INVESTMENT DECISION ANALYSIS OF INTERNATIONAL MEGAPROJECTS BASED ON COGNITIVE LINGUISTIC CLOUD MODELS

Xiaomei MI 1,2, Huchang LIAO 1,*, Xiao-Jun ZENG 2

1 Business School, Sichuan University, 610064, Chengdu, China
2 Department of Computer Science, The University of Manchester, M13 9PL, Manchester, UK

Received 03 April 2020; accepted 15 July 2020

Abstract. The investment decision analysis of international megaprojects is a major area of interest. The choice of international megaprojects usually depends on the multi-discipline knowledge from experts. Besides, experts may not be able to provide accurate or crisp evaluations such as deterministic numbers on each criterion because of the complexity of the decision problem. In this case, natural evaluation language, either single linguistic variable or multiple linguistic variables, is a good expression tool for experts to sharing their opinions freely and flexibly. To this end, this paper introduces a cognitive linguistic cloud model for the investment decision analysis of international megaprojects as a decision support system and provides a survey of the cloud model. Afterwards, the technique to tackle multi-granularity of cognitive linguistic information is proposed to capture personalized semantics. In addition, operators of the cognitive linguistic model are proposed to aggregate natural language. The proposed approach has the advantages of more accurate utilization of experts’ knowledge, reducing uncertainties, and more effective operations of cognitive clouds for decision analysis in comparing with the state of the art. Finally, a case study about the investment of international megaprojects is given to show the flexibility and understandability of the cognitive linguistic model.

Keywords: megaproject investment, decision analysis, cognitive linguistic cloud model, multi-granular information, personalized semantics.

Introduction

Megaprojects with no blueprint or shortcut for building are required to design and establish from the ground up. The construction of international megaprojects may face incredible amounts of risk, massive cost overruns and shortfall benefit (Priemus et al., 2008; Flyvbjerg, 2017). Before the start of an international megaproject, the investment decision analysis demands expertise from multiple dimensions (Lehtinen et al., 2019), and therefore requires the multiple criteria analysis (Giezen et al., 2015). However, the complexity of international megaprojects leads to the great difficulty in providing precise information in each dimension (Keith & Ahner, 2020) required by the conventional multiple criteria analysis. In this case, the natural language sometimes is the only available and feasible expression and often a good alternative to capture evaluations in the decision-making process (Liao et al., 2018).

Natural language, such as “very good”, “between fair and good”, “40% fair and 60% good”, can be modeled by hesitant fuzzy linguistic terms or probabilistic linguistic terms (Yu et al., 2016). To align the natural language with human cognitions, the cloud model was proposed as a useful tool in modelling possible membership functions of linguistic terms (Li et al., 1998, 2009). A cloud \( C = (Ex, En, He) \) can be expressed and determined by three parameters: the expectation \( Ex \), entropy \( En \) and hyper entropy \( He \). The values of these parameters associated with basic clouds vary with the different granularities of linguistic term sets. A visualization of the linguistic term such as “very good” in the form of a cloud can be seen in Section 1. The predefined values, the golden ratio method and linguistic scale functions are the existent ways to determine the expectation parameter values of basic clouds. But, the existing research often does not effectively use linguistic terms provided by experts, and result in overlapped basic clouds in terms of the uncertainty measure, entropy and hyper entropy. The overlaps of basic clouds increase the uncertainties and lead to the difficulty in differentiating linguistic terms. Therefore filling in this research gap is the first motivation of this study.
On the other hand, in some situations with high complexity, the uncertain linguistic evaluation is the only available information. Wang et al. (2015a, 2015b) presented the trapezium cloud model to model the interval integrated cloud, and the interval intuitionistic cloud model from the membership and non-membership aspects, respectively. Peng et al. (2017) determined the entropy and hyper entropy by the square root of the endpoint values’ squared sum, which may enlarge the uncertainty in the evaluation clouds. Wu et al. (2017) used the golden ratio to determine the three parameters of the 2-tuple linguistic cloud model. Zhou et al. (2019) put forward the hesitant fuzzy linguistic cloud model based on the definition of the synthetic cloud, which is more suitable than combining two contiguous clouds only. That is to say, it may be not appropriate for the hesitant fuzzy linguistic term which contains more than two linguistic terms. Moreover, Peng et al. (2018) came up with the probabilistic linguistic cloud model based on the assumption that no ignorance exists in the available probabilistic linguistic evaluations, which is a strict restriction for use. Therefore, filling in this research gap is the second motivation of this study.

For the natural evaluation language from multiple dimensions, the computation of these evaluations in forms of cognitive models requires fusing information. Most research used the formula \( C = (\lambda, \sqrt{E_x}, \sqrt{E_n}, \sqrt{H_e}) \) to aggregate a cloud \( C = (E_x, E_n, H_e) \) and a real number \( \lambda \). Such an operation contradicts the operations of clouds as stated in Li and Du (2007) (please see Section 2.2 for details) since such an operation of \( \lambda \times C \) may unnecessarily enlarge the uncertainty by increasing the values of entropy and hyper entropy. Therefore, filling in this research gap is the third motivation of this study.

To bridge the above-mentioned research gaps, this study will propose a new approach with the following contributions:

1. **A new way to generate basic clouds.** This study proposes the personalized linguistic scale function to obtain the expectations of basic clouds. As for the uncertain measure in forms of entropy and hyper entropy, they are only influenced by the adjacent basic clouds instead of the whole basic cloud. In this way, the linguistic terms provided by experts are more accurately and more effectively used as the irrelevant (far away) clouds are ignored and as a result, uncertainties are reduced.

