
Peer reviewed version

License (if available):
Other

Link to published version (if available):
10.1109/TFUZZ.2019.2906856

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at https://ieeexplore.ieee.org/document/8675512. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms
Linguistic distribution-based optimization approach for large-scale GDM with comparative linguistic information. An application on the selection of wastewater disinfection technology

Hengjie Zhang, Jing Xiao, Iván Palomares, Haiming Liang, and Yucheng Dong

Abstract: Managing comparative linguistic expressions (CLEs) information is a key issue in group decision-making (GDM). A transformation approach has been previously defined to convert CLEs into hesitant fuzzy linguistic terms sets (HFLTSs). However, it is noted that the occurring possibilities of the linguistic terms in the HFLTSs are assumed equal. This assumption might sometimes not capture the real opinions of the decision makers. Linguistic distribution assessments (LDAs) are an effective way to deal with this issue. This paper develops a linguistic distribution-based optimization approach for converting CLEs into LDAs, in which we assume that decision makers provide their opinions using preference relations with CLEs. Particularly, the proposed optimization approach is based on the use of a consistency-driven methodology, which seeks to minimize the inconsistency level of LDA preference relations obtained by transforming the original CLE preference relations elicited from decision makers. The linguistic distribution-based optimization approach is further developed to transform CLEs into interval LDAs to increase their flexibility. Moreover, society and technology trends make it possible to involve and manage large groups of decision makers in GDM environment. So, a large-scale GDM framework with CLE information is designed based on the linguistic distribution-based optimization approach. To justify the effectiveness and applicability of the proposed methodology, it is applied to solve a real large-scale GDM problem, pertaining the selection of best sustainable disinfection technique for wastewater reuse projects. A comparison against a baseline method is likewise provided to highlight the advantages and innovations of our proposal.

Keywords: Comparative linguistic expressions, linguistic distribution assessments, consistency, fuzzy and interval fuzzy preference relations, large-scale group decision making

I. INTRODUCTION

In our daily life, we are often faced with group decision making (abbreviated as GDM) problems [1-4]. In GDM problems, individuals are accustomed to convey their preferences using qualitative information, which is closer to human natural way of thinking and reasoning. The linguistic assessment method has been used in various areas [5-7]. Several different linguistic computation models have been suggested to handle decision problem with linguistic assessment information [8]. In the literature, the famous 2-tuple linguistic model was coined by Herrera and Martínez [9], which has proven to be useful in addressing linguistic information that is uniformly and symmetrically distributed. However, in many practical GDM problems, the linguistic term sets are asymmetrically distributed [10-12]. To deal with this, Herrera et al. [12] coined a linguistic decision making model based on a linguistic hierarchy. Meanwhile, Xu [13, 14] presented the linguistic symbolic computational model based on virtual linguistic terms, which can also avoid information loss in linguistic information processing, and Xu [15] also proposed several uncertain linguistic aggregation operators to fuse linguistic information based on the virtual linguistic model. Dong et al. [16, 17] developed a consistency-driven methodology to transform linguistic information into numerical information. Li et al. [18] proposed a personalized individual semantics model with linguistic preference relations based on the use of consistency-driven methodology. A comprehensive overview of the 2-tuple linguistic model has been made in Martínez and Herrera [19]. Additional linguistic computation models can be found in [20-23].

The above linguistic computation models have proven their usefulness to address these linguistic GDM problems, where a single linguistic label is used for conveying decision makers’ preferences. However, in many real-world linguistic GDM problems, more flexible linguistic expressions than a single linguistic label are needed due to lack of data, time pressures, and inherent vagueness exhibited by decision makers [8, 24-27]. In the literature, a context-free grammar-based approach was adopted by Rodríguez et al. [26] to elicit comparative linguistic expressions (abbreviated as CLEs). In particular, when CLEs are adopted in the pairwise comparisons method in GDM, preference relations with CLEs are constructed [28]. Recently, a consistency-driven methodology is proposed to set personalized numerical scales for linguistic terms with CLEs [29].

In the literature, CLE information is often transformed into hesitant linguistic information. Particularly, the concept of hesitant fuzzy linguistic term sets (HFLTSs) was proposed, and an approach to convert CLEs into HFLTSs was further designed [26]. However, it is noteworthy that the occurring possibilities of the linguistic terms in the HFLTSs are by default assumed to be equal, which is obviously not realistic in some practical situations. In other words, this default assumption might sometimes not capture the real opinion of the decision makers, each of whom might believe that some linguistic terms are more likely to best reflect such opinion than others. Thus, the obtained HFLTSs can be
The linguistic distribution assessments (abbreviated as LDAs) [30], which allow decision makers to assign different possibility degrees to different linguistic terms, are an effective way to deal with this issue. Wu et al. [27] developed the maximum support degree model to support linguistic GDM based on the use of LDAs and HFLTSs. It is sometimes difficult for decision makers to provide information about the possibilities of linguistic terms in a precise and exact way. The use of interval possibilities is a good way to deal with this issue, and interval LDAs were thus proposed by Dong et al. [11]. Moreover, the concept of probabilistic linguistic term sets (PLTSs) was developed to overcome the limitations of HFLTSs [31], in the case where the sum of linguistic term probabilities is not equal to 1 was considered. Gou and Xu [32] developed some operational laws for PLTSs. Further, probabilistic linguistic preference relations were developed based on the use of PLTSs [33], and a consensus process for GDM with probabilistic linguistic preference relations was designed [34]. Here, we offer the following example as an illustration to show the difference between LDAs and HFLTSs.

**Example 1:** Let \( S = \{s_0 = \text{extremely poor}, s_1 = \text{very poor}, s_2 = \text{poor}, s_3 = \text{sightly poor}, s_4 = \text{fair}, s_5 = \text{sightly good}, s_6 = \text{good}, s_7 = \text{very good}, s_8 = \text{extremely good}\} \). A football player participated in ten football games in last year, in which five times to draw, three times to win with a small score, two times to win with a big score. A football coach utilizes a CLE, greater than \( s_5 \), to evaluate a player’s level in last year. The football coach may believe that the most representative linguistic term to describe the football player’s performance is \( s_6 \) while the linguistic term \( s_5 \) is the least representative. In existing linguistic computation model, the CLE “greater than \( s_5 \)” is transformed into HFLTS \( H = \{s_6, s_7, s_8\} \). However, the HFLTS \( H = \{s_6, s_7, s_8\} \) cannot be used to describe this linguistic expression completely, because this implies that the possibilities of the linguistic terms in the HFLTS \( H = \{s_6, s_7, s_8\} \) are equal. The LDAs are an effective way to deal with this issue [30]. For example, LDA \( L = \{(s_0, 0), (s_1, 0), (s_2, 0), (s_3, 0), (s_4, 0), (s_5, 0.5), (s_6, 0.3), (s_7, 0.2)\} \) can faithfully reflect the honest opinion of the football coach. For simplification, the linguistic terms with a possibility degree of zero are omitted in the LDAs in this paper. So, \( L \) can be rewritten as \( L = \{(s_0, 0.5), (s_7, 0.3), (s_8, 0.2)\} \). Notably, LDAs can be generalized by PLTSs in the case where the sum of linguistic term probabilities is smaller than 1.

