

Discovering Weight Knowledge Based on Linguistic Information Aggregation

CHEN Shuguo¹, WANG Dongfa¹, ZHU Haijun², DENG Mingrong³ and MA Hong^{3*}

¹State Grid Zhejiang Electric Power co Ltd, China

²State Grid Zhejiang Procurement Company, China

³School of Management, Zhejiang University, China

*Corresponding author: MA Hong, School of Management, Zhejiang University, China, Tel: 86-571-88206827, E-mail: hongma@zju.edu.cn

Received Date: December 22, 2020 Accepted Date: January 22, 2021 Published Date: January 27, 2021

Citation: CHEN Shuguo (2021) Discovering Weight Knowledge Based on Linguistic Information Aggregation. J Comput Sci Software Dev 1: 1-8.

Abstract

Weight knowledge is important in many fields in decision science. For example, in an artificial neural network (ANN), knowledge resides in the weights on links between neural neurons, based on which output is obtained from given inputs. Although a meaningful output of an ANN is of great use, implications of individual weights are often difficult to decipher. In the context of qualitative decision making, linguistic variables often express information that is about ordinal ranking rather than an exact quantity, and therefore the challenge lies in how to determine the weights conforming to the reality and making sense of the decision. In this work, we discriminate two kinds of weights, where one kind is related to the importance of attributes, and the other is related to aggregation characteristics. We propose a weight estimation method, in which these two kinds of weights are calculated simultaneously from a set of known cases that provide additional aggregation information. Using a simulated annealing algorithm, weights satisfying certain conditions can be obtained, and results for new cases can be aggregated thereafter. We find that the weights developed in our method are more explanatory than other approaches, such as the artificial neural network.

Keywords: Linguistic Information; Weights Aggregation, Weight Discovering, Simulated Annealing Algorithm

Introduction

Aggregation of criteria functions to form an overall decision function is important in various fields in decision science. In the simplest cases such as when both outputs of the criterion functions and the overall decision function are numerical values, the knowledge lies in determining the criterion weights, by which one can get the overall grade by calculating a weighted average of all criteria grades. In more complex situations, the output can no longer be represented as a linear function of inputs. For example, when an artificial neural network (ANN) is trained by a set of instances with known results, the knowledge is embodied in weights on links between neural neurons. [14] However, the benefit of an ANN is achieved at a cost, which is the inherent inability to explain, in a comprehensible form, the process through which a decision or output is given. [1] To address this issue, some complementary techniques such as extracting rules from ANN have been studied (see [1, 8] for example).

Uncertainties in inputs and outputs also bring complexity. [21, 22] The Fuzzy Set theory, proposed and developed by Zadeh [24] is often used to deal with the situation where the assessment information is stated in linguistic terms. Membership functions can then be defined to denote the degree of truthfulness of the proposition. Evidence theory, of Shafer [17], is also useful in aggregating information when uncertainty is attributable to ignorance, instead of fuzziness [5]. Because of the difficulties in defining consensus membership function and approximation of the ultimate irregular fuzzy set, a method that directly operates on linguistic variables has been proposed in [9-11]. The method is based on the concepts of ordered weighted averaging (OWA) operators developed by Yager [20] and convex combination of linguistic labels defined by Degani and Bortolan. [3]

Aggregating linguistic information has large potential in practice. [15] For example, when assessing the attributes of cars such as acceleration, braking, handling, ride quality and powertrain, customers often use linguistic terms, or linguistic variables, to describe their opinions. Therefore, how these individual attribute assessments can lead to an overall ranking becomes a complicated problem. Although linguistic ordered weighted averaging (LOWA) operator [9] can provide an aggregation mechanism, it assumes that inputs are equally important. Therefore, it becomes difficult to develop the weights accordingly, which implicitly determine the degree of “anding” and “oring” in the aggregation.

In this work, we develop a method that extends the LOWA operation by removing the “equal importance” assumption

in the inputs. The knowledge about the aggregation mechanism and importance of each input variable can be calculated from a set of known cases with additional aggregation information.