2. **Novel expressions of natural evaluation language.** This study focuses on the uncertain evaluation language modeled by either hesitant fuzzy linguistic term sets or probabilistic linguistic term sets, and presents cloud models for these two kinds of information. The hesitant fuzzy linguistic cloud model is a special case of the probabilistic linguistic cloud model. The incompleteness of probability of linguistic terms is allowed in the probabilistic linguistic cloud model. In this case, the restriction on the experts for providing the complete information with the complete information or without ignorance is relaxed, which decreases the burden of experts in the decision-making process.

### 3. Effective operations of cognitive clouds

This study introduces an effective operation of a cloud and a number based on the degenerated operations of two clouds. Then, arithmetic and geometric weighted operators of clouds are further presented for aggregating multiple dimensions. The effective operation controls the uncertainty into a limited range, which significantly reduces the overlaps among clouds. This also helps the experts to differentiate their evaluations.

### 4. Application angle

The applicability of the proposed cognitive cloud model is demonstrated by a case study about the investment decision analysis of international megaprojects. The advantages of the model are highlighted by the comparative analyses with existing cloud models for natural evaluation language.

To implement these contributions, the rest of this paper is organized as follows: Section 1 reviews the literature regarding the cloud model. Section 2 proposes a cognitive linguistic cloud model to utilize natural language in evaluations. Section 3 solves a case study about the investment decision analysis of international megaprojects. Section 4 presents the concluding remarks.

## 1. Literature review

This section mainly reviews literature from two aspects: the megaproject investment and cloud models to express cognitive linguistic information.

### 1.1. Literature review on the megaproject investment

Investment in projects is directly related to a new project or renovation project. In recent years, scholars have adopted fruitful methods to investigate project investment. Boateng et al. (2015) combined analytical network process with the risk priority index to analyze social, technical, economic, environmental and political risks in the construction phase. Hu et al. (2016) used fuzzy synthetic evaluation analysis to examine an evaluation framework with twenty-four program organizational factors and key performance index. Turskis et al. (2017) considered the economic, social and historical-cultural criteria in renovation projects by analytic hierarchy process and evaluation based on distance from average solution method. Osei-Kyei et al. (2017) selected seven critical criteria from fifteen criteria summarized from the literature, such as the profitability, and satisfaction degree towards the public. Zhang et al. (2020) adopted the fuzzy system dynamic model to develop some insights of mechanisms of innovation diffusion in megaprojects.

In the above-mentioned research, project investments were considered by different ways. Most research only allows the certain information in the decision analysis process, while some studies (Rostamzadeh et al., 2014;
Hu et al., 2016; Zhang et al., 2020) accept fuzzy sets as input. In some situations, natural language given by experts are more flexible than the crisp information or fuzzy information due to the multiple dimensions in project investments. Therefore, this study pays attention to the cognitive linguistic information in megaproject investments.

1.2. Literature review on cloud models to express cognitive linguistic information

This section gives a bird’s eye view of the existing techniques for cognitive linguistic information from the expression perspective. Based on the review, three research gaps of modeling cognitive linguistic information in the existing literature are presented.

The techniques of modeling qualitative information can be divided into three categories: 1) symbol-based linguistic terms, 2) membership function-based semantics of linguistic terms, and 3) membership and randomness-included cloud model. Figure 1 shows an example of the “very good” cloud (Yu et al., 2018, 2019). The three parameters of this cloud model are calculated by the method to be introduced in Section 2. The brown “*”, called as a cloud drop, denotes the possible membership degrees of a specific value belonging to the linguistic term “very good”. For a given number, there are several potential membership values to the linguistic term “very good”, which naturally expresses the randomness in human cognition. The cloud of the linguistic term “very good” contains the fuzziness and randomness features of the membership function (Li et al., 1998, 2009).

To capture linguistic evaluations in a decision-making process, we find that building the basic clouds plays an important role. Different basic clouds can tackle the multiple granularities as well. After obtaining basic clouds, the natural language evaluations can be represented by the basic models and then operated in the calculation process. Generating basic clouds contains three steps to determine the three parameters of a cloud, i.e., the expectation $Ex$, entropy $En$ and hyper entropy $He$. We assume that $U$ is the universe of discourse and $T$ is a qualitative concept in the universe of discourse $U$. In the case that $x \in X$ is a random instantiation of the qualitative concept $T$, which satisfies $x \sim N(Ex, En^2)$ and $En' \sim N(En, He^2)$, the degree $y$ of $x$ belonging to the qualitative concept $T$ is $y = e^{-2En}$. The Gaussian distribution of $x$ is called as a cloud $C(Ex, En, He)$. The instantiation and associated membership degree consist a cloud drop, shown as the brown star in Figure 1. If we know these three parameters, a cloud can be produced by a forward normal cloud generator (Li et al., 2009).

Owning to the advantages of cloud models in transforming qualitative data into quantitative data, it has attracted lots of attention of scholars. The research related to cloud models are summarized in Table 1.

We find that there are three research gaps in the literature related cloud models. The first one is the methods to generate basic clouds. Most studies in Table 1 predetermined or used the golden ratio $\left(\frac{1+\sqrt{5}}{2}\right)$ to generate the values of expectation, entropy and hyper entropy, which did not consider the semantics of linguistic terms. Only nine papers (Wang et al., 2014, 2015b, 2018; Peng & Wang, 2017, 2018; M. X. Wang & J. Q. Wang, 2018; Peng et al., 2018, 2019; Liang & Wang, 2019) in Table 1 determined the expectation value $Ex$ based on the semantics of linguistic terms. But, when choosing the values of the entropy $En$ and hyper entropy $He$, these six papers regarded that the uncertainty of each basic linguistic term is related to all linguistic terms in the universe of discourse $U$. This leads to many unnecessary overlaps in the membership
Table 1. Literature about cloud models