The above analysis shows that LDAs are an effective way to model CLE information in an accurate and reliable manner. However, to our knowledge, no research to date focuses on devising methods for calculating the LDA possibility degrees whilst ensuring consistency in preferences. This proposal aims to fill this research gap with a threefold contribution:

1. This paper develops a novel linguistic distribution-based optimization approach for transforming CLEs into LDAs, where we assumed that decision makers provide their opinions utilizing preference relations with CLEs. Particularly, the proposed linguistic distribution-based optimization approach is inspired by the consistency-driven methodology presented in Dong et al. [16, 17], which seeks to minimize the inconsistency level of preference relations with LDAs obtained by transformation of a preference relation with CLEs.

2. In practice, it is not always easy for decision makers to offer the possibility degrees of linguistic terms by means of precise and exact values in an uncertain decision environment. Therefore, the second contribution in this work consists in extending the above linguistic distribution-based optimization approach into an interval-valued context. Based on this approach, a preference relation with interval LDAs can be generated from its associated preference relation with CLEs.

3. Nowadays, the development of information technology (such as E-government and E-commerce and social media) is causing a significant shift from conventional, small-group GDM towards large-scale GDM problems [35-39]. Large-scale GDM problems involve a larger number of decision makers [40], which has received wide attention in the decision-making field in recent years. Palomares in [40] defines a large-scale GDM problem as “a situation involving between several tens and thousands of participants with diversity in background, expertise level, behavior, attitudes and possibly conflicting interests/viewpoints, who must make a collective and acceptable decision pertaining a relevant problem to all of them”. Based on the linguistic distribution-based optimization approach, the third contribution presented in this work consists in presenting a large-scale GDM framework with CLE information. Particularly, the two proposed linguistic distribution-based optimization approaches are used to produce highly-consistent preference relations with LDAs or interval LDAs from their associated preference relations with CLEs in the proposed large-scale framework.

In addition to the three theoretical contributions listed above, an application on the selection of wastewater disinfection technology is provided to show the effectiveness of the proposal. In recent years, wastewater reuse is becoming a particularly important decision problem involving highly diverse groups of experts and stakeholders from diverse areas and disciplines, hence it can be potentially benefited from large-group informed decision-making, especially in the zones where the water resource is quantitatively and qualitatively scarce [41]. Choosing a sustainable treatment for wastewater reuse facilities presents a serious challenge for wastewater reuse project managers as well as for a large number of stakeholders and actors with highly diverse expertise and background in the process of obtaining a best solution [42, 43]. The problem of sustainable disinfection technique evaluation and selection often involves a large number of stakeholders (or decision makers). When evaluating sustainable disinfection techniques, the individuals may not provide their opinions using precise assessments, and the preference relations with CLEs are an adequate and intuitive tool for them to express their opinions and effectively capture the uncertainty underlying them. To help wastewater reuse project managers select a best sustainable disinfection technique, we present a case study in which the proposed large-scale GDM framework with comparative linguistic information is applied in a real wastewater disinfection large-scale GDM problem. Lastly, a detailed comparative analysis is provided to show the benefit of our proposed methodology with respect to existing approaches.

The remainder of this paper is arranged as follows. Section II introduces some basic knowledge. Then, Section III presents the linguistic distribution-based optimization approach to generate preference relations with LDAs or interval LDAs from the preference relations with CLEs. Following this, Section IV designs a large-scale GDM framework based on the linguistic
distribution-based optimization approach. Subsequently, Section V provides a case study of wastewater disinfection technology selection to show the application of the proposed framework, and a comparison study is also provided in this section. Finally, Section VI summarizes the paper and points out research directions for the future.

II. PRELIMINARIES

Some basic knowledge about the 2-tuple linguistic model, preference relations with CLEs, hesitant fuzzy linguistic preference relations (abbreviated as HFLPRs), preference relations with LADs and interval LADs, and fuzzy and interval fuzzy preference relations are presented in this section.

A. The 2-tuple linguistic model

Here, we use \( S = \{ s_0, \ldots, s_g \} \) to denote a linguistic term set, where \( g + 1 \) is a granularity of \( S \) and \( s_j \) signifies a possible linguistic value. Usually, the following two conditions should be satisfied: (1) \( S \) is ordered: \( s_i \leq s_j \) if and only if \( i \leq j \); (2) there is a negation function: \( \text{neg}(s_j) = s_{g-j} \). The detailed information about the linguistic variables can be found in Herrera and Martínez [9] and Herrera et al. [5].

The 2-tuple linguistic model is a famous linguistic computational model that is coined by proposed by Herrera and Martínez [9]. Let \( S = \{ s_0, \ldots, s_g \} \) be defined as the above. The 2-tuple that conveys equivalent evaluation information to \( \beta \in [0, g] \) can be yielded by the following formula:

\[
\Delta : [0, g] \rightarrow S \times (-0.5, 0.5),
\]

where

\[
\Delta(\beta) = (s_i, \alpha), \quad \text{with} \quad s_i = \text{round}(\beta) \\
\alpha = \beta - i, \quad \alpha \in (-0.5, 0.5).
\]

In the above linguistic model, \( \Delta \) is a one to one mapping function. Here, all linguistic 2-tuples associated with \( S \) are denoted by a set \( \tilde{S} \). An inverse operator for \( \Delta \) can be built, \( \Delta^{-1} : \tilde{S} \rightarrow [0, g] \) with \( \Delta^{-1}(s_i, \alpha) = i + \alpha \). In this paper, we set \( \Delta^{-1}(0, 0) = \Delta^{-1}(0, 0) \).

Here, we use \((s_i, \alpha)\) and \((s_j, \gamma)\) to represent two linguistic 2-tuples. If \( \Delta^{-1}(s_i, \alpha) > \Delta^{-1}(s_j, \gamma) \), then \((s_i, \alpha)\) is larger than \((s_j, \gamma)\).

B. Preference relations with CLEs

A context-free grammar-based approach is presented by Rodríguez et al. [26] to produce CLEs, which is introduced below.