This paper is structured as follows. In Section 2, the problem is briefly reviewed. The LOWA operator and the weights in different means are discussed in Section 3. In Section 4, we propose an operator allowing for two kinds of weights, and present how it can be used to aggregate linguistic information based on mining known behavior cases. Numerical examples and related computation are illustrated in Section 5. Finally, the conclusion together with some discussion are given in Section 6.

Problem Description

Let A_1, A_2, \dots, A_n be n criteria in a multi-criteria problem. X is the set of alternatives. For a proposed alternative $x \in X$, $A_j(x)$ indicates the performance of x according to Criterion A_j . The assessment of an alternative against other criterion or itself may be qualitative. Our problem is to develop a proper method to get the overall assessment of x from the set of $\{A_j(x) \mid j=1, \dots, n\}$.

A number of alternatives with known results are available, which act as a series of cases provided for case-based reasoning. There may also be other form of knowledge on hand such as comparative knowledge of two criteria's importance, etc.

Lowoperators and Aggregation Weights

When the performance of alternatives can be assessed quantitatively, an ANN can be expediently trained to learn the aggregation process. However, people still need the knowledge on weight allocation in the network to help comprehend and make sense out of the decision.

As Yager pointed out [20], when all criteria are equally important, the aggregation structure can be viewed as between two extremes, “anding” and “oring”. At one extreme, the overall result is good only if all the criteria are satisfied. Hence the output is the minimum of all the inputs. At the other extreme, the overall result is good as long as at least one of the criteria is satisfied. The output can then be calculated by maximizing all the inputs.

More generally, if the domain of each A_j is $[0,1]$, and all criteria are equally important, then the aggregation process can be expressed by an OWA operator F with weighting vector $v = (v_1, v_2, \dots, v_n)$

$$F(a_1, a_2, \dots, a_n) = v_1 b_1 + v_2 b_2 + \dots + v_n b_n$$

where $V_i \in [1, 0] \forall i, \sum_{i=1}^n V_i = 1$

$$(A_1(x), A_2(x), \dots, A_n(x)) = (a_1, a_2, \dots, a_n),$$

b_i is the i th largest element in the collection $\{a_1, a_2, \dots, a_n\}$.

Vector v can determine the “andness” or “orness” of the aggregation. $v = (1, 0, \dots, 0)$ corresponds to the pure “or” operator, while $v = (0, 0, \dots, 1)$ corresponds to the pure “and” operator.

In many real circumstances, it is more desirable that $A_j(x)$ s present a linguistic value. For example, their domain can be a linguistic expression set $S = \{s_i | i \in \{0, \dots, T\}\}$, used to provide performance value about alternatives according to various criteria. For instance, S can be the set of

$$\{s_0 = \text{None}, s_1 = \text{Very_Low}, s_2 = \text{Low}, s_3 = \text{Medium}, s_4 = \text{High}, s_5 = \text{Very_High}, s_6 = \text{Perfect}\}.$$

If the term set has the following characteristics:

- The set is ordered: $s_i \geq s_j$ if $i \geq j$;
- There is the negation operator: $Neg(s_j) = s_i$ such that $j = T - i$

An aggregation operator, LOWA ϕ can be defined to compute directly from linguistic labels [10]. For the label set

$$\begin{aligned} & \{a_1, a_2, \dots, a_n\}, \\ \phi(a_1, a_2, \dots, a_n) &= V \cdot B^T = C^n \{V_i, b_i, i = 1, \dots, n\} \\ &= V_1 \odot b_1 \oplus (1 - V_1) \odot C^{n-1} \{\beta_h, b_h, h = 2, \dots, n\} \end{aligned}$$

where, v and $B = \{b_1, \dots, b_n\}$ are as before, $\beta_h = V_h / \sum_{k=2}^n V_k, h = 2, \dots, n$, and C^m is the convex combination of m labels and if $m = 2, C^2$ is defined as

$$C^2 \{V_i, b_i, i = 1, 2\} = V_1 \odot S_j \otimes (1 - V_1) \odot S_i = S_k, \text{ for } S_j, S_i \in S, (j \geq i)$$

such that $k = \min \{T, i + \lfloor v_i \cdot (j - i) \rfloor\}$, where $b_1 = s_j, b_2 = s_i$

If $v_j = 1$ and $v_i = 0$ with $i \neq j, \forall i$, then the convex combination is defined as: $C^m \{v_p, b_p, i = 1, \dots, m\} = b_j$

Therefore, it is easy to see that how to calculate the weighting vector of LOWA operator v is a basic question to be resolved.