<table>
<thead>
<tr>
<th>No.</th>
<th>Reference</th>
<th>Expectation</th>
<th>Main idea</th>
<th>Qualitative information</th>
<th>Application area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Li et al. (1998)</td>
<td>Predetermined</td>
<td>Knowledge representation and discovery based on linguistic atoms</td>
<td>Single linguistic term</td>
<td>Predicting data mining</td>
</tr>
<tr>
<td>2</td>
<td>Li et al. (2009)</td>
<td>Predetermined</td>
<td>Presentation of the cognitive model: the cloud model</td>
<td>Single linguistic term</td>
<td>Shooting; Fractal evolution of plants</td>
</tr>
<tr>
<td>3</td>
<td>Qin et al. (2011)</td>
<td>Asymmetrical expectations</td>
<td>Histogram analysis utilizing the cloud</td>
<td>Image</td>
<td>Image segmentation</td>
</tr>
<tr>
<td>4</td>
<td>Wang et al. (2014)</td>
<td>Deviation increased linguistic scale function</td>
<td>Cloud model-based operators</td>
<td>Single linguistic term</td>
<td>Company investment</td>
</tr>
<tr>
<td>5</td>
<td>Li et al. (2014)</td>
<td>Golden ratio</td>
<td>Method to generate basic clouds</td>
<td>Single linguistic term</td>
<td>Risk assessment of enterprise resource planning</td>
</tr>
<tr>
<td>6</td>
<td>Wang et al. (2015a)</td>
<td>Golden ratio</td>
<td>Interval integrated cloud</td>
<td>Interval linguistic variable</td>
<td>Mineral investment</td>
</tr>
<tr>
<td>7</td>
<td>Wang et al. (2015b)</td>
<td>Deviation increased linguistic scale function</td>
<td>Interval-valued intuitionistic linguistic information</td>
<td>Interval-valued intuitionistic linguistic information</td>
<td>Company investment</td>
</tr>
<tr>
<td>8</td>
<td>Zhang et al. (2016)</td>
<td>Predetermined</td>
<td>Neutrosophic normal cloud</td>
<td>Neutrosophic set</td>
<td>Backpack selection</td>
</tr>
<tr>
<td>9</td>
<td>Wang et al. (2016)</td>
<td>Predetermined</td>
<td>Application</td>
<td>Single linguistic term</td>
<td>Water quality assessment</td>
</tr>
<tr>
<td>10</td>
<td>Chang and Wang (2016)</td>
<td>Deviation increased linguistic scale function</td>
<td>Application</td>
<td>Single linguistic term</td>
<td>Supplier selection</td>
</tr>
<tr>
<td>11</td>
<td>Peng and Wang (2017)</td>
<td>Personalized linguistic scale functions</td>
<td>Linguistic intuitionistic cloud; asymmetric normal clouds</td>
<td>Linguistic intuitionistic fuzzy number</td>
<td>Sustainable energy crop selection</td>
</tr>
<tr>
<td>12</td>
<td>Gao and Yi (2017)</td>
<td>Golden ratio</td>
<td>Cloud generalized power ordered weighted operator</td>
<td>Single linguistic term</td>
<td>Company investment</td>
</tr>
<tr>
<td>13</td>
<td>Peng et al. (2017)</td>
<td>Golden ratio</td>
<td>Interval integrated cloud</td>
<td>Uncertain linguistic variables</td>
<td>Supplier selection</td>
</tr>
<tr>
<td>14</td>
<td>Wu et al. (2017)</td>
<td>Golden ratio</td>
<td>2-tuple linguistic cloud</td>
<td>2-tuple linguistic information</td>
<td>Low-speed wind farm site selection</td>
</tr>
<tr>
<td>17</td>
<td>M. X. Wang and J. Q. Wang (2018)</td>
<td>Deviation increased linguistic scale function</td>
<td>Unbalanced linguistic label</td>
<td>Asymmetric and personalized linguistic information</td>
<td>Online recommendation</td>
</tr>
<tr>
<td>18</td>
<td>Wang et al. (2018)</td>
<td>Deviation increased linguistic scale function</td>
<td>Similarity measure between clouds based on alpha cuts</td>
<td>Single linguistic term</td>
<td>Logistics park selection</td>
</tr>
<tr>
<td>19</td>
<td>Peng et al. (2018)</td>
<td>Personalized linguistic scale function</td>
<td>Probabilistic linguistic integrated cloud</td>
<td>Probabilistic linguistic information</td>
<td>Hotel selection</td>
</tr>
<tr>
<td>20</td>
<td>Peng and Wang (2018)</td>
<td>Deviation increased linguistic scale function</td>
<td>Z cloud</td>
<td>Z number</td>
<td>Air pollution potential evaluation</td>
</tr>
<tr>
<td>22</td>
<td>Peng et al. (2019)</td>
<td>Personalized linguistic scale function</td>
<td>Z-trapezium-normal clouds</td>
<td>Z number</td>
<td>New energy evaluation</td>
</tr>
<tr>
<td>23</td>
<td>Zhou et al. (2019)</td>
<td>Predetermined</td>
<td>Combination with evidence theory</td>
<td>Hesitant fuzzy linguistic terms</td>
<td>Information warfare system evaluation</td>
</tr>
<tr>
<td>25</td>
<td>Liu et al. (2020)</td>
<td>Golden ratio</td>
<td>Combination with PROMETHEE II</td>
<td>2 dimensional uncertain linguistic terms</td>
<td>Choose renewable energy source</td>
</tr>
</tbody>
</table>
functions of these linguistic terms. This is because, when experts give their evaluations in linguistic terms, they usually use the adjacent linguistic terms as the references and the more far away terms are in fact irrelevant. For this reason, involving all linguistic terms is an inaccurate expression of experts’ evaluation and unnecessarily increases the uncertainties. To eliminate overlaps, motivated by the asymmetrical cloud model (Qin et al., 2011), Peng et al. (2018) adjusted the formula to determine the entropy \( En \) by only taking into account the differences of adjacent linguistic terms’ expectation values. However, the value of hyper entropy \( He \) is influenced by the set of basic linguistic terms as well, which may increase the uncertainty and reduce the accuracy but this is not considered by Peng et al. (2018). In short, the gap in the literature about generating basic clouds, the experts’ evaluations in each basic linguistic term are not accurately expressed because overlaps of clouds still exist.

