Information Fusion 24 (2015) 93–107

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Project evaluation method using non-formatted text information based on multi-granular linguistic labels

Zhu Jian-jun [a,*], Wang He-hua [b], Chen Ye [a], Qin Lang [a]

[a] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, PR China
[b] Jinling Institute of Technology, Nanjing, Jiangsu 211169, PR China

ARTICLE INFO

Article history:
Received 26 January 2014
Received in revised form 7 September 2014
Accepted 9 September 2014
Available online 2 October 2014

Keywords:
Decision analysis
Project evaluation
Multiple-criteria decision analysis
Chinese text evaluation information
Multi-granular linguistic label

ABSTRACT

We propose a novel project evaluation method for non-formatted Chinese text evaluation information. First, the non-formatted Chinese text evaluation information is determined and expressed using extensible markup language and a hypertext preprocessor. Then, the evaluation problem is transformed into a multiple-criteria decision-analysis problem based on multi-granular linguistic labels, including a comprehensive evaluation score for alternatives and an evaluation criteria point score for incomplete items. Next, we propose a weighting model for the criteria based on the minimal difference between the comprehensive evaluation score and the evaluation criterion point score of decision-makers. We establish an estimation model for incomplete evaluation items with the minimal evidence distance of Dempster–Shafer theory using maximal group consistency. In addition, we calculate a weighting for the decision-makers using the similarity of the group. Finally, we present a score modification method for alternatives based on weights of the criteria and the decision-maker. We use a soft science project evaluation and selection to illustrate the application process and feasibility of the proposed method.

1. Introduction

It is often necessary to evaluate projects comparatively by giving a result based on a particular project, criteria, and achievements [1,2]. Although this problem often arises during project evaluation in China, it has not been the focus of much research. Projects are generally evaluated using:

(1) a comprehensive evaluation score of the alternatives, given by experts;
(2) a detailed text evaluation regarding the key points;
(3) an assessment of the experts themselves.

The problem we are concerned with is how to use this information to obtain a more rational evaluation result. Generally, the first and third types of information have been used in the decision-making process, while the second type has been largely ignored. Text evaluation information, which expresses the experts' viewpoints, contains quantitative evaluation results and holds important decision-making information. The text can be formatted or non-formatted. In formatted text, the experts give an evaluation point score based on pre-arranged criteria; that is, the evaluation process has already been determined. For example, a regular evaluation table may be provided before the evaluation process. In the case of non-formatted text, the experts are unaware of the evaluation process, or, no evaluation table exists. Instead, experts give their opinion using plain Chinese text. In this case, there can be large differences in, for example, the evaluation content, evaluation criteria, and expression style. Formatted and non-formatted texts, as well as combinations of both, are widely used in practical evaluation procedures.

Formatted text information is easy to handle, yet rigid, whereas non-formatted text information is flexible, yet difficult to quantify. Non-formatted text contains all the important information, which is critical in the evaluation process. However, the text evaluation information is typically only used as additional material during the evaluation procedure. There has been much research on text information problems, which fall mainly into the categories of text extraction [3,4], clustering [5,6], text mining [7], integration of text information [8,9], text reasoning [10,11], recommending systems on the basis of text [12], and analyzing techniques for text abstraction [13]. From our analysis of existing research, we have found many methods for analyzing formatted text information for decision-making, which can be regarded as a classic multiple-criteria decision-making process [14]. The analysis frame is given in Table 1. Here $a_{ij}^k, i = 1,\ldots,m, j = 1,\ldots,n, k = 1,\ldots,K$ denotes the

* Corresponding author. Tel.: +86 13505152630, +86 02584892701.
E-mail address: zhujianjun@nuaa.edu.cn (J.-j. Zhu).

http://dx.doi.org/10.1016/j.inffus.2014.09.006
1566-2535/

Table 1
Decision-making analysis based on formatted text evaluation information.

Alternative name	Formatted text evaluation information matrix			
	Criterion 1	Criterion 2	...	Criterion n
Project 1	a_{11}^k	a_{12}^k	...	a_{1n}^k
...	...	...	...	...
Project m	a_{m1}^k	a_{m2}^k	...	a_{mn}^k

evaluation value of decision-maker (DM) k for project i and criterion j. Many methods can be used to solve this problem [15].

However, few studies have focused on non-formatted text information for decision-making. Existing text-processing techniques are not sufficiently developed for multi-attribute decision-making. However, by not using the unformatted text and only using their quantitative evaluation results, we cannot take full advantage of the experts' evaluation information. This may lead to information loss and incorrect judgments. Considering the universality and complexity of the non-formatted text information, and having analyzed the problem, we propose the following method. A small change in an expert's rating and evaluation scores may lead to an entirely different result, and thus, a more reasonable selection and evaluation method has a great theoretical and practical value in the case of non-formatted text information.