On the other hand, weights in LOWA operators determine the structure of aggregation, so can be viewed as aggregation weights. If the attributes or criteria are not really of equal importance, another kind of weights should be taken into account. Issues of weighted ag-

gregation operators have been studied in [6, 7, 9-12, 16, 19, 23, 25]. In the following section, how the importance weight vector w of criteria can be integrated into the aggregation process, and how they can be identified from a set of cases are studied.

Attribute Weights and Discovering By Learning

LOWA operator with determinate weights

The weights bound with attributes are more intuitive for assessing alternatives. Most of the information aggregation methodologies in literature presume the weights are assigned in advance, but in reality, it is not easy for the decision-maker to give a certain and consistent weight vector. Some improvement ideas include attribute weights as fuzzy numbers (e.g. [2, 24]) or linguistic variables (e.g. [4, 18]).

In our method these weights are regarded as interior quantitative variables implied in the aggregation results. They are initially unknown, but can be defined and calculated, and used to aggregate information thereafter.

Let the importance weight vector of criteria be $w = (w_1, w_2, \dots, w_n)$: $w_i \in [0, 1] \forall i, \sum_{i=1}^n w_i = 1$.

Considering both the aggregation weights and the attributed weights, for some alternative $x: (A_1(x), A_2(x), \dots, A_n(x)) = (a_1, a_2, \dots, a_n)$, an aggregation operator ϕ can be defined as:

$$\phi_{v,w}(a_1, a_2, \dots, a_n) = C^n \{u_i, b_i, i = 1, \dots, n\}$$

where $u_i = V_i \cdot W_{\sigma(i)} / \sum_{i=1}^n (V_i \cdot W_{\sigma(i)})$, $i = 1, \dots, n$, σ is a permutation over $\{1, 2, \dots, n\}$ such that $(b_1, b_2, \dots, b_n) = (a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)})$.

For the operator $\phi_{v,w}$ (we can call it LOW²A), we have simple properties as follows.

$$(a) \text{ If } w = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \text{ then } \phi_{v,w} = \phi$$

(b) If $w_i = 1$ for some $i \in \{1, 2, \dots, n\}$ and $v_i \neq 0$ for all i , then $\phi_{v,w}(a_1, a_2, \dots, a_n) = a_i$. Because $u_{\sigma^{-1}(i)} = 1$, and $u_j = 0$ with $j \neq \sigma^{-1}(i)$, $\forall j$, we have $\phi_{v,w}(a_1, a_2, \dots, a_n) = b_{\sigma^{-1}(i)} = a_i$.

(c) If $w_i = 0$ for some $i \in \{1, 2, \dots, n\}$, then $\phi_{v,w}(a_1, a_2, \dots, a_n)$ is independent of a_i .

This is obvious from the fact that $u_{\sigma^{-1}(i)} = 0$ and the definition of the convex combination C^m .

The LOW²A operator $\phi_{v,w}$ can be illustrated by a simple example.

Assume the aggregation problem involves 4 criteria. $v = (0.25, 0.25, 0.25, 0.25)$, which implies both the “orness” and the “andness” of aggregation are 0.5. An alternative x has criteria assessments (s_0, s_3, s_2, s_4) .

If four criteria are equally important, then the aggregation result for x by ordinary LOWA operator will be $\phi = s_3$

However, if the criteria weight vector is given, for instance, $w = (0.2, 0.3, 0.4, 0.1)$, then one can calculate the composite weight $u = (0.1, 0.3, 0.4, 0.2)$, and $\phi = s_2$. This comes from the fact that importance of the third largest element has been improved.