The second challenge is the representation techniques for uncertain linguistic terms (not single linguistic term), such as hesitant fuzzy linguistic terms and probabilistic linguistic terms. It is worth noticing that the hesitant fuzzy linguistic terms can be regarded as the probabilistic linguistic terms with the same probability for all linguistic terms. As we know, each certain linguistic term can be represented by a cloud. Further, it is possible that the complex linguistic evaluations with uncertain linguistic terms could be modeled by certain linguistic terms with coefficients. In other words, the uncertain linguistic terms can be denoted by the operation results of basic clouds. However, such a possibility has not been investigated in the literature. Due to the uncertainty, the ignorance in uncertain linguistic terms, despite its possibility and existence naturally, has not been allowed in the existing literature.

The third challenge is the computation techniques of the multiplication between a crisp value \( \lambda \) and a cloud \( C(En,\alpha,\lambda) \). The cloud model is a generalization of crisp numbers and fuzzy numbers. If the hyper entropy \( He \) is equal to zero, the cloud reduces to a fuzzy number with a Gaussian membership function. If the entropy \( En \) and the hyper entropy \( He \) are zero at the same time, the cloud model reduces to a crisp number \( Ex \). Most studies (Wang et al., 2014, 2015a, 2015b, 2018) used the equation \( \lambda \times C = \{ \lambda \times Ex, \sqrt{\lambda \times En}, \sqrt{\lambda \times He} \} \), which is not accordance with the multiplication operation of two clouds as proposed in Li and Du (2007), and enlarges the uncertainty. We will discuss the details of this issue in Section 2.3. To bridge these research gaps, Section 2 presents a procedure of capturing natural evaluation language by cognitive linguistic cloud models.

2. Cognitive linguistic cloud models: capturing natural language

In this section, cognitive cloud models are established in three stages: generating personalized basic cloud models of multi-granularity linguistic term sets, visualization of natural evaluation language in forms of cognitive cloud models, and calculation by cognitive linguistic cloud model-based operators. The procedure of the cloud model-based decision-making process is summarized in Figure 2.

2.1. Generate basic clouds

The basic cloud has a pivotal role in modeling personalized semantics. The number of basic clouds is related to the number of linguistic terms in the linguistic term set. The common linguistic term set \( S = \{ s_{a} | a = 0,1,\ldots,\tau,\ldots,2\tau \} \) can own multi-granularity, i.e., five, seven, or nine linguistic terms (Yu et al., 2020). The more linguistic terms in a linguistic term set, the finer the granularity of the linguistic term set is. This can be shown in Figure 3.

In Figure 3, the linguistic terms are assumed to obey the uniform distribution in the universe of discourse. In these cases, the uniform semantics \( f_{UN}(s_{a}) \) of the linguistic term \( s_{a} \) can be obtained by \( f_{UN}(s_{a}) = \alpha/2\tau \). Besides the uniform distribution, the deviations of two adjacent linguistic terms may be increased or decreased according to the psychological behaviors of experts, which can be mathematically represented as follows (Wu & Liao, 2019):

\[
\begin{align*}
\bar{f}_{IN}(s_{a}) = & \frac{t^{\tau} - t^{\tau-a}}{2t^{\tau} - 2} \times \frac{1}{|a|} + \frac{t^{\tau} + t^{\tau-a} - 2}{2t^{\tau} - 2} \times \frac{1}{|\tau|} \\
\bar{f}_{DE}(s_{a}) = & \frac{t^{\tau} - (\tau - \alpha)^{\gamma}}{2t^{\tau}} \times \frac{1}{|a|} + \frac{t^{\tau} + (\alpha - \tau)^{\gamma}}{2t^{\tau}} \times \frac{1}{|\tau|}
\end{align*}
\]

where: \( 1_{|a|} + 1_{|\tau|} = 1 \) and \( 1_{|a|} \times 1_{|\tau|} = 0 \) are devised to unify the piecewise functions. Further \( t = 2^{t^{0.88}} \), \( \gamma = 0.88 \) and \( \psi = 0.88 \) are the recommended values in the linguistic scale functions (Peng et al., 2018).

In this paper, without loss of generality, we set the granularity of the linguistic term set as nine to compare our work with the recent study (Peng et al., 2018). After determining the number of basic clouds, we also take the personalized semantics of linguistic terms into
As for all asymmetrical clouds in Peng et al. (2018), the right cloud of \( C_a \) is symmetrical to the left cloud of \( C_{a+1} \) for \( a = 0, 1, \cdots, \tau, \cdots, 2\tau - 1 \), which is inconsistent with human cognitions. In our approach to generate basic clouds, we only set the endpoint clouds as asymmetric clouds \((C_{0,0,0,0})\), where the values of \( He \) and \( En \) are the hyper entropy and entropy of the left cloud, respectively, and \( En \) and \( He \) are the hyper entropy and entropy of the right cloud, respectively. That is to say, \( C_0(0,0,0,0,0) \) is the half downcloud and \( C_{2\tau}(He,En,Ex,0,0) \) is the half upcloud. The other basic clouds are regarded as symmetrical clouds with only three parameters \( Ex, En \) and \( He \) since the symmetricity exists naturally.