An increasing number of decision-makers (DMs) refer to the decision-making problem. We cannot manually interpret the non-formatted evaluation information, which has a certain linguistic style, where the linguistic label method is frequently used to evaluate the key points during the decision process. Thus, converting non-formatted text information into a linguistic label format may be a feasible method. Previous linguistic information studies have focused on linguistic label form selection [16–19], computing and operation rules for linguistic variables [20–23], mathematical processing of linguistic variables [24] consistency of linguistic variables [25–26], multi-granular linguistic information [27–31], multi-attribute decision-making based on linguistic variables [32–35], and group decision-making [36–40], with many valuable results having been obtained. However, there are few quantitative models of linguistic labels for non-formatted text evaluation information. If we can automatically solve the evaluation problem, we can increase the rationality and decrease the workload [41]. In this paper, we consider non-formatted text information and its application to the decision-making problem. First, we must solve the important problem of how to deal with the text information. We propose a decision-making model for non-formatted text evaluation information based on linguistic labels and Dempster–Shafer theory (DST) [42,43]. We have developed a new project evaluation method that considers non-formatted evaluation information, and a multiple-criteria decision-making (MCDA) method with multi-granular linguistic labels. The method is most applicable to the management decision problem.

Based on the properties of the problem, we obtain and store the evaluation information using information technology. Then, we format a particular MCDA problem with incomplete evaluation values adopting the method of multi-granular linguistic labels. Thus, the key points of the paper focus on the particular MCDA problem allowing us to present a method for this complex problem. The main contributions of the paper are as follows. First, text evaluation information is considered during the evaluation process. According to existing research, a comprehensive evaluation value has been adopted by many researchers. However, text evaluation information has not been considered, especially non-formatted text information. Second, a mining method for non-formatted text comments is proposed using extensible markup language (XML). This method analyzes the text information automatically. Many important evaluation processes require text information of experts; however, analyzing this information is difficult. Moreover, various important evaluation processes relate to a large number of applicants; therefore, relying on manual processing is not a reality. The process used by the National Natural Science Foundation of China to evaluate projects is one such example. Third, an integrated decision-making approach is proposed based on text evaluation information and the comprehensive evaluation value. The problem of non-formatted text information has its own characteristics. On the one hand, inconsistencies may arise between the text evaluation information and the comprehensive evaluation result. How to solve these inconsistencies is a key problem. On the other hand, one can obtain much evaluation information from many experts' evaluations. Thus, how to determine the weights of the criteria and DMs is another important concern. Our viewpoints differ from existing literature.

The structure of the paper is as follows. Section 2, we define the linguistic variable and DST. The main result is included in Section 3. We propose the multi-granularity linguistic processing and transformation framework for non-formatted text information using an extensible markup language (XML) method. And then, we present the weight model for the evaluation criterion. Following that, we use a numerical estimation model for incomplete evaluation items relating to incomplete text evaluation information. In addition, we suggest a weight model of the experts and a final score adjusting method for the alternatives. Section 4, an example is given to illustrate the concrete application of the method and to demonstrate its feasibility and practicality. Conclusions are made in Section 5.

2. Basic concepts and definitions

In this section, we give some basic definitions to explain our idea clearly, including the linguistic label decision method and DST.

(1) Fuzzy linguistic approach

The fuzzy linguistic approach is modeled using linguistic variables. The values of a linguistic variable are not numbers, but words or sentences expressed in a natural language. In general, words or sentences are less precise than numbers. In a way, the fuzzy linguistic approach is an approximate technique for appropriately representing the qualitative aspects of these problems. The concept of a linguistic variable provides an approximate characterization of the phenomenon [17–19].

A linguistic variable is characterized by the quintuple $((H, T(H), U, G, M))$, where H is the name of the variable; $T(H)$ denotes the term set of H (the set of names of linguistic values of H, with each value being a fuzzy variable denoted generically by X and ranging across a universe of discourse U associated with the base variable u); G is a syntactic rule (which usually takes the form of a grammar) for generating the names of the values of H; and M is a semantic rule for associating its meaning with each H, which is a fuzzy subset of U.

The ordinal fuzzy linguistic approach, which is a special kind of linguistic approach, is defined by considering a finite totally ordered label set $S = \{s_\beta\}$, where $\beta = 0, \ldots, T$, with odd cardinality. The middle term represents an assessment of "approximately equal", while the remaining terms are arranged symmetrically around it. Suppose a DM uses a linguistic term set S to express his/her preferences for the qualitative and quantitative criteria. A linguistic term set can be used to express a preferred measurement process. For example, a set with seven terms is given as [27,44]: