Counted Linguistic Variable in Decision-Making

Wu-E Yang, Xin-Fan Wang, and Jian-Qiang Wang

Abstract

To help decision-makers better understand the decision procedure and result is a major objective of decision analysis. In linguistic decision, symbolic scheme thereby is more attractive than set-based scheme. The counted (expanded) linguistic variable (CLV/CELV) model is proposed for representing and aggregating linguistic decision information. The model is developed from the ordinary linguistic variable by attaching a counting label. This label separately contains the quantitative information that may influence the comparison and aggregation of linguistic variables. Using the CLVs (CELVs) and their operations for computing with words, the linguistic information can be handled more intelligibly.

Keywords: Computing with words, information fusion, linguistic modeling.

1. Introduction

Linguistic variables [1] are words or sentences in a natural or artificial language. Such variables are usually taken from a set of linguistic terms. The semantics of these terms is given early and pre-ordered as a scale. This linguistic ordered scale is more expressive in decision-making than a simple ordered scale because there exist the meaningful landmarks [2]. Much research attention thereby was attracted to use linguistic variables to express decision information [3-6]. These decision problems are called as linguistic decision problems.

The linguistic decision implies the reasoning, computing, and other operations with linguistic information, namely computing with words (CWW) [7, 8]. Fig. 1 illustrates the procedure of a linguistic decision [9]. We can find that the CWW is the key process during this procedure.

There are two main CWW schemes in literatures. First is the set-based scheme, where linguistic variables are represented by pre-determined fuzzy sets [10-12], namely translating linguistic variables to fuzzy sets. Different fuzzy sets are used in this scheme [13-16]. Translating a linguistic variable into which type of fuzzy set remains debates. Moreover, the decision result highly depends on the defuzzication procedure. Second is the symbolic scheme. This scheme directly computes on linguistic labels [17-19]. Compare to the set-based scheme, the debatable transformation process is avoided in the symbolic scheme. Therefore, the linguistic decision models in symbolic CWW frame are rapidly developed in recent years [20, 21].

Rodríguez and Martínez [9] analyzed the symbolic CWW scheme by comparing three typical symbolic computing models, i.e., the virtual linguistic model [22, 23], the 2-tuple linguistic model [17, 24], and the proportional 2-tuple linguistic model [25]. They were analyzed from the representation, comparison and computation, respectively. The analysis shows that a successful CWW model should keep the computation result accurate and understandable.

However, a considerable problem in the existing CWW is that aggregating linguistic variables sometimes does not result in a pre-identified linguistic term [2], especially when we aggregate linguistic variables without normalized weights. The CWW result would be hard to understand in these cases. This issue will decrease the efficiency of a linguistic decision model, because helping decision-makers to better understand the decision procedure and consequence is a major objective of decision analysis. But if we retranslate the CWW result to a pre-identified linguistic variable approximately, the information will be lost unavoidably [26].

This semantics confusion probably derives from that the existing models jumble the quantitative factors with...
qualitative aspect in decision. These quantitative factors, such as weights, are encountered during the information processing. And they may further influence comparing and aggregating linguistic variables.

The counted linguistic variable (CLV) model thereby is introduced in this paper. Generally, it is a symbolic model. An additive counting label is attached to the original linguistic variable. Thus, the qualitative part of decision information can be separated from the quantitative part, which is contained in the counting label. By operating on CLVs, the CWW outputs meaningful results without information losing.

Section 2 and 3 give the definition and some operation rules of CLVs. Several properties of the operation are also proposed. In Section 4, an example is established to illustrate the validity of the CLV model. We further discuss the model in Section 5 and conclude in Section 6.