The deviation-increased linguistic scale function \( INf \) is taken into consideration to capture personalized semantics. The parameter of nine basic clouds is calculated by

\[
t = \frac{2}{10} \approx 1.2765.
\]

Based on the parameter \( t \), we can obtain two asymmetric clouds and nine symmetric clouds, which are shown in Figure 4. The deviation-increased linguistic scale function \( f_{IN} \) is taken into consideration to capture personalized semantics. The parameter of nine basic clouds is calculated by

\[
t = \frac{2}{10} \approx 1.2765.
\]
Compared with the nine basic clouds in Peng et al. (2018), this paper uses the right ratio rather than the predetermined value $\sqrt{10} \approx 3.16$ given that there are nine basic linguistic terms in the linguistic term set. Besides, the values of entropy $E_n$ and hyper entropy $H_e$ are smaller than or equal to those in Peng et al. (2018), which shows the less uncertainty in the nine clouds generated in this paper. Previous research (Wang et al., 2014; Peng et al., 2018) took all clouds with respect to the linguistic term as $s_a$ when determining the uncertainty measures, entropy and the hyper entropy. The reasons why our method to generate the basic clouds with the less uncertainty or the less overlaps are the effective ways to determine the entropy and the hyper entropy, Eqs. (2) and (3). Eqs. (2) and (3) determine the basic cloud $C_a$ of the linguistic term $s_a$ by focusing on the adjacent two linguistic terms, $s_a-1$ and $s_a+1$, and ignoring the irrelevant (far away) clouds related to the linguistic term $s_a$.

2.2. Cloud expressions of natural evaluation language

Under the assumption that the basic clouds are obtained, this section reviews the existing algebraic operations of clouds to give the cloud expressions of natural evaluation language. The operations of clouds proposed by Li and Du (2007) are as follows:

$$C_1 + C_2 = \left( E_{x_1} + E_{x_2}, \sqrt{E_{n_1}^2 + E_{n_2}^2}, \sqrt{H_{e_1}^2 + H_{e_2}^2} \right);$$

$$C_1 - C_2 = \left( E_{x_1} - E_{x_2}, \sqrt{E_{n_1}^2 + E_{n_2}^2}, \sqrt{H_{e_1}^2 + H_{e_2}^2} \right);$$

$$C_1 \times C_2 = \left( E_{x_1} \cdot E_{x_2}, \sqrt{E_{n_1}^2 E_{n_2}^2 + E_{n_1}^2 E_{n_2}^2}, \sqrt{E_{e_1}^2 H_{e_1}^2 + E_{e_1}^2 H_{e_2}^2} \right);$$

$$C^\lambda = \left( E_{x_1}^\lambda, \sqrt{E_{n_1}^\lambda E_{n_2}^\lambda}, \sqrt{H_{e_1}^\lambda + H_{e_2}^\lambda} \right),$$

where: $C_1 \left( E_{x_1}, E_{n_1}, H_{e_1} \right)$ and $C_2 \left( E_{x_2}, E_{n_2}, H_{e_2} \right)$ are two clouds and $\lambda$ is a crisp number.

In the above operations, the operations of clouds would reduce to the operation of a cloud and a crisp value when one cloud's entropy and hyper entropy are zero at the same time. For example, the multiplication of a crisp number $\lambda$ and a cloud $C \left( E, H \right)$ is $\lambda \times C = \left( \lambda E, \lambda H \right)$. In Peng et al. (2018), the operation is modified into $\lambda \times C = \left( \lambda E, \sqrt{\lambda E} \right)$, which is inconsistent with or contradictory to the algebraic operations of clouds. For this reason, new and better operations will need to be developed.

After expressing natural evaluation language by cloud models, we need to calculate the linguistic information based on the operations of clouds. The section mainly focuses on the uncertain linguistic information in the forms of hesitant fuzzy linguistic term sets and probabilistic linguistic term sets. In our view, a hesitant fuzzy linguistic term set $h_{S}$ is a special probabilistic linguistic term set with the uniform probability distribution to each linguistic term, that is, $h_{S} = \left\{ \left( \frac{1}{\#S}, s_l \right) \mid s_l \in S, L = 1, 2, \ldots, \#S \right\}$, where $\#S$ represents the cardinality of the linguistic terms.

A hesitant fuzzy linguistic term set can be described by the following cloud model based on the operations of clouds:

$$h_j = \sum_{L=1}^{\#S} \frac{1}{\#S} E_{n_L}^{2j} = \left( \sum_{L=1}^{\#S} \frac{1}{\#S} E_{n_L}^{2j}, \sum_{L=1}^{\#S} \frac{1}{\#S} E_{n_L}^{2j}, \sum_{L=1}^{\#S} \frac{1}{\#S} E_{n_L}^{2j} \right) \times \sqrt{\sum_{L=1}^{\#S} \frac{1}{\#S} H_{e_L}^2};$$

where: $s_j$ is modeled by the cloud $C_L \left( E_{n_L}, E_{x_L}, H_{e_L} \right)$.

If a hesitant fuzzy linguistic term set $h_{S_1}$ “between slightly good and good” is given on a nine-valued linguistic term set, then the natural evaluation language can be translated into “between slightly good and good” which means $\#S_1 = 2$ and $h_{S_1} = \left\{ \left( 0.5, s_6 \right), \left( 0.5, s_7 \right) \right\}$. Based on the aforementioned operations of clouds, $h_{S_1}$ can be described as $C_{S_1} = \left( 6.3684, 0.2257, 0.0208 \right)$, visually shown as the blue points in Figure 5. The green plus symbols are possible cloud drops calculated by the operation in Peng et al. (2018).
It is not very hard to find that the green cloud involves more uncertainty than the blue one, with larger value ranges of possible entropy and hyper entropy. That is to say, the operations of clouds and the expression technique developed in this paper capture the natural evaluation language more accurately than Peng et al. (2018)'s model.

As for probabilistic linguistic term sets, Peng et al. (2018) assumed that the probability sum of these elements equals to one. In other words, their model does not allow the existence of ignorance regarding the probability of linguistic terms. Actually, there are two categories of linguistic evaluations with the forms of complete probability and incomplete probability. For example, the complete probabilistic linguistic term set "80% slightly good and 20% good", and "60% slightly good and 40% good" are shown as the red x symbol and the magenta diamond symbol, respectively in Figure 6. Besides the complete probability in the probabilistic linguistic term sets, the ignorance denoted by incomplete probability sum should not be forbidden. For example, if the probabilistic linguistic term "60% slightly good and 20% good" is provided, we can assign the ignorance probability to endpoints, the full linguistic term set, or the envelop containing the lower and upper linguistic terms (Fang et al., 2020). The red and magenta clouds are lower and upper bounds of the incomplete probabilistic linguistic term. The clouds of these latter two common parts can be calculated by the addition and multiplication operations of clouds. An example of the cloud of the incomplete probabilistic linguistic term set is illustrated in Figure 7.

![Figure 6. An example of the cloud of complete probabilistic linguistic term set](image6)

![Figure 7. An example of the cloud of the incomplete probabilistic linguistic term set](image7)
Based on the above analyses, natural evaluation language can be represented by cloud models regardless of the length of linguistic evaluations. If only one certain linguistic term is given in a defined linguistic term set, the basic cloud is a good alternative to express the natural evaluation language. If uncertain linguistic terms are provided, we can classify the evaluations into hesitant fuzzy linguistic term sets, probabilistic linguistic term sets with or without complete probability. Generally speaking, the cognitive cloud model expresses the natural evaluation language of an expert in two steps: generating basic clouds with the granularity, and expressing the natural evaluation language based on the generated basic clouds and the operations of clouds. In the next subsection, the cognitive cloud model-based operators are designed to fuse information from multiple dimensions.

2.3. Cognitive cloud model-based operators

In this section, the weighted cloud model-based arithmetic and geometric operators are presented for aggregating multiple dimensional information, which are shown as:

$$WCA(C_1, C_2, \ldots, C_n) = \sum_{i=1}^{n} w_i C_i = \left( \sqrt{\sum_{i=1}^{n} w_i^2 E_{i1}^2}, \sum_{i=1}^{n} w_i E_{i1}, \sqrt{\sum_{i=1}^{n} w_i^2 E_{i2}^2} \right);$$

$$WC(G(C_1, C_2, \ldots, C_n)) = \prod_{i=1}^{n} C_i^{w_i} = \left[ \prod_{i=1}^{n} E_{i1}^{w_i}, \prod_{i=1}^{n} E_{i2}^{w_i}, \sqrt{\sum_{i=1}^{n} w_i^2 \left( \frac{E_{i1}}{E_{i2}} \right)^2}, \prod_{i=1}^{n} E_{i2}^{w_i}, \sqrt{\sum_{i=1}^{n} w_i^2 \left( \frac{E_{i2}}{E_{i1}} \right)^2} \right].$$

where: $w_i \in [0,1], i = 1, 2, \ldots, n$ and the weights are normalized as $\sum_{i=1}^{n} w_i = 1$.

The weighted operators are calculated by the operations of cognitive cloud models, i.e., $\lambda \times C = (\lambda E_{x1}, \lambda E_{x2}, \lambda E_{x3}, \lambda E_{x4})$, $C_1 \times C_2 = (E_{x1} \times E_{x2}, E_{x3} \times E_{x4}, E_{x5} \times E_{x6}, E_{x7} \times E_{x8})$ and $C^\lambda = (E^\lambda, \sqrt{\lambda} E^{\lambda^2}, E^{\lambda^2}, \sqrt{\lambda} E^{\lambda^3})$. Hence, the aggregated results by the WCA and WCG operators are also a cloud. Compared with the previous cloud model-related operators proposed in Peng et al. (2018), less uncertainty are produced by the WCA and WCG operators.

In the process of capturing natural language by cognitive linguistic cloud models, generating basic clouds with different granularities of linguistic term sets is the preliminary step, which was addressed in Section 2.1. Based on the established basic units, certain or uncertain linguistic terms can be represented based on the operations of corresponding clouds, which was justified in Section 2.2. If the evaluation language comes from multiple dimensions, the weighted cloud model-based operators given in Section 2.3 can fuse information into one dimension.

3. Case study: investment decision analysis of international megaprojects

This section gives a case study about the investment decision analysis of international megaprojects by the proposed cognitive cloud models. Further, we will compare the presented model with other similar models to show its advantages.

3.1. Case description

The McKinsey Global Institute (Woetzel et al., 2016) estimated that, during the time period from 2013 to 2030, about 4% of total global gross domestic product (3.4 US dollars trillion / year) will be served for large-scale projects. Megaprojects are usually accompanied with a large amount of committed investment, long-lasting impact on economy, environment and society (Flyvbjerg, 2011, 2014; Brookes & Locatelli, 2015). Investigating megaprojects is a continuing concern within the research area of economy. Boateng et al. (2015) analyzed the risks in megaprojects using the analytical network process from the multiple criteria perspective. He et al. (2015) measured the construction complexity of megaprojects from technological, organizational, goal, environmental, cultural and information aspects with fuzzy numbers. Chapman (2016) studied the inherent multiple dimensions and complexity within the framework of megaprojects. Lin et al. (2017) presented the indicator system for evaluating the social responsibility of megaprojects at different organizational levels.

Moreover, any definition of megaprojects contains big business, which is less influenced by the economy recession. In some situations, the investment of megaprojects may be stimulated by the economy downturn, such as 2008 financial crisis (Flyvbjerg, 2014). In this way, the construction of megaprojects usually contributes to the development of economy. However, the required amount of investment in megaprojects is large. Hence, understanding the complexity of a megaproject and then smartly investing are significantly important from the aspect of limited resources. To this end, Flyvbjerg (2014) provided an overview about the knowledge and reasons for investing megaprojects.