Definition 1 (Context-free grammar) [26]. Let \( S \) be as above, a context-free grammar is a 4-tuple \( G_{H} = (V_S, V_T, I, P) \), where \( V_S \) indicates a set of nonterminal symbols, \( V_T \) is a set of terminal symbols, \( I \) is the starting symbol, and \( P \) denotes the production rules. The elements of \( G_{H} \) are defined as follows:

\[
V_S = \{ \langle \text{primary term} \rangle, \langle \text{composite term} \rangle, \langle \text{unary relation} \rangle, \langle \text{binary relation} \rangle, \langle \text{conjunction} \rangle \};
\]
\[
V_T = \{ \text{lower than}, \text{greater than}, \text{at least}, \text{at most}, \text{and}, s_0, s_1, \ldots, s_g \};
\]
\[
P = \{ \langle I := \langle \text{primary term} \rangle \rangle, \langle \text{composite term} \rangle \}
\]
\[
\langle \text{composite term} \rangle := \langle \text{unary relation} \rangle \langle \text{primary term} \rangle
\]

The preference degree of the alternative \( s_i \) over \( s_j \) is computed as follows:

\[
E(s_i, s_j) = \sum_{p_i=0.1}^{g} NS(s_i) \cdot p_i
\]

where \( NS(s_i) \) is the numerical scale of linguistic term \( s_i \). For detailed information about the numerical scale, see Dong et al. [45].
For convenience, we use $LD_2$ to denote a set of LDAs associated with linguistic term set $S$. Based on the use of LDAs, the preference relation with LDAs is constructed.

**Definition 7** [30]. Let $LD_2$ be as above. A preference relation with LDAs is represented using a matrix $C=(c_{ij})_{n\times n}$, where $c_{ij} \in LD_2$ denotes the preference intensity of alternative $x_i$ against $x_j$.

It is sometimes difficult for decision makers to express exact possibilities of linguistic terms in LDAs. To deal with this issue, the interval LDAs were developed [11].

**Definition 8.** Let $S=\{s_0, ..., s_g\}$ be as above. An interval LDA is denoted by $IL=(\{s_0, [p_0^l, p_0^u]\}, (s_1, [p_1^l, p_1^u]), ..., (s_g, [p_g^l, p_g^u]))$, where $[p_i^l, p_i^u] \in [0, 1]$, $(i=0, ..., g)$, denotes the possibility degree of linguistic term $s_i$ satisfying $\sum_{i=0}^{g} p_i^l + p_i^u \leq 1$ and $p_i^l + \sum_{i=0}^{g} p_i^u \geq 1$ ($s=0, ..., g$).

The expectation of interval LAD $IL$ is defined by:

$$E(IL) = \left[ \sum_{i=0}^{g} NS(s_i) \cdot p_i^l \cdot \sum_{j=0}^{g} NS(s_j) \cdot p_j^l \right]$$  \hspace{1cm} (4)

For convenience, we use $ILD_2$ to denote a set of interval LDAs associated with linguistic term set $S$. The concept of preference relation with interval LDAs is formally defined below.

**Definition 9.** Let $ILD_2$ be as above. A preference relation with interval LDAs is denoted using a matrix $D=(d_{ij})_{n\times n}$, where $d_{ij} \in ILD_2$ signifies the preference intensity of alternative $x_i$ against $x_j$.

### E. Fuzzy and interval fuzzy preference relations

There are several different numerical preference representation structures, including multiplicative preference relations, fuzzy and interval fuzzy preference relations [46-48]. Herrera-Viedma et al. [49] designed the transformation laws among several distinct numerical preference relations. This study assumes that the preference relations with LDAs are converted into fuzzy and interval fuzzy preference relations. The fuzzy and interval fuzzy preference relations are introduced below.

**Definition 10** [50, 51]. The matrix $F=(f_{ij})_{n\times n}$ is known as a fuzzy (or additive) preference relation, where $f_{ij} \geq 0$ denotes the preference intensity of the $x_i$ over $x_j$ and $f_{ij} + f_{ji} = 1$.

**Definition 11** [52]. The consistency level of fuzzy preference relation $F=(f_{ij})_{n\times n}$ is defined as follows:

$$CL(F)=\frac{2}{3n(n-1)(n-2)} \sum_{i,j,k=1,i \neq j \neq k} f_{ij} + f_{ji} - f_{ik} - f_{ki} - 0.5$$  \hspace{1cm} (5)

Clearly, $CL(F) \in [0, 1]$. $CL(F)=0$ indicates the consistency level of $F$ is completely acceptable; otherwise, the smaller $CL(F)$ value signifies the better consistency level of $F$.

Due to the complexity and uncertainty involved in real-world decision problems, sometimes it is unrealistic to acquire exact judgments. Thus, fuzzy preference relations are extended to interval fuzzy preference relations [47].

**Definition 12** [47]: The matrix $F=(f_{ij})_{n\times n}$ is defined as an interval fuzzy preference relation, where $F_{ij}=[f_{ij}^l, f_{ij}^u] \subseteq [0, 1]$ and $f_{ij}^l + f_{ij}^u = 1$ for $i, j = 1, 2, ..., n$.

**Definition 13:** An interval fuzzy preference relation $F=(f_{ij})_{n\times n}$ is called additive consistent, if the following additive transitivity is satisfied

$$f_{ij}^l + f_{jk}^u + f_{ki}^u = f_{ij}^u + f_{jk}^l + f_{ki}^l; \hspace{1cm} i, j, k=1, 2, ..., n$$  \hspace{1cm} (6)

In terms of left and right limits of interval-valued preferences, additive transitivity is defined as

$$f_{ij}^l + f_{ij}^u - 0.5,$$  \hspace{1cm} (7)

and

$$f_{ij}^u + f_{ij}^u - 0.5.$$  \hspace{1cm} (8)

**Definition 14:** Let $F=(f_{ij})_{n\times n}$ be as the above, we define the consistency level of $F=(f_{ij})_{n\times n}$ as follows:

$$CL(F) = \frac{1}{3n(n-1)(n-2)} \sum_{i,j,k=1,i \neq j \neq k} \left( |f_{ij}^l + f_{ij}^u - f_{ij} - 0.5| + |f_{ij}^l + f_{ij}^u - f_{ij} - 0.5| \right)$$  \hspace{1cm} (9)

Clearly, $CL(F) \subseteq [0, 1]$. $CL(F)=0$ indicates the consistency level of $F$ is completely acceptable; otherwise, a smaller $CL(F)$ value signifies a better consistency level of the fuzzy preference relation $F$.

III. THE LINGUISTIC DISTRIBUTION-BASED OPTIMIZATION APPROACH

As analyzed in the Introduction section, when transforming CLEs information into LDAs or interval LDAs, it is difficult to obtain possibilities that describe the occurring possibilities of the linguistic terms in the LDAs or interval LDAs. To deal with this issue, this section presents a consistency-driven optimization model to handle CLEs information.

A. Generate preference relations with LDAs from preference relations with CLEs

Let $N \subseteq \{1, 2, ..., n\}$. Recall that $A=(a_{ij})_{n\times n}$ is a preference relation with CLEs, and $C=(c_{ij})_{n\times n}$ is the preference relation with LDAs transformed from $A=(a_{ij})_{n\times n}$. By employing the following method, $C=(c_{ij})_{n\times n}$ can be converted into $F=(f_{ij})_{n\times n}$, where

$$p_{ij} = E(c_{ij}) = \sum_{s \in S} NS(c_{ij}) \cdot p_{ij}, \hspace{1cm} i, j \in N,$$  \hspace{1cm} (10)

where

$$p_{ij} \geq 0, \hspace{1cm} i, j \in N; \hspace{1cm} t=1, 2, ..., #c_{ij}$$  \hspace{1cm} (11)

and

$$\sum_{i=1}^{#c_{ij}} p_{ij} = 1, \hspace{1cm} i, j \in N.$$  \hspace{1cm} (12)

Naturally, we hope the consistency level of $F=(f_{ij})_{n\times n}$ that transformed from $C=(c_{ij})_{n\times n}$ is as high as possible, that is

$$\min F \left( CL(F) = \frac{2}{3n(n-1)(n-2)} \sum_{i,j,k=1,i \neq j \neq k} |f_{ij}^l + f_{ij}^u - f_{ij} - 0.5| \right)$$

$$= \frac{2}{3n(n-1)(n-2)} \sum_{i,j,k=1,i \neq j \neq k} \left( |\sum_{s \in S} NS(c_{ij}) \cdot p_{ij} + \sum_{s \in S} NS(c_{ij}) \cdot p_{ij} - 0.5| \right)$$

$$- \sum_{s \in S} NS(c_{ij}) \cdot p_{ij}$$  \hspace{1cm} (13)

The following optimization model is proposed according to the above analysis:
In model (14), \( p_{ij} \) (i, j \in N; t = 1, ..., #c) are decision variables. The optimal solutions to \( p_{ij} \) (i, j \in N; t = 1, ..., #c) can be produced by solving model (14). Further, a preference relation with LDAs (i.e. \( C = (c_{ij})_{n,n} \)) and a fuzzy preference relation (i.e. \( F = (f_{ij})_{n,n} \)) associated with \( A = (a_{ij})_{n,n} \) can be generated. If the consistency level of \( F = (f_{ij})_{n,n} \) is unacceptable, then the associated preference relation with CLEs \( A \) should be adjusted, until the consistency level of \( F = (f_{ij})_{n,n} \) is acceptable. There are many approaches to improve the consistency level of the preference relation with CLEs [28].

Clearly, model (14) is a non-linear programming model, which is difficult to solve. To solve model (14) easily, we present a theorem (Theorem 1) to convert it into a linear programming model.

**Theorem 1:** The following linear programming model can be generated from model (14):

\[
\begin{align*}
\min \ cI(F) & = \frac{2}{3(n-1)(n-2)} \sum_{i,j,k: i \neq j \neq k} \left| f_{ij} + f_{jk} - 0.5 \right| \\
\text{s.t.} & \quad p_{ij} \geq 0, \quad i, j \in N \\
& \quad \sum_{i} p_{ij} = 1, \quad i \in N \\
& \quad f_{ij} + f_{jk} - f_{ik} - 0.5 \leq u_{ijk}, \quad i, j, k \in N \\
& \quad f_{ij} \geq 0, \quad i, j \in N \\
& \quad f_{ij} \geq 0, \quad i, j \in N \\
& \quad f_{ij} \geq 0, \quad i, j \in N \\
& \quad f_{ij} \geq 0, \quad i, j \in N \\
\end{align*}
\]

**Proof:** In model (15), \( f_{ij} + f_{ji} - f_{ik} \leq 0.5 \leq u_{ijk} \) and \(- f_{ik} + f_{ji} + 0.5 \leq u_{ijk} \) guarantee that \( | f_{ij} + f_{ji} - f_{ik} - 0.5 | \leq u_{ijk} \). As a result, model (14) can be equally converted into linear programming model (15).

**B. Generate preference relation with interval LDAs from preference relations with CLEs**

Let \( A = (a_{ij})_{n,n} \) be defined as the above, and recall that \( D = (d_{ij})_{n,n} \) is the preference relation with interval LDAs transformed from \( A \). By using the following approach, \( D = (d_{ij})_{n,n} \) can be converted into \( F = (f_{ij})_{n,n} = (f_{ij}, f_{ji})_{n,n} \), where \( (f_{ij}, f_{ji}) = E(d_{ij}) = \sum_{i} a_{ij} d_{ij} \), \( i, j \in N \).

Meanwhile,

\[
\begin{align*}
0 \leq p_{ij} & \leq p_{ji} \leq 1, \quad i, j \in N; \ t = 1, 2, ..., #d \\
\sum_{i,j} p_{ij} & = 1, \quad i, j \in N; \ s = 1, 2, ..., #d \\
\end{align*}
\]

Thus, we present the following optimization model,

\[
\begin{align*}
\min \ cI(F) & = \frac{1}{3(n-1)(n-2)} \sum_{i,j,k: i \neq j \neq k} \left| f_{ij} + f_{ji} - f_{ik} - 0.5 \right| + \left| f_{ij} + f_{ji} - f_{ik} - 0.5 \right| \\
\text{s.t.} & \quad p_{ij} \geq 0, \quad i, j \in N \\
& \quad \sum_{i} p_{ij} = 1, \quad i \in N \\
& \quad f_{ij} + f_{ji} - f_{ik} - 0.5 \leq u_{ijk}, \quad i, j, k \in N \\
& \quad - f_{ik} + f_{ji} + 0.5 \leq u_{ijk}, \quad i, j, k \in N; \ i \neq j \neq k \\
& \quad f_{ij} + f_{ji} - f_{ik} - 0.5 \leq u_{ijk}, \quad i, j, k \in N; \ i \neq j \neq k \\
\end{align*}
\]

**Proof:** In model (20), \( p_{ij} \) and \( p_{ji} \) (i, j \in N; t = 1, 2, ..., #d) are decision variables. Solving model (20), we can obtain the optimum solutions to \( p_{ij} \) and \( p_{ji} \) (i, j \in N; t = 1, 2, ... , #d). Further, preference relation with interval LDAs (i.e. \( D \) ) and interval fuzzy preference relation (i.e. \( F \) ) associated with \( A = (a_{ij})_{n,n} \) can be generated.

Model (20) is also a non-linear programming model, and we propose a theorem (Theorem 2) to decrease the solving complexity of model (20).