2. Counted Linguistic Variables

The linguistic descriptors are pre-determined by the decision-maker (DM) in a linguistic ordered scale \( S = \{s_0, s_1, \ldots, s_g\} \), which is a set of \( g + 1 \) ordered terms. \( s_g < s_b \iff a < b \), \( a, b \in \{0, 1, \ldots, g\} \) [27]. A seven-term-scale could be 

\[
S = \{s_0: \text{VeryPoor}, s_1: \text{Poor}, s_2: \text{MediumPoor}, s_3: \text{Fair}, s_4: \text{MediumGood}, s_5: \text{Good}, s_6: \text{VeryGood}\}
\]

**Definition 1:** Let \( S_E = \{s_0, s_j, s_1, s_2, \ldots, s_{g-1}, s_g\} \) be an expanded linguistic ordered scale with \( 2g + 1 \) ordered terms. Comparing to the original linguistic scale \( S \), \( g \) intermediate terms have been inserted into each pair of original adjacent terms.

The way to build \( S_E \) is quite similar to that in the [28] by Herrera and Martinez. Every term in the \( S_E \) can be expressed in a uniform formation by \( [s_i, s_j] \), \( j \in \{i, i+1\} \), because \( s_i \) can also be denoted by \( [s_i, s_j] \). If denoting \( S_E = \{e_0, e_1, \ldots, e_{2g+1}\} \), there is \( e_a < e_b \iff a < b \), \( a, b \in \{0, 1, \ldots, 2g+1\} \).

To keep the evaluation structured, the inserted terms are not used in direct evaluation by DMS but to express the aggregated evaluations.

However, the inserted intermediate terms also have clear semantic meanings like the original terms. Each inserted term in fact is a linguistic interval [29] whose semantics lies between two original terms. For instance, provided an alternative is finally evaluated by \( [s_i: \text{Good}, s_{i+1}: \text{VeryGood}] \), the evaluation means that the performance of the alternative is higher than “Good” but lower than “VeryGood.”

The ordering feature of the scale does not be broken after the intersection. All the terms are still comparable and have a strict complete-order.

Note that: First, expanding the scale further in the same way may be not appropriate. This is because that, if so, the semantics of the further-added terms will be ambiguous. Second, adding linguistic intervals that cross one or more original terms, such as \([s_1, s_4]\) (crossing \( s_2 \) and \( s_3 \)), should also be avoided. This is because the positions of these terms in the scale are unclear, i.e., the strict ordering feature of the scale would not be preserved.

**Definition 2:** By attaching an additive label to the original linguistic variable, a counted linguistic variable (CLV) is an object like \( s_i^k \), \( k \geq 0 \). We call \( k \) the counting label. This variable means there are \( k \) linguistic terms \( s_i \), e.g., representing that \( k \) persons evaluate an alternative with the linguistic term \( s_i \).

A counted expanded linguistic variable (CELV) is denoted by \( [s_i^k, s_j^l], \quad k \geq 0, l \geq 0, \quad j \in \{i, i+1\} \). It means there are \( k + l \) objects whose semantics lies in the linguistic interval \([s_i, s_j]\), and equivalent to \( k \) linguistic terms \( s_i \) and \( l \) terms \( s_j \). We also denote it by \( [s_i^k, s_j^l] \) for short.

**Interpretation:** The CLV and CELV models are used for representing and aggregating the linguistic decision information more completely and meaningfully. These models are not for a particular decision-maker to express the direct evaluations. In practice, decision-makers still evaluate alternatives with original linguistic terms.

During the information processing, much quantitative factors will be encountered. These factors may influence comparing and aggregating linguistic variables in the subsequent process. The counting label separately keeps such quantitative aspect.

**Example 1:** 1) Five experts all believe the performance of an alternative is “s6: VeryGood.” The analyst can denote such information by \( s_6^5 \).

2) For evaluating the taste of a new species of apple, a sample set that contains ten fruits of such species is evaluated. Six of them are evaluated by an expert with “s5: Good” the others “s6: VeryGood.” This information can be denoted by \( [s_5^6, s_6^4] \).