In this paper, we take four sublimes of megaprojects as four criteria related to the investment decision analysis. Moreover, the social responsibility of megaprojects is adopted as the fifth criterion (Zeng et al., 2015; Ma et al., 2017; Lin et al., 2017). The detailed criteria and related descriptions are tabulated in Table 2.

Based on the above five criteria in the investment decision analysis of international megaprojects, we adopt the five typical megaprojects in Flyvbjerg (2011) as alternatives for this study. The five alternatives and related natural evaluation language on the nine-valued linguistic term set are listed in Table 3.

Based on the obtained evaluations of the alternatives on each criterion, below we solve this investment decision problem by the proposed cognitive cloud models.
As the evaluations are given on the nine-valued linguistic term set, nine basic clouds are generated by the person-alized semantics as presented in Section 2.1 and visually shown in Figure 4. Next, we transform the evaluations into clouds based on the operations of clouds and the representation techniques. The transformed clouds of alternatives are listed in Table 4.

After obtaining the cloud expressions of alternatives' performances, we need to aggregate the clouds on five dimensions by the proposed weighted cloud model-based arithmetic operator WCA. The last column in Table 4 shows the five overall clouds of the alternatives produced by the normal cloud generator, which is illustrated in Figure 8.

In Figure 8, the blue snow symbols are possible cloud drops of the third alternative's performance cloud. We can observe that the investment ranking result of the alternatives is \( x_3 > x_2 > x_4 > x_5 > x_1 \), which implies that the expressway networks > high speed railways > gas pipeline projects > long-span bridges > large-scale hydropower projects, where “>” refers “prior to”.

### Table 2. Five criteria in investment decision analysis of international megaprojects

<table>
<thead>
<tr>
<th>Alternatives/Criteria</th>
<th>Criteria</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1 Large-scale hydropower projects</td>
<td>c_1</td>
<td>Technological Engineers and experts in technologies are extremely excited for pushing forward “international-longest-highest-fastest” types of projects</td>
</tr>
<tr>
<td>x_2 High speed railways</td>
<td>c_2</td>
<td>Political Politicians from contained countries or regions benefit from establishing monuments in international megaprojects, and the visibility from the public and the media</td>
</tr>
<tr>
<td>x_3 Expressway networks</td>
<td>c_3</td>
<td>Economic A lot of money and job opportunities can be earned by businessmen and trade unions; such as contractors, consultants, bankers, construction and transportation workers, lawyers and developers</td>
</tr>
<tr>
<td>x_4 Gas pipeline projects</td>
<td>c_4</td>
<td>Aesthetic The public gains happiness from the construction and use of large, symbolic and beautiful designs</td>
</tr>
<tr>
<td>x_5 Long-span bridges</td>
<td>c_5</td>
<td>Social responsibility Policies and practices of stakeholders’ responsibility for wider social welfare in the life cycle of international megaprojects</td>
</tr>
</tbody>
</table>

### Table 3. Evaluation language of five typical megaprojects on five criteria

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1 Large-scale hydropower projects</td>
<td>c_1</td>
<td>Technological Engineers and experts in technologies are extremely excited for pushing forward “international-longest-highest-fastest” types of projects</td>
</tr>
<tr>
<td>x_2 High speed railways</td>
<td>c_2</td>
<td>Political Politicians from contained countries or regions benefit from establishing monuments in international megaprojects, and the visibility from the public and the media</td>
</tr>
<tr>
<td>x_3 Expressway networks</td>
<td>c_3</td>
<td>Economic A lot of money and job opportunities can be earned by businessmen and trade unions; such as contractors, consultants, bankers, construction and transportation workers, lawyers and developers</td>
</tr>
<tr>
<td>x_4 Gas pipeline projects</td>
<td>c_4</td>
<td>Aesthetic The public gains happiness from the construction and use of large, symbolic and beautiful designs</td>
</tr>
<tr>
<td>x_5 Long-span bridges</td>
<td>c_5</td>
<td>Social responsibility Policies and practices of stakeholders’ responsibility for wider social welfare in the life cycle of international megaprojects</td>
</tr>
</tbody>
</table>

### Table 4. The alternatives’ performances in forms of clouds

<table>
<thead>
<tr>
<th>Alternatives/Criteria</th>
<th>c_1</th>
<th>c_2</th>
<th>c_3</th>
<th>c_4</th>
<th>c_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1 Large-scale hydropower projects</td>
<td>(4.5824, 0.1969, 0.0128)</td>
<td>(3.8449, 0.1871, 0.0172)</td>
<td>(7.5820, 0.2882, 0.0266)</td>
<td>(5.6230, 0.1128, 0.0087)</td>
<td>(7.1057, 0.2894, 0.0267)</td>
</tr>
<tr>
<td>x_2 High speed railways</td>
<td>(7.5820, 0.2882, 0.0266)</td>
<td>(4.5406, 0.1794, 0.0129)</td>
<td>(5.4176, 0.1969, 0.0128)</td>
<td>(6.1352, 0.1018, 0.0084)</td>
<td>(7.7181, 0.2572, 0.0129)</td>
</tr>
<tr>
<td>x_3 Expressway networks</td>
<td>(7.4459, 0.2043, 0.0188)</td>
<td>(7.2849, 0.1419, 0.0131)</td>
<td>(7.4459, 0.2043, 0.0188)</td>
<td>(8.2626, 0.4537, 0.0418)</td>
<td>(7.8988, 0.1319, 0.0107)</td>
</tr>
<tr>
<td>x_4 Gas pipeline projects</td>
<td>(5.3341, 0.1671, 0.0081)</td>
<td>(5.0000, 0.2784, 0.0000)</td>
<td>(7.4459, 0.2043, 0.0188)</td>
<td>(7.5820, 0.2882, 0.0266)</td>
<td>(5.6230, 0.1128, 0.0087)</td>
</tr>
<tr>
<td>x_5 Long-span bridges</td>
<td>(5.3341, 0.1671, 0.0081)</td>
<td>(5.0000, 0.2784, 0.0000)</td>
<td>(7.1057, 0.2894, 0.0267)</td>
<td>(5.8988, 0.1319, 0.0107)</td>
<td>(6.1352, 0.1018, 0.0084)</td>
</tr>
</tbody>
</table>