**Theorem 2:** The following linear programming model (i.e., model (21)) is equivalent to model (20).

\[
\begin{align*}
\min \ cI(F) & = \frac{2}{3(n-1)(n-2)} \sum_{i,j,k: i \neq j \neq k} \left| a_{ij} + v_{ijk} \right| \\
\text{s.t.} & \quad p_{ij} \geq 0, \quad i, j \in N \\
& \quad \sum_{i} p_{ij} = 1, \quad i \in N \\
& \quad a_{ij} + v_{ijk} \leq 1, \quad i, j \in N; \ s = 1, 2, ..., #d \\
& \quad p_{ij} \quad + \sum_{s=1}^{#d} p_{ij}^{s} \geq 1, \quad i, j \in N; \ s = 1, 2, ..., #d \\
\end{align*}
\]

**Proof:** In model (21), \( a_{ij} + v_{ijk} \leq 1 \) and \(- a_{ij} - v_{ijk} \leq 1 \) guarantee that \( | a_{ij} + v_{ijk} - 1 | \leq v_{ijk} \) and \( | a_{ij} + v_{ijk} + 1 | \leq v_{ijk} \). The objective function achieves the optimal value only when \( | a_{ij} + v_{ijk} - 1 | = v_{ijk} \). As a result, model (21) is equivalent to model (20).

**IV. LARGE-SCALE GDM FRAMEWORK BASED ON LINGUISTIC DISTRIBUTION-BASED OPTIMIZATION APPROACH**

Classically, GDM problems have been solved by a small number of decision makers, and the number of decision makers in the most effective GDM context is less than 7 (see [53]). As introduced in section I, research on large-scale GDM has attracted wide attention in decision-making area due to the growing need of undertaking large-group decision making processes in various
real-life domains (see [40]). Usually, when the number of decision makers in a GDM problem exceeds 11, the GDM problem can be defined as a large-scale GDM problem (see [54, 55]). In this section, a large-scale GDM framework based on linguistic distribution-based optimization approach is presented (see Fig. 1).

This large-scale GDM framework consists of the following steps:

(1) Generating numerical preference relations from the preference relations with CLEs.

In this step, the linguistic distribution-based approach is used to produce numerical preference relations from the preference relations with CLEs. For those obtained numerical preference relations with unacceptable consistency levels, their associated preference relations with CLEs should be adjusted. There are many approaches to improve the preference relations with CLEs [8, 28].

Let \( A^{(k)} = (a^{(k)}_{ij})_{mn} \) be the preference relation with CLEs associated with \( e_k \). Using the linguistic distribution-based approach presented in Section III.A generates the preference relation with LDAs (\( C^{(k)} = (c^{(k)}_{ij})_{mn} \)) and fuzzy preference relation (\( F^{(k)} = (f^{(k)}_{ij})_{mn} \)) associated with \( A^{(k)} = (a^{(k)}_{ij})_{mn} \). Applying the linguistic distribution-based approach presented in Section III.B produces the preference relation with interval LDAs (\( D^{(k)} \)) and interval fuzzy preference relation (\( \overline{F}^{(k)} \)) associated with \( A^{(k)} = (a^{(k)}_{ij})_{mn} \).

![Fig. 1. Scheme of the proposed large-scale GDM framework](image)

(2) Preference clustering and aggregation.

Here, a preference clustering approach is put forward to classify individuals into a number of small clusters. Then, the large group’s preference is generated by aggregating individual preferences.

In large-scale GDM, where analyzing all the individual preferences at large-group level can become a complex and time-consuming task due to a large amount of available information, preference clustering is an effective way to analyze and manage preferences associated with the members of the large group. Several clustering approaches for carrying out preference clustering have been reported [35, 36, 54, 56]. The use of the preference clustering approach does not change any of the essence of the proposed decision framework. The preference clustering approach that presented in [56] is employed in this study.

Let \( R = (r_{ij})_{mn} \) be a similarity matrix among decision makers \( E \), where \( r_{ij} \in [0, 1] \) denotes the similarity degree between decision makers \( e_p \) and \( e_q \) and it can be calculated as below:

**Case A:** \( F^{(k)} = (f^{(k)}_{ij})_{mn} \) are fuzzy preference relations

\[
r_{ij} = 1 - \frac{1}{n(n-1)} \sum_{p=1}^{n} \sum_{q=1, q \neq p}^{n} \left| f^{(k)}_{pq} - f^{(k)}_{qp} \right|
\]

**Case B:** \( F^{(k)} = (f^{(k)}_{ij})_{mn} \) are interval fuzzy preference relations

\[
r_{ij} = 1 - \frac{1}{2n(n-1)} \sum_{p=1}^{n} \sum_{q=1, q \neq p}^{n} \left( | f^{(k-1)}_{pq} - f^{(k-1)}_{qp} | + | f^{(k)}_{pq} - f^{(k)}_{qp} | \right)
\]

The larger \( r_{ij} \) value implies the higher similarity between \( e_p \) and \( e_q \).

**Definition 15:** Let \( \alpha \) be a parameter, \( 0 < \alpha < 1 \). If \( r_{ij} \geq \alpha \), then \( e_i \) is the direct neighbor of \( e_j \). The set of all direct neighbors of decision maker \( e_i \) is denoted as the set \( DN_i \), i.e.,\( DN_i = \{ e_j | r_{ij} \geq \alpha \} \).

**Definition 16:** Let \( e_1, e_2, \ldots, e_T \) be \( T \) decision makers, if decision maker \( e_T \) is only a direct neighbor of \( e_{T-1} \), decision maker \( e_k \) is only a direct neighbor of the decision maker \( e_{k-1} \), and decision maker \( e_1 \) is the direct neighbors of \( e_{k-1} \) and \( e_{k+1} \) (\( 1 < k < T \)), then \( e_1, e_2, \ldots, e_T \) are all the indirect neighbors of decision maker \( e_1 \). In this case, the all indirect neighbors of decision maker \( e_1 \) are denoted as \( ID_1 \).

Let \( \beta \ (\beta \in (0, 1]) \) be a parameter, which is used to judge whether a decision maker belongs to a cluster or not. For cluster \( CE \) and decision maker \( e_i \) that needs to be classified, if the proportion of the neighbors of \( e_i \) in \( CE \) is larger than or equal to \( \beta \), then \( e_i \) can be classified into cluster \( CE \). Based on this idea, a clustering method is presented, which is described as Algorithm I.

**Algorithm I**

**INPUT:** Decision makers that need to be classified, \( E \). The similarity degree matrix among decision makers \( R \), the parameters \( \alpha \) and \( \beta \).

**OUTPUT:** The clusters of decision makers.

**BEGIN:** Let \( t = 1 \);

While \( E \neq \emptyset \)

Create an empty cluster \( CE_t \). Choose a decision maker \( e \) from \( E \), and put it into \( CE_t \).

Delete \( e \) from \( E \).

For each decision maker \( e_i \in E \)

**IF** \( r_{ij} \geq \alpha \)

For each decision maker \( e_j \in CE_t \)


\( q_j = 1 \).