**Definition 3:** Given any pair of CELVs \([s_i^k, s_j^l], \ j \in \{i, i+1\}\) and \([s_p^m, s_q^n], \ q \in \{p, p+1\}\), if

1) \( k + l \geq m + n \)
2) \( \frac{k+l}{k} \geq \frac{mp+nq}{k+m+n} \)

and at least a strong inequality holds, then \([s_i^k, s_j^l]\) is
superior than \([s_p^m, s_q^n]\), denoted by
\[
[s_p^k, s_q^l] > [s_p^m, s_q^n].
\]

If both of them are equal, then \([s_p^k, s_q^l]\) is equivalent to \([s_p^m, s_q^n]\), denoted by
\[
[s_p^k, s_q^l] = [s_p^m, s_q^n].
\]

When there is no above-mentioned relations between \([s_p^k, s_q^l]\) and \([s_p^m, s_q^n]\), we define that they are incomparable, denoted by
\[
[s_p^k, s_q^l] \perp [s_p^m, s_q^n].
\]

Particularly, from the definition, we have \([s_p^k, s_q^l] = [s_p^m, s_q^n]\) iff \(k + l = m + n \). Therefore, we can denote \(s_w^m = [s_p^k, s_q^l]\), \(a + b = v\).

**Interpretation:** The incomparability is necessary for comparing linguistic variables because linguistic information is naturally uncertain. We cannot get a certain comparison result in every framework. The counting labels reflect such frameworks partially.

For example, certainly, one big apple is heavier than one small apple, three big apples are heavier than two small apples and etc. However, whether “one big apple is heavier than two small apples”, or “two big apples are heavier than three small ones” is hard or even impossible to answer. The numerical part of these evaluations can be considered as counting labels.

### 3. Operations on CLVs

**Definition 4:** Given any pair of CELVs \([s_p^k, s_q^l]\), \(j \in \{i, i+1\}\), \([s_p^m, s_q^n]\), \(q \in \{p, p+1\}\) and a real number \(\rho\), we define

1) \([s_p^k, s_q^l] + [s_p^m, s_q^n] = [s_p^{(k+l+m+n)}, s_q^{(k+l+mp+nq)}]\), \(M(\frac{ki+lj+mp+nq}{k+l+m+n})\), \(C(\frac{ki+lj+mp+nq}{k+l+m+n})\)

where \(M(.)\), \(F(.)\) and \(C(.)\) represent the taking reminder, rounding down and rounding up operations, respectively.

Note that \(M((ra)/(rb)) = rM(a/b)\), such that we can not cancel the common factors of \((ki+lj+mp+nq)\) and \((k+l+m+n)\) during the calculation of \(M(.)\).

2) \(\rho(s_p^k, s_q^l) = [s_p^{\rho k}, s_q^{\rho l}]\)

Although we use the rounding functions \(F(.)\) and \(C(.)\), the operation will not result in the losing of information because the other necessary information is represented by the counting labels.

**Figure 2.** The aggregation procedure of two CLVs.

**Interpretation:** The basic idea for aggregating linguistic variables is that two linguistic terms can be equivalent to a pair of the central terms between them. If such term does not exist, then be equivalent to the two adjacent terms in the center.

For example, with a 7-term-scale, two experts evaluating an alternative “VeryGood” and “VeryPoor” respectively is equivalent to two experts evaluating this alternative “Fair.” In addition, a “Good” and a “VeryPoor” is equivalent to a “Fair” and a “MediumPoor.” Following this logic, \(x \) terms \(s_a\) and \(x \) terms \(s_b\) can rationally be equivalent to \(2x\) central terms between them even if \(x\) is not an integer.

The “add” operation of two CLEVs reflects such idea. We try to find the equivalent CLEV of two CLEVs as their aggregation result. Fig. 1 illustrates the three conversion steps for \([s_p^k, s_q^l] + [s_x^i, s_y^j] = [s_z^k, s_y^l]\) with a 7-term-scale. These steps are marked by I, II and III in the Fig. 2.

**Property 1:** The results of above operations are also CLEVs, namely the linguistic parts of the operation results fall in the expanded linguistic ordered scale \(S_E\).

**Proof:**

1) Since the linguistic parts of \([s_p^k, s_q^l]\) and \([s_x^i, s_y^j]\) are both \([s_z^k, s_y^l] \in S_E\), this property must hold for \(\rho(s_p^k, s_q^l) = [s_z^k, s_y^l]\)

2) To abridge the space, we also use “X” to denote \(\frac{ki+lj+mp+nq}{k+l+m+n} \) in the following proofs.

For the add operation, we need to prove that \(F(X)\) and \(C(X)\) are integers in \([0, g]\) and \(C(X) - F(X) \in [0, 1]\).

Obviously, since \(F(.)\) and \(C(.)\) represent the rounding down and rounding up operations, then \(F(X)\) and \(C(X)\) must be integers and there is \(C(X) - F(X) \in [0, 1]\).

Therefore, what we need to prove is
\[ 0 \leq F(X) \leq C(X) \leq g \]

Since \( i, j, p, q, k, l, m, n \geq 0 \), then \( X \geq 0 \), thus
\[ 0 \leq F(X) \leq C(X) \]

Further, notice that \( j \in [i, i+1] \) and \( q \in [p, p+1] \), therefore,
\[ X = \frac{ki + lj + mp + nq}{k + l + m + n} \leq \frac{ki + lj + mq + nq}{k + l + m + n} = \frac{(k + l)j + (m + n)q}{k + l + m + n} \]

Because \( j, q \in [0, g] \), there must be
\[ (k + l)j + (m + n)q \leq (k + l)g + (m + n)g = \frac{k + l + m + n}{k + l + m + n} = g \]

Hence, \( X \leq g \). Since \( g \) is a positive integer, therefore,
\[ F(X) \leq C(X) \leq g \]

That is, we get \( 0 \leq F(X) \leq C(X) \leq g \) and completed the proof of Property 1.

**Property 2:** For any CELVs \([s_j^k, s_j^l], \quad j \in [i, i+1], \quad [s_{ij}^m, s_{ij}^n], q \in [p, p+1], \quad [s_x^u, s_x^v], \quad y \in [x, x+1] \) and positive numbers \( \rho, \sigma \), there are

1) **(commutativity)**
\[ [s_j^k, s_j^l] + [s_{ij}^m, s_{ij}^n] = [s_{ij}^m, s_{ij}^n] + [s_j^k, s_j^l] \]

2) **(associativity)**
\[ ([s_j^k, s_j^l] + [s_{ij}^m, s_{ij}^n]) + [s_x^u, s_x^v] = [s_x^u, s_x^v] + ([s_j^k, s_j^l] + [s_{ij}^m, s_{ij}^n]) \]

3) **(distributivity I)**
\[ \rho([s_j^k, s_j^l] + [s_{ij}^m, s_{ij}^n]) = \rho([s_j^k, s_j^l]) + \rho([s_{ij}^m, s_{ij}^n]) \]

4) **(distributivity II)**
\[ (\rho + \sigma)[s_j^k, s_j^l] = \rho[s_j^k, s_j^l] + \sigma[s_j^k, s_j^l] \]

**Proof:**

1) The commutativity can be proven from the definition of the add operation directly, such that it is omitted here. Other properties are proven in following:

2) **Associativity**

Because
\[ ([s_j^k, s_j^l] + [s_{ij}^m, s_{ij}^n]) + [s_x^u, s_x^v] = [s_{ij}^m, s_{ij}^n] + ([s_j^k, s_j^l] + [s_x^u, s_x^v]) = [s_j^k, s_j^l] + [s_{ij}^m, s_{ij}^n] \]

when \( C(X) = F(X) \), i.e. \( F(X) = X \) and \( M(X) = 0 \), then
\[ \alpha = F((k + l + m + n) - M(X))F(X) + M(X)C(X) + ux + vy \]
\[ = F((k + l + m + n) + u + v) \]
\[ = F((k + l + m + n) + u + v) \]
\[ = F((k + l + m + n) + u + v) \]

and
\[ \lambda = M((k + l + m + n) - M(X))F(X) + M(X)C(X) + ux + vy \]
\[ = M((k + l + m + n) + u + v) \]