### Table 4. The alternatives’ performances in forms of clouds

<table>
<thead>
<tr>
<th>Alternatives/Criteria</th>
<th>c_4</th>
<th>c_5</th>
<th>Overall clouds</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1 Large-scale hydropower projects</td>
<td>(5.0000, 0.2784, 0.0000)</td>
<td>(5.6230, 0.1128, 0.0087)</td>
<td>(5.6230, 0.1128, 0.0087)</td>
</tr>
<tr>
<td>x_2 High speed railways</td>
<td>(5.4176, 0.1969, 0.0128)</td>
<td>(6.1352, 0.1018, 0.0084)</td>
<td>(6.1352, 0.1018, 0.0084)</td>
</tr>
<tr>
<td>x_3 Expressway networks</td>
<td>(7.4459, 0.2043, 0.0188)</td>
<td>(7.2849, 0.1419, 0.0131)</td>
<td>(7.2849, 0.1419, 0.0131)</td>
</tr>
<tr>
<td>x_4 Gas pipeline projects</td>
<td>(7.5820, 0.2882, 0.0266)</td>
<td>(5.9633, 0.1214, 0.0089)</td>
<td>(5.9633, 0.1214, 0.0089)</td>
</tr>
<tr>
<td>x_5 Long-span bridges</td>
<td>(8.2626, 0.4537, 0.0418)</td>
<td>(7.1057, 0.2894, 0.0267)</td>
<td>(7.1057, 0.2894, 0.0267)</td>
</tr>
</tbody>
</table>
3.3. Comparative analyses

This section compares the results deduced by other models with the one derived in Section 3.2 from three aspects, i.e., basic clouds generation, natural evaluation language expression in clouds, and multiple-dimensional clouds aggregation.

1. As for the basic clouds’ generation, this paper takes the personalized semantics into consideration instead of the fixed golden ratio value. The expectation values can be determined by three types of linguistic scale functions. For the values of entropy and hyper entropy, the differences between the expectations of adjacent linguistic terms are considered rather than the maximal adjacent difference among the whole clouds. This reduces the uncertainty and ensures the smaller values of entropy and hyper entropy, which can be visually demonstrated by the less overlaps of clouds.

2. Expressing natural evaluation language can be divided into two scenarios: the certain evaluation language representation and the uncertain evaluation language representation. As for the former case, the generated basic clouds are useful tools. With respect to the latter case, this paper models the uncertain evaluation language in the forms of hesitant fuzzy linguistic term sets and probabilistic linguistic term sets. In particular, the ignorance of probability is accepted in this study compared with Peng et al. (2018) which gave a strict restriction that the sum of probabilities of linguistic terms should be equal to one.

3. From the information fusion process of cognitive clouds, different operators based on the operations of clouds are presented. Compared with the previous research in Peng et al. (2018), the uncertainty measures would not enlarge too much by our aggregation operators. If we assume that the incomplete probabilistic linguistic term sets can be replenished by the envelope of related linguistic terms, then five clouds with larger uncertainty can be obtained by the operator in Peng et al. (2018). The larger uncertainty can be observed by the overlaps in Figure 9.
In conclusion, the cognitive cloud model proposed has the advantages in modeling natural evaluation languages and fusing multiple language evaluations. It controls the original uncertainty in the computation process without enlarging the uncertain part in the entropy and hyper-entropy.

Conclusions

To invest international megaprojects with natural evaluation languages, this paper presented a cognitive cloud model to capture information modeled as hesitant fuzzy linguistic term sets and probabilistic linguistic term sets. The cognitive cloud model generated basic clouds on a linguistic term set by considering the personalized semantics and the granularity of a linguistic term set at the same time. Based on the basic clouds, the cognitive linguistic evaluation information can be modeled by hesitant fuzzy linguistic clouds and probabilistic linguistic clouds. The former was a special case of the latter with no ignorance and equal probability for each linguistic term. With respect to the computation process of clouds, two weighted cloud model-based operators were proposed from the arithmetic and geometric aspects. A case study regarding the investment decision analysis of international megaprojects was implemented to show the application and advantage of the presented cognitive cloud model.

In this paper, the probabilistic linguistic information can be viewed as the payoff value of the expert if (s)he chooses the specific alternative on the criterion in one time. This can be regarded as the game between the expert and the nature (Wu & Seidmann, 2018). In the future, the matrix games with hybrid strategies using clouds may be interesting research topics with challenges. The main disadvantage of the proposed cognitive linguistic cloud model is the determination of the hyper-parameters in the personalized linguistic scale function. We will consider to adopt machine learning and data mining techniques to analysis the personalities of decision-makers based on the available information.

Funding

The work was supported by the National Natural Science Foundation of China (Nos. 71771156, 71971145) and the Scholarship from China Scholarship Council (No. 201906240161).

Author contributions

Xiaomei Mi and Huchang Liao proposed the original idea, conceived the study, and wrote the first draft of the article. Xiao-Jun Zeng revised the paper.

Disclosure statement

The authors have no competing financial, professional, or personal interests from other parties that are related to this paper.

References