ELSE
\( q_j = 0 \).
END

IF \((e_i \in (O_{G1} \cup J_{G1}))) \& \& \left( \sum_{e_i \in e_{CE}} \geq \beta \right), \) then \( CE_i = CE_j \cup \{e_i\} \) and \( E = E \setminus \{e_i\} \)
END
END

Output cluster \( CE_i \).

\( t = t + 1 \).

END

**Theorem 3:** The time complexity of Algorithm I is no more than \( O(n^2) \).

**Proof.** We consider the worst case: in each iteration \( t \), only one decision maker \( e_i \) enters into the cluster \( CE_i \), then the frequency \( g(n) \) of the Algorithm I is determined as:

\[
g(n) = 1(n-1) + 2(n-2) + \ldots + n \left( \frac{n}{2} - \frac{n}{2} \right) + \ldots + (n-1)! \text{ if } n \text{ is an even number.}
\]

\[
g(n) = 1(n-1) + 2(n-2) + \ldots + n \left( \frac{n-1}{2} - \frac{n-1}{2} \right) + \ldots + (n-1)! \text{ if } n \text{ is an odd number.}
\]

For both two cases on the values of \( n \), we can easily obtain:

\( O(g(n)) = O(n^2) \).

This completes the proof of Theorem 3.

Using Algorithm I, a large-scale group can be divided into \( K \) \((1 \leq K \leq m)\) clusters, that are \( CE_1, CE_2, \ldots, CE_K \). According to the principle that the larger-scale clusters should be assigned larger weights, the weights of the clusters are determined. Let \( \lambda_i \) be the weight of cluster \( CE_i \). Without loss of generality, we use the following way to calculate \( \lambda_i \), i.e.,

\[ \lambda_i = \frac{|CE_i|}{\sum_{i \in CE} |CE_i|^2}, \]

where \(|CE_i|\) is the number of individuals in \( CE_i \).

Decision makers in the same cluster can be assigned the same weight because they have the similar individual preference information and the individual concerns on alternatives. Therefore, the weight of decision maker \( e_i \) in cluster \( CE_k \) is calculated as

\[ \theta_i = \frac{1}{|CE_k|}. \]

Let \( F^{(k)} = (f^{(k)}_{ij})_{i,j \in CE_k} \) be as above, let \( F^{(c, k)} = (f^{(c, k)}_{ij})_{i,j \in CE_k} \) be the collective numerical preference relation of cluster \( CE_k \), where

**Case A:** \( F^{(k)} \) are fuzzy preference relations,

\[ f^{(c, k)}_{ij} = \sum_{e_i \in CE_k} \theta_i \cdot f^{(k)}_{ij} \]

**Case B:** \( F^{(k)} \) are interval fuzzy preference relations,

\[
\begin{align*}
    f^{(c, k)}_{ij} &= \sum_{e_i \in CE_k} \theta_i \cdot f^{(k)}_{ij}, \\
    f^{(c, k)}_{ij} &= \sum_{e_i \in CE_k} \theta_i \cdot f^{(k)}_{ij}
\end{align*}
\]

Let \( F^{(c)} = (f^{(c)}_{ij})_{i,j \in CE_k} \) be the large group’s preference, where \( f^{(c)}_{ij} \) is computed as follows:

**Case A:** \( F^{(k)} \) are fuzzy preference relations,

\[ f^{(c)}_{ij} = \sum_{k} \lambda_k \cdot f^{(k)}_{ij} \]

**Case B:** \( F^{(k)} \) are interval fuzzy preference relations,

\[
\begin{align*}
    f^{(c, k)}_{ij} &= \sum_{k} \lambda_k \cdot f^{(k)}_{ij}, \\
    f^{(c, k)}_{ij} &= \sum_{k} \lambda_k \cdot f^{(k)}_{ij}
\end{align*}
\]

(3) Exploitation process to generate the ranking of alternatives from large group’s preferences

In this process, the alternatives are ranked from best to worst based on the large group’s numerical preference relation.

Here, the collective preference vector \( PV^{(c)}=(p_{v1}^{(c)}, \ldots, p_{vn}^{(c)}) \) is produced from \( F^{(c)} \) using the following way:

**Case A:** \( F^{(c)} \) is fuzzy preference relation

\[ PV^{(c)} = \sum_{j=1}^{n} w_{j} \cdot f^{(c)}_{ij} \]

**Case B:** \( F^{(c)} \) is interval fuzzy preference relation

\[ PV^{(c)} = \left[ \sum_{j=1}^{n} w_{j} \cdot f^{(c, k)}_{ij}, \sum_{j=1}^{n} w_{j} \cdot f^{(c, k)}_{ij} \right] \]

and \( w = (w_1, \ldots, w_n)^T \) is a weight vector that satisfies \( 0 \leq w_j \leq 1 \) and \( \sum_{j=1}^{n} w_j = 1 \).

From the values of \( PV^{(c)} \), alternatives \( \{x_1, x_2, \ldots, x_n\} \) can be ranked from best to worst. The larger \( PV^{(c)} \) value, the better alternative \( x_i \) is. Using comparison laws of interval numbers presented in Wang et al. [57], the ranking of alternatives can be generated.

V. APPLICATION EXAMPLE AND COMPARISON ANALYSIS

In this section, the proposed large-scale GDM framework is applied in the selection of a sustainable disinfection technique for wastewater reuse projects. Furthermore, a comparison analysis is conducted to validate the effectiveness of the proposal.

A. Application example

During the last few years, we have witnessed growing water stress, both in terms of water scarcity and quality deterioration. Looking for a more efficient use of water resources, including a more widespread acceptance of water reuse practices, is a key issue for relieving the water stress. Particularly, selecting a sustainable treatment for wastewater reuse facilities presents a serious challenge for wastewater reuse project managers as well as a large number of stakeholders and actors with highly diverse expertise and background in the decision-making process (Curiel-Esparza et al. [42]). This situation can be modeled as a large-scale GDM framework.

A city is faced with a problem of water shortage and pollution, and the wastewater reuse project managers of this city invite twenty decision makers (denoted as \( \{e_1, e_2, \ldots, e_{20}\} \)) to evaluate wastewater treatment technologies. These decision makers from different departments include experts of the water resources bureau, professors of wastewater resource management, local resident representatives, and experts of third-party water
technology management company. After a pre-evaluation, the following four technologies for the disinfection of treated wastewater are selected for further discussion and evaluation:

**Chlorination (CHL).** Water chlorination is the process of adding chlorine (Cl\(_2\)) or hypochlorite to water. This method is used to kill certain bacteria and other microbes in tap water as chlorine is highly toxic. The required quantity depends on the water and on the disinfection requirements. Chlorine is a disinfectant with strong disinfection capability and low cost, so it is widely applied around the world.