Otherwise, if \( C(X) = F(X) + 1 \), thus
\[ \alpha = F((k + l + m + n) - M(X))F(X) + M(X)C(X) + ux + vy \]
\[ = F((k + l + m + n) - M(X) + u + v) \]
\[ = F((k + l + m + n) - M(X) + u + v) \]
\[ = F((k + l + m + n) + u + v) \]
\[ = F((k + l + m + n) + u + v) \]
\[ = F((k + l + m + n) + u + v) \]
\[ = F((k + l + m + n) + u + v) \]
\[ = F((k + l + m + n) + u + v) \]

So, in both cases, there must be
\[ \alpha = F((k + l + m + n) + u + v) \]
\[ \lambda = M((k + l + m + n) + u + v) \]

Hence, according to the definition of “add” operation,
\[ y = ((k + l + m + n) - M(X) + u + v) \]
\[ = k + l + m + n + u + v \]

Similarly, we also can prove that
\[ \beta = C((k + l + m + n) - M(X) + u + v) \]
\[ = C((k + l + m + n) + u + v) \]

Such that
\[ ([s_j^k, s_j^l] + [s_{ij}^m, s_{ij}^n]) + [s_x^u, s_x^v] = \]

\[ F((k + l + m + n) + u + v) \]
\[ M((k + l + m + n) + u + v) \]
\[ M((k + l + m + n) + u + v) \]

In a like manner, we can prove
\[ [s^k_j, s^l_j] + ((s^m_p, s^n_q) + (s^u_x, s^v_y)) = [s^k_j, s^l_j] + [s^m_p, s^n_q] + [s^u_x, s^v_y] \]
\[ = [s^m_{k+l+m+n}, s^n_{k+l+m+n}] + M_{kl}(k+l+m+n) - M_{kl}(k+l+m+n) \]
\[ = [s^m_{k+l+m+n}, s^n_{k+l+m+n}] + M_{kl}(k+l+m+n) - M_{kl}(k+l+m+n) \]
\[ = [s^m_{k+l+m+n}, s^n_{k+l+m+n}] + M_{kl}(k+l+m+n) - M_{kl}(k+l+m+n) \]

Thus, in both cases,
\[ T = [s^k_j, s^l_j] + [s^m_p, s^n_q] + [s^u_x, s^v_y] \]

3) Distributivity I
\[ \rho([s^k_j, s^l_j] + [s^m_p, s^n_q]) = \rho(s^m_p, s^n_q) \]
\[ = \rho(s^m_p, s^n_q) \]
\[ = \rho(s^m_p, s^n_q) \]

4) Distributivity II

And the term in the right hand of the equation to be proved is
\[ T = \rho(s^k_j, s^l_j) + \sigma(s^k_j, s^l_j) = \rho(s^k_j, s^l_j) + \sigma(s^k_j, s^l_j) \]

4. Application Example

Marketing decision often need analyze the incomes of potential customers. However, in some cultures such as in the Chinese culture, income information may be the most private. Thereby, the analyst has to estimate it in some cases. Using linguistic variables represent such estimations may be more convenient.

Provided there are two families:
1) The first family (F_1) has two members. One has a VeryGood income and the other is a homemaker with a MediumPoor income from a SOHO work.
2) Three members live in the second family (F_2) including a retiree with Poor pension. The others are both school teachers whose incomes are Fair and MediumGood in the local, respectively.

Question:
1) Which family has a higher total income?
2) Which family has a higher average personal income?

Instinctively, these questions are not easy to answer. By using CELV model, we have
\[ T_1 = \text{(one)WG + (one)MP} = s^1_k + s^2_j \]
\[ = [s^1_k, s^2_j] \]
\[ = [s^1_k, s^2_j] \]

The result shows that the family F_1 has a MediumGood total income. However, the counting label “2” indicates that such result is with reference to a two-member family.

\[ T_2 = \text{(one)P + (one)F + (one)MG} \]
\[ = s^1_k + s^3_j + s^4_l = [s^1_k, s^3_j] + [s^4_l, s^0_0] + [s^1_k, s^3_j] \]
\[ = [s^1_k, s^3_j] \]
\[ = [s^1_k, s^3_j] \]

Similarly, the result shows that the total income of the family F_2 is MediumPoor to Fair with reference to a three-member family. From definition 3, \( T_1 \perp T_2 \).

Furthermore
\[ \text{Average}_1 = \frac{T_1}{2} = \frac{1}{2} [s^1_k, s^3_j] = [s^1_k, s^3_j] = s^1_k \]
The income of the family \( F_1 \) appears \textit{MediumGood}. The income of the family \( F_2 \) is between \textit{MediumPoor} and \textit{Fair} with reference to a three-member family. This income also like the total incomes of three members, one has \textit{MediumPoor} income and the others have \textit{Fair} incomes.

2) Averagely, a member in the family \( F_1 \) has a \textit{MediumGood} income. It is higher than the family \( F_2 \) whose average personal income is \textit{MediumPoor} to \textit{Fair}.

5. Discussion

From the results of the example, there are two major characteristics:

1) For the question 1, the CLV model gives the incompatible result rather than a deterministic one. This incompatibility of the linguistic information was mentioned in definition 3.

2) The aggregation results of the CLVs appear clear meaning which would give the analyst or DM more expressive details.

Table 1 shows the comparison of the results from different computing models for the example in section 4. For the Question 1, the 2-tuple linguistic models in [17] and [25] do not give the answer since the “add” operator was not defined in these models. Meanwhile, the virtual linguistic model [22] gives a deterministic result. But this result perhaps is doubtful from the actual background of the example. It is hard to assert that both families have the same total income based on the given linguistic information.

For the Question 2 in the example, we get Average\(_1\) > Average\(_2\) by using CLV. It is same as the results of other models. This result illustrates the efficiency of the CLV model partially. First, the order of the CLV model is not bizarre comparing to the existing models in their successful used field. Second, the semantics of the aggregation results are clear. A CWW result still contains a meaningful linguistic term from the original scale \( S_F \).

We used a symmetrical conversion in aggregating CLVs into the equivalent CLV or CELV. Such conversion implies that a linguistic scale has the uniform semantics structure. Although this structure is the default structure in many studies, it needs more proofs on such uniformity. If the semantics structure is not uniform, the aggregation of CLVs should be updated accordingly.

6. Conclusions

Generally, symbolic scheme is more convenient than set-based scheme in linguistic decision making. However, the existing symbolic methods are troubled from how to improve the aggregation results to have more semantic meaning. The CLV (CELV) model, which composes a linguistic term and a counting label, is thereby proposed to treat the problem. The information losing and semantics issue are handled when the CWW is performed on the CLVs (CELVs). The effectiveness of the proposed method was illustrated with a example. It should also be applicable to treat other linguistic problems.

Acknowledgment

We would like to thank the editors and anonymous referees for their valuable advices. This work was supported by MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Nos. 13YJC630200 and 12YJA630114), Hunan Provincial Natural Science Foundation of China (No. 14JJ4050) and China Post-doctoral Science Foundation (No. 2014M552135).

References


[27] F. Herrera, E. Herrera-Viedma, and L. Martínez, “A fusion approach for managing multi-granularity...


**Wu-E Yang** received the Ph.D. degree in management science and engineering from Central South University, Changsha, China, in 2013. He currently is a lecturer with the School of Economics, Central South University of Forestry and Technology, Changsha, China. His research interests include uncertainty, decision analysis, and risk management.

**Xin-Fan Wang** received the Ph.D. degree in management science and engineering from Central South University, Changsha, China, in 2014. He currently is a professor of School of Science, Hunan University of Technology, Zhuzhou, China. His research interests include aggregation operators, decision making and uncertainty.

**Jian-Qiang Wang** received his Ph.D. degree in management science and engineering from Central South University, Changsha, China, in 2005. He is currently a professor with the School of Business, Central South University. His research interests include decision-making theory and application, risk management and control, and information management.