Weight discovering model

By means of the LOW²A $\phi_{v,w}$ operator defined in 4.1, the aggregation mechanism can be mapped to configuration of the weights v and w . Taking $v_1, \dots, v_2, w_1, \dots, w_n$ as variables, according to the cases and information given, a mathematical programming can be built to estimate the weights.

a) Each alternative case (a_1, a_2, \dots, a_n) with known result S_k can be transformed into a constraint equation.

$$\phi_{v,w}(a_1, a_2, \dots, a_n) = S_k$$

b) The information about the comparison of criteria importance should be expressed by equation or inequality in weight variables.

c) The objective function can be maximizing or minimizing the dispersion of v , or w , or both, where

$$dispersion(W) = -\sum_i W_i \ln W_i$$

$$dispersion(V) = -\sum_i V_i \ln V_i$$

Numerical Examples

Model and numerical examples

In this section, we make use of an example to demonstrate how the approach can be applied.

For a collection of alternatives, each can be assessed based on a set of attributes, to get an overall assessment. In group decision making context, information aggregation can refer to producing a comprehensive evaluation from a group of decision makers’ evaluations. In this example, we have four alternatives, and for each one, four linguistic assessments can be obtained

from 4 experts according to their own specialties. We are interested in finding aggregation knowledge in terms of aggregation weights and expert weights.

Suppose we have a set of m alternatives, whose assessments results are according to the experts and their actual performance ranks are known. Let the alternative set be $\{c_1, c_2, c_3, c_4\}$. For each alternative c_i , we have its attribute linguistic evaluation $A(c_i)$, and actual total performance $D(c_i)$, as follows.

a) For $c_1 : \{a_{11} = s_1, a_{21} = s_2, a_{31} = s_3, a_{41} = s_4\}$, actual performance is s_5 .

b) For $c_2 : \{a_{12} = s_1, a_{22} = s_3, a_{32} = s_3, a_{42} = s_1\}$, actual performance is s_3 .

c) For $c_3 : \{a_{13} = s_4, a_{23} = s_2, a_{33} = s_3, a_{43} = s_2\}$, actual performance is s_2 .

d) For $c_4 : \{a_{14} = s_1, a_{24} = s_4, a_{34} = s_1, a_{44} = s_0\}$, actual performance is s_1 .

In addition, provided it is also known that the 3rd expert is more knowledgeable than the second expert.

Then, a mathematical programming model for minimizing the sum of dispersion of v and of w can be described as:

$$\min G(v, w) = \sum_{i=1}^4 V_i \ln V_i + \sum_{i=1}^4 W_i \ln W_i \quad (5.1)$$

S. t.

$$\phi_{v,w}(S_1, S_5, S_4, S_6) = S_5$$

$$\phi_{v,w}(S_1, S_3, S_3, S_1) = S_3$$

$$\phi_{v,w}(S_4, S_5, S_3, S_2) = S_2$$

$$\phi_{v,w}(S_1, S_6, S_1, S_0) = S_1$$

$$V_3 \geq V_2$$

$$V_i \geq 0, W_i \geq 0, (i = 1, 2, 3, 4); \sum_{i=1}^4 V_i = 1, \sum_{i=1}^4 W_i = 1$$

A simulated annealing solution approach

As we can see, the model is formulated as a nonlinear mathematical programming model, which is difficult to solve to

obtain exact optimal solutions. Therefore, in this paper, we propose a meta-heuristic approach, simulated annealing [13], to generate feasible solutions and then optimize heuristically.

The simulated annealing algorithm minimizes the non-linear objective function $G(v, w)$ (Eq.(5.1)) from an initial vector based on a series of random moves. A group of parameters determine how a move is generated, and whether it is accepted, and how many times moves occur before temperature gets changed. Among the parameters, p defines the probability used in the generation of a random move. γ determines the initial temperature T_0 as $\gamma F(X_0)$ where X_0 is the initial solution. r refers to the number of random moves before temperature is changed. β specifies the temperature reduction rate, where $0 < \beta < 1$. The simulated annealing algorithm is illustrated as follows.

1) Select an initial solution X_0 . Set the temperature $T_0 = \gamma F(X_0)$.

