and \( s_q = [s_q', s_q'], i \in I \), where

\[ s_p = \max_{i \in I} \{s_{a_i}, s_{a_i}'\}, \]
\[ s_p' = \max_{i \in I} \{s_{a_i}, s_{a_i}'\} \setminus s_p, \]
\[ s_q = \min_{i \in I} \{s_{a_i}, s_{a_i}'\}, \]
\[ s_q' = \min_{i \in I} \{s_{a_i}, s_{a_i}'\} \setminus s_q. \]

**Definition 5** Suppose \( s_i = [s_{a_i}, s_{a_i}'], i \in I \) is a group of uncertain linguistic variables, the best and the worst reference uncertain linguistic valuable as a datum uncertain linguistic variables are \( s_p = [s_p', s_p'], i \in I \) and \( s_q = [s_q', s_q'], i \in I \), then the relative correlation coefficient of \( s_i \) is defined as:

\[ \alpha_i = \frac{\frac{d(s_{a_i}, s_q) - d(s_{a_i}, s_p)}{d(s_p, s_q)} - d(s_{a_i}, s_{a_i}')} \frac{d(s_{a_i}, s_{a_i}')} {d(s_p, s_q)} \]

Obviously,

1. \( -1 \leq \alpha_i \leq 1, \quad i \in I. \)
2. If \( \alpha_i > \alpha_j, \alpha_j > \alpha_k \), then \( \alpha_i > \alpha_k, i, j, k \in I. \)

For \( \alpha_i > \alpha_j, \alpha_j > \alpha_k, i, j, k \in I \), then

\[
\frac{d(s_{a_i}, s_q) - d(s_{a_i}, s_p)}{d(s_p, s_q)} > \frac{d(s_{a_i}, s_q) - d(s_{a_i}, s_p)}{d(s_p, s_q)} > \frac{d(s_{a_i}, s_q) - d(s_{a_i}, s_p)}{d(s_p, s_q)}
\]
We get $\alpha_i > \alpha_k$, so can rank uncertain linguistic variables based on their relative correlation coefficient. **Theorem 1**

Suppose $s_a=[x_{a_i}, s_{a_i}], i \in I$ is a group of linguistic interval valuable, the best and the worst reference uncertain linguistic valuable as a datum uncertain linguistic variables are $s_p=[s_{p_i}, s_{p_i}], i \in I$ and

$$s_q = [s_q, s_q], i \in I, \text{ then } s_p \succeq s_q \succeq s_q.$$ 

**Proof.** For $\forall k \in I$, according to the definition of $s_p$, we get $s_{a_i} \preceq s_{p_i}$. So if $s_{a_i} \preceq s_{p_i}$, suppose $s_{a_i} \preceq s_{p_i}$, $l \in I$, then

$$s_p = \max_{i \in I} \{ \{s_{a_i} \preceq s_{a_i} \} \setminus s_{p_i} \}, \text{ so } s_{a_i} \preceq s_p.$$ 

If $s_{a_i} = s_{p_i}$, then $s_p = \max_{i \in I} \{ \{s_{a_i} \preceq s_{a_i} \} \setminus s_{a_i} \}$. Obviously, $s_{a_i} \preceq s_{p_i}$. Overall, $s_{a_i} \preceq s_p$.

So, for $\forall i \in I, s_{a_i} \preceq s_{p_i}$. In the same way, $s_{a_i} \preceq s_p$.

**Definition 6** Suppose $C = \{c_1, c_2, \ldots, c_m\}$ is the set of attributes, the score function of attribute $c_i \in C$ is defined as:

$$s_i = \sum_{i, j=1}^{m} r_{ij}$$ 

where

$$r_{ij} = \begin{cases} 
1 & c_i > c_j \\
0.5 & c_i = c_j \quad c_i, c_j \in C \\
0 & c_i < c_j
\end{cases} \quad (4)$$ 

### 3 The solution approach

Let $A = \{A_1, A_2, \ldots, A_n\}$ be the alternative set and $C = \{c_1, c_2, \ldots, c_m\}$ be the set of all attributes. Assume $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ is the weight vector of attributes, such that $\sum_{j=1}^{m} \omega_j = 1$, $\omega_j \geq 0$ and $\omega_j$ denotes the weight of attribute $c_j$. Let $E = \{e_1, e_2, \ldots, e_n\}$ be the set of experts who give the preference of attributes and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_i)$ be the weight vector of the experts, where $\lambda_i \in (0, 1), \sum_{i=1}^{n} \lambda_i = 1$. Suppose $\tilde{S} = (s_{a_i})_{n \times m}$ is the uncertain linguistic decision matrix given by the experts, where $s_{a_i} = [x_{a_i}, s_{a_i}]$ represents the performance of the alternative $A_i$ with respect to the attribute $c_j$.

Now, we will give the process of decision based on the mentioned method before.

**Step 1.** Based on Eq. (4), compute the weight vector of attributes as:

$$\omega_j = \frac{\sum_{i=1}^{n} \lambda_i s_{j}^{k}}{\sum_{j=1}^{n} \sum_{i=1}^{m} \lambda_i s_{j}^{k}} \quad (j = 1, 2, \ldots, m) \quad (5)$$

where $s_{j}^{k}$ denote the score of $c_j$ given by expert $e_k$, $k = 1, 2, \ldots, t$.