**Ozonization (OZO).** Ozone is one of the most powerful disinfectants, due to its high oxidizing capacity, suitable for the treatment of water. Ozone emerged as a popular alternative to chlorine. Its greatest advantage is that not produce unwanted by-products, since ozone becomes oxygen. Disinfection by ozone has increased popularity in recent years

**Ultraviolet radiation (UVR) technology** is one of the most applied in wastewater treatment plants, as tertiary treatment for disinfection of effluent. This is because of its ability to inactivate a wide range of pathogens without the formation of harmful byproducts. In ultraviolet disinfection, water is exposed to shortwave ultraviolet light. This is an effective germicide and does not affect the water quality. This is a technology that applies both to drinking water treatment and disinfection of treated wastewater.

**Membrane filtration (MFI)** can be used instead of the decanter to separate solids from the liquid. In wastewater treatments, membrane filtration can be defined as a separation process that uses semi-permeable membranes to divide the treated wastewater into two portions: a permeate with the material passing through the membranes, and a retentate consisting of residues that do not pass through the filter. The main types of membrane filtration are: microfiltration, ultrafiltration, nanofiltration, and reverse osmosis.

For the sake of convenience, the technologies CHL, OZO, UVR, and MFI are denoted as \( x_1 \), \( x_2 \), \( x_3 \), and \( x_4 \), respectively. For the detailed information regarding these four technologies, please refer to Curiel-Esparza et al. [42]. Due to the complexity and uncertainty of the technology evaluation process, the twenty decision makers find it difficult to provide their opinions over the four identified technologies in an accurate manner. They use preference relations with CLEs to express their opinions, and a nine-grade linguistic term set \( S \) is used by them, which is provided below:

\[ S = \{s_0 = \text{absolutely worse}, s_1 = \text{much worse}, s_2 = \text{worse}, s_3 = \text{slightly worse}, s_4 = \text{indifferent}, s_5 = \text{slightly better}, s_6 = \text{better}, s_7 = \text{much better}, s_8 = \text{absolutely better} \} \]

The numerical scales are set as: \( NS(s_0) = 0 \), \( NS(s_1) = 1/8 \), \( NS(s_2) = 2/8 \), \( NS(s_3) = 3/8 \), \( NS(s_4) = 4/8 \), \( NS(s_5) = 5/8 \), \( NS(s_6) = 6/8 \), \( NS(s_7) = 7/8 \), and \( NS(s_8) = 1 \).

The preference relations with CLEs over the four technologies provided by the twenty decision makers are provided below:
In the following, we use the proposed large-scale framework to assist the twenty decision makers in selecting the most suitable disinfection technology. First, we convert preference relations with CLEs $A_k$ ($k = 1, 2, ..., 20$) into preference relations with interval LDAs $D_k$ using the proposed linguistic distribution-based optimization model. It should be noted that we can also transform preference relations with CLEs into preference relations with LDAs in this example, and the relevant results are not provided due to space limitations. Then, the preference clustering and aggregation is used to generate the whole preference of the twenty decision makers. Finally, the exploitation process is employed to generate the ranking of four technologies based on the obtained large group preference.

(1) Applying the proposed linguistic distribution-based optimization model

Using model (20), we can obtain preference relations with interval LDAs $D_k$ from $A_k$ ($k = 1, 2, ..., 20$), and they are provided below:

$$D_1 = \begin{bmatrix}
null & null & null & null & null \\
null & null & null & null & null \\
null & null & null & null & null \\
null & null & null & null & null \\
null & null & null & null & null \\
\end{bmatrix},$$

$$D_2 = \begin{bmatrix}
null & \{s_i^{[0.006,0.0778]}\} & \{s_i^{[0.9222,0.9992]}\} & null & null \\
null & \{s_i^{[0.4983,0.5151]}\} & \{s_i^{[0.05]}\} & \{s_i^{[0.95]}\} & null \\
null & \{s_i^{[0.4885,0.5602]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.05]}\} & \{s_i^{[0.95]}\} \\
null & null & \{s_i^{[0.95]}\} & \{s_i^{[0.05]}\} & \{s_i^{[0.95]}\} \\
null & null & null & \{s_i^{[0.95]}\} & \{s_i^{[0.05]}\} \\
\end{bmatrix},$$

$$D_3 = \begin{bmatrix}
null & \{s_i^{[0.4783,0.5399]}\} & \{s_i^{[0.4961,0.5217]}\} & null & null \\
null & \{s_i^{[0.4865,0.8678]}\} & \{s_i^{[0.1322,0.1935]}\} & \{s_i^{[0.4925,0.5485]}\} & null \\
null & null & \{s_i^{[0.95]}\} & \{s_i^{[0.05]}\} & \{s_i^{[0.95]}\} \\
null & null & null & \{s_i^{[0.95]}\} & \{s_i^{[0.05]}\} \\
null & null & null & null & \{s_i^{[0.95]}\} \\
\end{bmatrix},$$

$$D_4 = \begin{bmatrix}
null & \{s_i^{[0.35,0.4]}\} & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
null & \{s_i^{[0.35,0.4]}\} & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
null & \{s_i^{[0.95]}\} & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
null & \{s_i^{[0.95]}\} & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
null & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
\end{bmatrix},$$

$$D_5 = \begin{bmatrix}
null & \{s_i^{[0.341,0.4454]}\} & \{s_i^{[0.3801,0.3789]}\} & \{s_i^{[0.2282,0.3206]}\} & \{s_i^{[0.2133,0.3289]}\} \\
null & \{s_i^{[0.35,0.4]}\} & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
null & \{s_i^{[0.95]}\} & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
null & \{s_i^{[0.95]}\} & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
null & \{s_i^{[0,0.05]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} & \{s_i^{[0.95]}\} \\
\end{bmatrix}. $$
Meanwhile, the interval fuzzy preference relations \( F^k \) associated with \( D^k \) \((k=1, 2, \ldots, 20)\) are generated using Eq. (16). Here, we take \( f_{12} = [f_{12}^{0.5}, f_{12}^{1.0}] \) as an example to show the following computation process: \( f_{12}^{0.5} = p_{12}^{0.5} \times \text{NS}(s_i) + p_{12}^{1.0} \times \text{NS}(s_i) = 0 \times (3/8) + 0.95 \times (4/8) = 0.475 \) and \( f_{12}^{1.0} = p_{12}^{0.5} \times \text{NS}(s_i) + p_{12}^{1.0} \times \text{NS}(s_i) = 0.05 \times (3/8) + 1 \times (4/8) = 0.5188 \).
According to Eq. (23), we obtain the weights associated with the three clusters: \( \lambda_1 = 0.526 \), \( \lambda_2 = 0.4156 \), and \( \lambda_3 = 0.0584 \). Further, using Eq. (28) produces the large group’s interval fuzzy preference relation \( F^{(1)} = (f^{(1)}_{ij})_{k=1} \), where

\[
f^{(1)}_{ij} = \lambda_1 \times f^{(0)}_{ij} + \lambda_2 \times f^{(3)}_{ij} + \lambda_3 \times f^{(2)}_{ij},
\]

and we hope that the ranking of the four technologies employing the approach presented by Wang et al. [57], that is \( x_2 > x_3 > x_4 > x_1 \).