2) Repeat the following steps r times:

a) Generate a random move. Each element in the vector has probability p unchanged and probability $1-p$ replaced by a new random value.

b) Calculate the objective value of the new vector X .

c) Update the current configuration to the new vector X if a better objective value is found.

d) Otherwise, let ΔF denote the increase in the value of the objective function. Update the current configuration to the new vector X with probability $p_1 = \exp(\Delta F / T)$. Keep the current configuration with probability $1 - p_1$.

3) Reduce the temperature by multiplying the time reduction parameter β .

4) Terminate when an optimum value is obtained or the objective value cannot get improved after certain number of temperature changes.

In order to find a feasible solution weights, we set the objective function as follows.

$$F(v, w) = \sum_{i=1}^m (u(\varphi_{v,w}(A(c_i))) - u(D(c_i)))^2 + \max\{v_2 - v_3, 0\} \quad (5.2)$$

In Equation (5.2), u is the utility for evaluation rank, which is set as $u(s_i) = i$ for symmetric and even linguistic term set. For asymmetric terms set, u is set according to [26].

We notice in a standard simulated annealing algorithm, a random move on the weight element w_i may dramatically change the whole weight vector (w_1, w_2, \dots, w_n) , and therefore we apply the following customized procedure to keep some characteristics of the weight vector unchanged when implementing a random move. We select a non-zero w_j , say, the first one greater than 0.2 in our example, and keep each w_i / w_j for $i \neq j$ unchanged with probability p and change it to $\max\{w_i / w_j + \text{uniform}(-1, 1), 0\}$ with probability $1-p$.

Once a (v, w) satisfies $F(v, w) = 0$, we switch to a procedure for minimizing another objective function $G(v, w)$. The simulated annealing algorithm is the same as the above, except that for each random move, $F(v, w)$ is calculated and feasibility is kept during the whole second phase.

In this experiment, we set $p_1 = 0.6$, $\gamma_1 = 1$, $\beta_1 = 0.98$, $r_1 = 20$, $v_0 = (0.25, 0.25, 0.25, 0.25)$, $w_0 = (0.25, 0.25, 0.25, 0.25)$ in the first phase. After 13 temperature changes, a set of feasible weights is obtained $v = (0.278, 0., 0.424, 0.298)$, $w = (0., 0.25, 0.58, 0.17)$

In the second phase, $p_2 = 0.8$, $\gamma_2 = 100$, $\beta_2 = 0.99$, $r_2 = 40$, we can get an optimized solution, $v = (0.05, 0., 0.95, 0.)$, $w = (0., 0.025, 0.244, 0.734)$

Using these weights, aggregation over a new alternative can be processed conveniently.

It should be noted that a small learning set leads to a large number of feasible weights, and the procedure easily terminates with a solution. While the learning set gets bigger, the searching process becomes harder. But the solution obtained makes more sense.

Conclusion

In this paper, we propose a new method for discovering a mechanism of aggregating linguistic information from a set of cases. Two kinds of weights, namely, aggregation weights and attributed weights are distinguished. They play key roles in aggregation and can be learned from a learning case set composed of alternatives with known aggregation information. Any known alternative provides actual comprehensive evaluations, or only some order relations between them for example, an alternative with attribute performances $\{a_{11} = S_1, a_{21} = S_5, a_{31} = S_4, a_{41} = S_6\}$ will be better than the alternative $\{a_{11} = S_2, a_{21} = S_4, a_{31} = S_5, a_{41} = S_5\}$ as a whole, can be converted into equality or inequality constraints.

Since the learning set is the foundation of model composition and serves as the base of weights discovering, the alternatives should be chosen according to typicality and rightness. Improper alternatives may result in conflicts among constraints and result in the model yielding no solution. Of course, an amendment of the model can be a goal for programming with only some of the cases satisfying constraints equation. Both some constraints and objectives can be assigned by priority parameters. The constraints with more certainty and accuracy should be with higher priorities and be satisfied first.

Since finding the exact optimal weight solution can be very difficult, we present a procedure beginning with searching feasible solution and then optimizing according to the pre-defined objective function. The simulated annealing algorithm is employed because it is not dependent on the initial point and shown convergent to the global optimum point. As illustrated in an example, the weights obtained using our method are more explanatory and meaningful when compared with the weights learned from other approaches, such as the traditional ANN.