**Step 2.** Based on the decision matrix and Eq.(1), for every alternative $A_i$, compute the corresponding weighted decision value $s_{a_i}$ as follows:
Where

\[ [a_i^*, a_i^+] = [\sum_{j=1}^{m} \omega_j I(s^*_i), \sum_{j=1}^{m} \omega_j I(s^+_i)] = [\sum_{j=1}^{m} \omega_j a^-_j, \sum_{j=1}^{m} \omega_j a^+_j] \]

Step 3. For the set \( \{s_{a_1}, s_{a_2}, \ldots, s_{a_n}\} \), decide the the best and the worst reference uncertain linguistic valuable as a datum linguistic interval \( s_p = [s_p^-, s_p^+] \) and \( s_q = [s_q^-, s_q^+] \), based on definition 4, that is

\[ s_p^* = \max_{i=1, \ldots, n} \{s_{p_i}\}, \]

\[ s_q^* = \max_{i=1, \ldots, n} \{\{s_{a_i}^-, s_{a_i}^+\}\}, \]

\[ s_q^* = \min_{i=1, \ldots, n} \{s_{q_i}\}, \]

\[ s_q^* = \max_{i=1, \ldots, n} \{\{s_{a_i}^-, s_{a_i}^+\}\}. \]

Step 4. Then, compute the relative correlation coefficient \( \alpha_i \) of \( s_{a_i}, (i = 1, \ldots, n) \) by Eq.(3), that is

\[ \alpha_i = \frac{d(s_{a_i}, s_q) - d(s_{a_i}, s_p)}{d(s_p, s_q)}, \]

then the bigger the \( \alpha_i \), the better the alternative \( A_i \), so the overall ranking is acquired.

4 The establishment of the evaluation

At present, the tourism big data visualization has gradually been widely used in industry. The characteristics of tourism big data visualization alternative can be described by a group of factor attributes, and experts can evaluate it with these attributes. Based on the National Tourism Administration website and the latest research results, experts summarized the evaluation rules and related factors generally considered in the process of evaluating the tourism big data visualization alternative, then construct the evaluation system, which will be presented as follows:

Let \( A = \{A_1, A_2, \ldots, A_5\} \) be the set of alternatives, where \( A_1 \) is statistic graph tool, \( A_2 \) is information graph tool, \( A_3 \) is map tool, \( A_4 \) is vectored graph tool, \( A_5 \) is statistic graph measured tool.

Let \( C = \{c_1, c_2, c_3, c_4\} \) be the set of attributes, where \( c_1 \) is for the government, \( c_2 \) is for the enterprise, \( c_3 \) is for the tourism practitioners, \( c_4 \) is for the broad masses of the visitor.

Suppose there are four experts who give the preference order of attributes as follows:

\[ c_4 > c_3 > c_2 > c_1 \]
\[ c_2 > c_1 > c_3 > c_4 \]
\[ c_4 > c_2 > c_1 > c_3 \]
\[ c_1 > c_2 > c_3 > c_4 \]

After many years research analysis by many experts, we can construct the decision matrix as follows:
In the following, the proposed method is applied to solve this problem and the computational procedure is summarized in the following.

Step 1. Using Eq.(5), compute the weight vector of attribute:
\[
\omega_1 = 0.333, \\
\omega_2 = 0.308, \\
\omega_3 = 0.112, \\
\omega_4 = 0.247
\]

Step 2. Compute the weighted decision value by Eq.(6):
\[
s_{a_1} = [(s_0, 0.282), (s_8, 0.035)], s_{a_2} = [(s_5, -0.343), (s_6, 0.349)], \\
s_{a_3} = [(s_5, 0.494), (s_7, 0.180)], \\
s_{a_4} = [(s_5, -0.435), (s_6, 0.145)], \\
s_{a_5} = [(s_5, 0.470), (s_7, -0.220)]
\]

Step 3. Decide the best and the worst reference uncertain linguistic valuables by Eq.(7):
\[
s_p = [(s_5, 0.186), (s_8, 0.035)], \\
s_q = [(s_5, -0.343)]
\]

Step 4. Compute the relative correlation coefficient \( \alpha_i \) of \( z_i (i = 1, 2, \ldots, n) \) by Eq.(8):
\[
\alpha_1 = 0.785, \\
\alpha_2 = -0.348, \\
\alpha_3 = 0.201, \\
\alpha_4 = -0.435
\]

Then, we can rank all the alternatives:
\[ A_1 \succ A_3 \succ A_5 \succ A_2 \succ A_4 \]

and thus the most desirable alternative is \( A_1 \).

5 Conclusion

There are many types of visualization alternatives faced by traveler, so it is hard to see which one is better. In order to deal with this problem, visualization alternative evaluations should be done to help people to decide on the better option for tourism. In this paper, a new multiple attribute decision method based on uncertain theory is given to evaluate visualization alternatives. Compared to traditional way of assessment, the new selection model proposed by this article is less liable to fault and more feasible to practice. This method can not only be used to solve visualization alternatives selection problem, but it can also be used to deal with select problem of many similar issues. In the practice application, people can not only adopt the most desirable alternative, but also combine several desirable alternative. In future research, the decision model presented in this paper will be extended according to the frequency of the application in different scenarios and group decision [3,8].

Acknowledgment

The National Natural Science Foundation Project of China (No. 61203285; No.61673285; No.11671284)  
The Province Department of Science Soft Science Project in Sichuan (2016ZR0095)
The high level research team of the major projects division of Sichuan province (Sichuan letter [2015] no.17-5)
The Project of Chengdu University of Information Technology (CRFKYTZ201644)

References