**Remark 1**: In the extant literature the examples with 12–20 decision makers are often utilized to show the application process of the large-scale GDM frameworks due to space limitations, and they will not violate the basic assumption of large-scale GDM because a GDM is considered the large-scale GDM problem when the number of decision makers in the GDM problem exceeds 11 (see [54, 55]). For instance, the large-scale GDM examples with 20 and 15 decision makers are respectively considered in the Xu et al. [54] and Zhu et al. [58]. Meanwhile, we need to emphasize that our proposal is a general framework for large-scale GDM. When the number of decision makers is large enough (e.g., 100 or 1000), our proposal is still applicable.

### B. Comparison analysis

In the existing approach, the preference relations with CLEs are converted into HFLPRs. In our proposal, the preference relations with CLEs are converted into preference relations with LDAs or interval LDAs. The consistency is the basis of the preference relations, and we hope that the relevant preference relations transformed from the preference relations with CLEs are as consistent as possible. Thus, the consistency index of the transformed preference relations is an important criterion to evaluate the performance of different approaches.

Here, we compare the consistency levels of the HFLPRs and preference relations with LDAs transformed from the preference relations with CLEs, respectively. Particularly, when measuring the consistency level of the HFLPR, it is transformed into a preference relation with LDAs by assigning equal possibility degrees of linguistic terms in each HFLTS. The comparison

Particularly, the consistency levels of \( F^{(k)} (k = 1, 2, \ldots, 20) \) can be generated: \( CL(F^{(1)}) = 0.0229 \), \( CL(F^{(2)}) = 0.0297 \), \( CL(F^{(3)}) = 0.0439 \), \( CL(F^{(4)}) = 0.0445 \), \( CL(F^{(5)}) = 0.0236 \), \( CL(F^{(6)}) = 0.0432 \), \( CL(F^{(7)}) = 0.044 \), \( CL(F^{(8)}) = 0.0247 \), \( CL(F^{(9)}) = 0.003 \), \( CL(F^{(10)}) = 0.0029 \), \( CL(F^{(11)}) = 0 \), \( CL(F^{(12)}) = 0 \), \( CL(F^{(13)}) = 0.0219 \), \( CL(F^{(14)}) = 0 \), \( CL(F^{(15)}) = 0 \), \( CL(F^{(16)}) = 0 \), \( CL(F^{(17)}) = 0 \), \( CL(F^{(18)}) = 0 \), \( CL(F^{(19)}) = 0.0354 \), and \( CL(F^{(20)}) = 0.0448 \).

Here, the consistency threshold is set as 0.5. The consistency levels of all interval fuzzy preference relations are acceptable.

(2) Preference clustering and aggregation

When setting \( \alpha = 0.8 \) and \( \beta = 0.85 \), three clusters of decision makers are yielded employing the proposed preference clustering approach:

\[
CE_1 = \{e_1, e_2, e_3, e_5, e_{12}, e_{16}, e_{17}, e_{18}\}, CE_2 = \{e_4, e_5, e_8, e_9, e_{13}, e_{14}, e_{15}, e_{20}\}, \text{ and } CE_3 = \{e_6, e_7, e_{10}\}.
\]

Based on Eq. (24), we obtain the weights of all decision makers, where \( \theta_1 = 1/9 \) if \( e_k \in CE_1 \); \( \theta_2 = 1/8 \) if \( e_k \in CE_2 \); \( \theta_3 = 1/3 \) if \( e_k \in CE_3 \). According to Eq. (26), we obtain the interval fuzzy preference relations \( F^{(3, z)} = (f^{(3, z)}_{ij})_{k=1} (z = 1, 2, 3) \) associated with the three clusters, where

\[
f^{(3, z)}_{ij} = \sum_{k=1}^{3} \theta_k \times f^{(3, z)}_{ij}.
\]
results regarding the preference relations with CLEs $A^{(k)}$ ($k = 1, 2, \ldots, 20$) used in section V.A are offered in Fig.2.

From Fig. 2, we can see that the consistency levels of the transformed preference relations in our proposal are better than those in the existing approach, which shows that the linguistic-distribution optimization approach has a good performance in the criterion of consistency level.

VI. CONCLUSION

This paper presented a linguistic distribution-based optimization approach for large-scale GDM with comparative linguistic information. The main contributions presented are:

1. A pioneering linguistic distribution-based optimization approach is presented to transform CLEs into LDAs, and preference relations with CLEs are then converted into preference relations with LDAs.

2. The above linguistic distribution-based optimization approach is further extended into an interval-valued context to increase its flexibility in which CLEs are transformed into interval LDAs.

3. Based on the linguistic distribution-based optimization approach, a large-scale GDM framework with CLEs information is developed.

4. A case study on a large-scale GDM problem about selecting sustainable disinfection technique for wastewater reuse projects, along with a comparison study that shows the effectiveness of the proposed large-scale GDM framework.

Meanwhile, we point out some future research directions:

1. Consensus building is a hot research topic in decision-making field [4, 59-65]. Therefore, investigating the consensus issue in large-scale GDM problem based on comparative linguistic expressions is a very interesting research direction for the future.

2. The social trust relationships among decision makers play a key role in the large-scale GDM [66-68]. So, we also consider that it would be important to examine the social trust relationships in the proposed large-scale GDM framework.

3. The solution for the large-scale GDM involves not only mathematical issues but also the psychology issues of decision makers [38, 69]. It could be an interesting research topic to incorporate the psychology issues in the proposed large-scale GDM framework.

REFERENCES


J. Wu, R. Xiong, F. Chiclana, Uninorm trust propagation and aggregation methods for group decision making in social network with four tuple information, Knowledge-Based Systems 96 (2016) 29-39.

Yucheng Dong received the B.S. and M.S. degrees both in mathematics from Chongqing University, Chongqing, China, in 2002 and 2004, respectively, and the Ph.D. degree in management from Xi’an Jiaotong University, Xi’an, China, in 2008. He is currently a Professor at the Business School, Sichuan University, Chengdu, China. He is currently a Professor at the Business School, Sichuan University, Chengdu, China. His current research interests include consensus process, computing with words, opinion dynamics, and social network decision making. He has published more than 70 international journal papers in Decision Support Systems, European Journal of Operational Research, IEEE Transactions on Big Data, IEEE Transactions on Cybernetics, IEEE Transactions on Fuzzy Systems, IEEE Transactions on Systems, Man, and Cybernetics, Omega, among others.

Dr. Dong is a member of the editorial board of Information Fusion, an Area Editor of Computers & Industrial Engineering, and an Associate Editor of Group Decision and Negotiation.