References

1. Andrews R, J Diederich, AB Tickle (1995) Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-based Systems*, 1995: 373-89.
2. Cholewa W (1985) Aggregation of fuzzy opinions—an axiomatic approach. *Fuzzy Sets and Systems* 17: 249-58.
3. Degani R, G Bortolan (1988) The problem of linguistic approximation in clinical decision making. *Int J Approximate Reasoning* 2: 143-62.
4. Delgado M, JL Verdegay MA Vila (1993) On aggregation operations of linguistic labels. *Int J of Intelligent Systems* 8: 351-70.
5. Deng M, W Xu, JB Yang (2004) Estimating the attribute weights through evidential reasoning and mathematical programming. *International Journal of Information Technology & Decision Making* 3: 419-28.
6. Dubois, D. and H. Prade, Weighted minimum and maximum operations in fuzzy set theory. *Information Sciences*, 39: 205-10.
7. Dubois D, H Prade C Testemale (1988) Weighted fuzzy pattern matching. *Fuzzy Sets and Systems* 28: 313-31.
8. Elalfi AE, R Haque, ME Elalami (2004) Extracting rules from trained neural network using GA for managing E-business. *Applied Soft Computing* 4: 65-77.
9. Herrera F, E Herrera-Viedma, J Verdegay (1996) Direct approach processes in group decision making using linguistic OWA operators. *Fuzzy Sets and Systems* 79: 175-90.
10. Herrera F, E Herrera-Viedma (1997) Aggregation operators for linguistic weighted information. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 27: 646-56.
11. Herrera, F. and E. Herrera-Viedma, Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 2000. 115(1): p. 67-82.
12. Ju Y (2014) A new method for multiple criteria group decision making with incomplete weight information under linguistic environment. *Applied Mathematical Modelling* 38: 5256-68.
13. Kirkpatrick S, CD Gelatt, MP Vecchi (1983) Optimization by simulated annealing. *Science* 220: 671-80.

14. Rosenblatt F (1962) Principles of Neurodynamics.
15. Sanchez E (1989) Importance in knowledge systems. Information Systems 14: 455-64.
16. Schjaer-Jacobsen H (2002) Representation and calculation of economic uncertainties: Intervals, fuzzy numbers, and probabilities. International Journal of Production Economics 78: 91-98.
17. Shafer G (1976) A mathematical theory of evidence. Vol. 1. 1976: Princeton university press Princeton.
18. Tong RM, PP Bonissone (1982) A linguistic approach to decisionmaking with fuzzy sets. IEEE Transactions on Systems, Man, and Cybernetics 10: 716-23.
19. Wang YM, Y Luo, YS Xu (2013) Cross-weight evaluation for pairwise comparison matrices. Group Decision and Negotiation 2013: 1-15.
20. Yager RR (1988) On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man, and Cybernetics. 18: 183-190.
21. Yager RR (1994) On weighted median aggregation. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2: 101-13.
22. Yager RR (1999) Modeling uncertainty using partial information. Information Sciences 121: 271-94.
23. Yang Gl, JB Yang, DL Xu, M Khoveyni (2016) A three-stage hybrid approach for weight assignment in MADM. Omega.
24. Zadeh LA (1965) Fuzzy sets. Information and Control 8: 338-53.
25. Zhang F, J Ignatius, CP Lim, M Goh (2014) A two-stage dynamic group decision making method for processing ordinal information. Knowledge-Based Systems 70: 189-202.
26. Zhou W, Z Xu (2016) Generalized asymmetric linguistic term set and its application to qualitative decision making involving risk appetites. European J Operational Res 254: 610-21.

Submit your manuscript to a JScholar journal and benefit from:

- ¶ Convenient online submission
- ¶ Rigorous peer review
- ¶ Immediate publication on acceptance
- ¶ Open access: articles freely available online
- ¶ High visibility within the field
- ¶ Better discount for your subsequent articles

Submit your manuscript at
<http://www.jscholaronline.org/submit-manuscript.php>