Developing an Evaluation Approach for Software Trustworthiness Using Combination Weights and TOPSIS

Li Shi\textsuperscript{ab}, Shan-lin Yang\textsuperscript{abc}, Kai Li\textsuperscript{ac}, Ben-gong Yu\textsuperscript{ab}
\textsuperscript{a}School of Management, Hefei University of Technology, Hefei 230009, China
\textsuperscript{b}Institute of Computer Network System, Hefei University of Technology, Hefei 230009, China
\textsuperscript{c}Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei 230009, China

Email: \{Sally_shili@yahoo.com.cn, slyang@mail.hf.ah.cn, hfulk@gmail.com, bgyu19@163.com\}

Abstract—Software is everywhere. However, software is not always trustworthy. Confronting the demand of software trustworthiness evaluation, this paper proposes a novel software trustworthiness evaluation approach based on combination weights (CW) and improved TOPSIS methods. The determination of CW relies on experts’ judgments and mathematical computation together. FAHP method is used to determine the importance of degree of criteria according to the experts’ judgment. To avoid the subjective one-sidedness of weights, entropy weighting method is employed to calculate objective weights. Then, the paper adopts combination weighting method to obtain the CW of evaluation criterion. Next, the CW is applied to the improved TOPSIS method and the trustworthiness rating of software is calculated. The result contributes to provide decision makers more decision information. Finally, the proposed approach is demonstrated with the evaluation of PLM software for an aircraft equipment manufacturer in China, which is followed by the sensitivity analysis and the results illustrate the robustness of the presented evaluation approach.

Index Terms—trustworthy software; FAHP; entropy weighting method; TOPSIS; evaluation model; PLM software

I. INTRODUCTION

Modern society is irrevocably and virtually dependent on information technology, and software systems or products in particular. Software has been a key technology in national security, government, finance, business, manufacturing, energy, and permeated services of all kinds. However, software is not always trustworthy. The failure of software can lead to serious consequences, some of which are extensive and even disastrous damage [1]. The problems of trustworthy software have become worldwide in scope, among which, how to evaluate Software Trustworthiness (ST) for providing decision makers more decision information is becoming a desiderated problem and challenger.

Up to now, much effort has been expended in methods for different aspects of trustworthy software [2,3,4,5,6,7,8,9,10], which primarily involve basic concepts, terminology, systems, research plan, measurement and software process trustworthiness evaluation, etc. Based on these studies, such criteria as functionality, co-existence, reliability, safety, maintainability, and usability, etc., have been studied as major trustworthiness properties that contribute to the trustworthiness of a software product. Although there is lack of standard term of “software trustworthiness” for the present, the standpoint that ST is a holistic property encompassing a set of criteria has been an increasing consensus. Therefore, in this paper, software trustworthiness evaluation (STE) is regarded as a Multi-Criteria Decision Making (MCDM) problem that consists of both quantitative criteria and qualitative criteria.

The MCDM approach is suited to deal with two kinds of problems [11,12]. One is the classical MCDM problems [13], among which the ratings and the weights of criteria are measured in crisp numbers. The other MCDM category is the Fuzzy Multi-Criteria Decision Making (FMCDM) problems [14], among which the ratings and the weights of criteria evaluated on imprecision, uncertainty and vagueness are usually expressed by linguistic variables and then set into fuzzy numbers [15].

The purpose of this paper is to formulate software trustworthiness evaluation as a FMCDM model, and then presents a novel approach based on FMCDM methods for evaluating ST. The evaluation results assist the decision makers to select a trustworthy software product from proposed alternatives. Among existing FMCDM methods, Fuzzy Analytic Hierarchy Process (FAHP) [16,17], entropy theory [18] and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [19] method have been widely used to deal with multiple criteria evaluation or selection problems in many fields [20,21,22,23,24,25,26,27,28,29,30,31,32]. However, according to the authors’ knowledge, they are rarely used in the field of software trustworthiness evaluation. Therefore, different from the practical applications reported in the above literatures, this study adopts combination weighting method and improved TOPSIS together based on fuzzy set theory to evaluate the software trustworthiness.

The rest of this paper is structured as follows: Section II presents the evaluation framework. In Section III, the research methodology used in the proposed model is reviewed. Section IV illustrates an application in the evalua-
tion of Product Lifecycle Management (PLM) software for an aircraft equipment manufacturer in China. The results of evaluation are discussed and the model parameters are evaluated with the sensitivity analysis in Section V. The final section draws conclusions and makes suggestions for the future research.

II. EVALUATION FRAMEWORK

The evaluation procedure proposed model consists of three stages as shown in Fig. 1. Considering that main factors contribute to the trustworthiness of a software system, the criteria of STE are chosen in stage 1. Stage 2 determines the weights of criteria and includes two parts. Part one is to obtain subjective weights with FAHP, which is a kind of subjective weighting method widely used to express the judgment of evaluators in dealing with the multiple-criteria evaluation. Part two identifies objective weights by using of Shannon’s entropy theory [18], which is well suited for measuring the related contrast intensities of attributes to represent the average intrinsic information transmitted to the DM [33]. Next, the comprehensive weights of criteria are determined by utilizing a combination weighting method [34]. In stage 3, an improved TOPSIS based on vertical projection distance is used to obtain the final rating of software trustworthiness. The details of stepwise procedure can be found in each of the following sub-sections.

III. RESEARCH METHODOLOGY

A. Evaluation Criteria

Since software trustworthiness does not appear in any glossary, finding what is often termed “software trustworthiness” is an elusive goal. Presently, there are competing definitions of “trustworthy software” found in various literatures [1,2,3,7,35,36,37]. No matter which definition the researchers gravitate to, the viewpoint that ST is a holistic property encompassing a set of criteria has reaching a consensus. Basic research on criteria is not green-field research. Stefan et al. consider that trustworthiness of software systems is determined by correctness, safety, quality of service (performance, reliability, availability), security, and privacy [6]. Tan et al. discuss the key properties that are mentioned when talking about trustworthiness, which includes functionality, security, usability, safety, portability, maintainability, and reliability [7]. According to ISO 9126-1, six major characteristics of software quality attributes such as functionality, reliability, efficiency, maintainability, usability, and portability are defined [38]. Fenton and Pfleeger provide rigorous and practical approaches to measure such criteria as maintenance, reliability, learnability, and operability [39]. Zhao et al. consider the software trustworthiness consists of 5 criteria: availability, reliability, maintainability security and safety [37].

Based on research achievements in existing literatures, the criteria of trustworthiness can be summarized, and a criterion set denoted as $C = \{C_j | j = 1, 2, \ldots, n\}$ is yielded. For our research, the criterion set provides us a good reference list of evaluation criteria of software trustworthiness. Since different organizations in all walks of life have the diverse requirements and expectations on software trustworthiness, in an actual evaluation, only important criteria are taken into consideration. The key trustworthy criteria for specific evaluation can be derived from comprehensive investigation and consultation with different experts from diverse fields.

B. Fuzzy Set Theory

In order to model the uncertainty or inherent imprecision of human judgments and decision-making processes, Zadeh originally introduced fuzzy set theory [40]. In assessment procedures, such terms of expression as “poor”, “very high”, “satisfied”, can be heard very often, and their communality is that they are more or less tainted with ambiguity and vagueness, which makes it difficult to make an exact evaluation in numerical data. Fuzzy theory can play a significant role in dealing with this kind of evaluation situation. The preliminary notions are described as follows.

1) Fuzzy Sets

A fuzzy set is an extension of a crisp set. Crisp sets only allow full membership or no membership at all. Different from the crisp set, fuzzy sets allow partial membership, that is, an element may partially belong to a fuzzy set [21]. One major contribution of fuzzy set theory is its capability of modeling the vagueness of human recognition by using of vague data. Such a set is characterized by a membership function, which uses values ranging from 0 to 1 for showing the membership of the object in a fuzzy set. Complete non-membership is represented by 0, and complete membership as 1. Values between 0 and 1 represent intermediate degrees of membership [41].

2) Fuzzy Numbers

<table>
<thead>
<tr>
<th>Stage</th>
<th>Evaluation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Constructing the evaluation criteria</td>
</tr>
<tr>
<td>2</td>
<td>FAHP</td>
</tr>
<tr>
<td>3</td>
<td>Improved TOPSIS</td>
</tr>
</tbody>
</table>

![Figure 1. Evaluation framework of ST](image-url)
Fuzzy numbers are a fuzzy subset of real numbers, and they represent the expansion of the idea of confidence interval [29]. Among various types of fuzzy numbers, the use of triangular fuzzy numbers (TFNs) is fairly common in literatures [24,29,41,42] because of their computational simplicity and their usefulness in promoting representation and information processing in a fuzzy situation. So TFNs are adopted in this study.

A triangular fuzzy number (TFN) denoted as $\tilde{A} = (l, m, u)$ has the following type membership function:

$$u(x) = \begin{cases} (x - l)/(m - l), & l \leq x \leq m, \\ (u - x)/(u - m), & m \leq x \leq u, \\ 0, & \text{otherwise}. \end{cases}$$

where the parameters $l$, $m$, and $u$, respectively, denote the smallest possible value, the most promising value, and the largest possible value that describe a fuzzy event.

Four important operations on TFN are put forward by Chen and Hwang [43]. The following is the operational laws of two TFNs $\tilde{A}_1 = (l_1, m_1, u_1)$ and $\tilde{A}_2 = (l_2, m_2, u_2)$.

Addition of a fuzzy number $\ominus$

$$\tilde{A} \ominus \tilde{A}_j = (l_j, m_j, u_j) \ominus (l_1, m_1, u_1) = (l_1 + l_j, m_1 + m_j, u_1 + u_j).$$

Multiplication of a fuzzy number $\ominus$

$$\tilde{A} \otimes \tilde{A}_j = (l_1, m_1, u_1) \otimes (l_j, m_j, u_j) = (l_1 l_j, m_1 m_j, u_1 u_j).$$

Reciprocal of a fuzzy number $\ominus$

$$\tilde{A}^{-1} = (l, m, u)^{-1} = (1/l, 1/m, 1/u).$$

3) Linguistic Variables

A linguistic variable is a variable values of which are linguistic terms [44]. In this study, one example for the linguistic variable is “relative importance of two criteria”. Such five linguistic terms as “equally important”, “weakly important”, “essentially important”, “very strongly important” and “absolutely important” could be the possible values for this variable. Each linguistic expression values can be denoted by a TFN within the scale range of 0-1.

Table 1. shows the linguistic terms and their corresponding TFNs.

Table 1: Fuzzy Linguistic Variable and Corresponding TFNs of Linguistic Terms for Criteria

<table>
<thead>
<tr>
<th>Relative importance of criteria</th>
<th>TFNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equally important (Eq)</td>
<td>(1,1)</td>
</tr>
<tr>
<td>Weakly important (Wk)</td>
<td>(1,2)</td>
</tr>
<tr>
<td>Essentially important (Es)</td>
<td>(2,3)</td>
</tr>
<tr>
<td>Very strongly important (Vs)</td>
<td>(3,5)</td>
</tr>
<tr>
<td>Absolutely important (Ab)</td>
<td>(5,7)</td>
</tr>
</tbody>
</table>

Table 2: Fuzzy Linguistic Variable and Corresponding TFNs of Linguistic Terms for Candidate Software

<table>
<thead>
<tr>
<th>Software trustworthiness / performance</th>
<th>TFNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low (Vl) / Very bad (Vb)</td>
<td>(0,0,0,0,3)</td>
</tr>
<tr>
<td>Low (L) / Not good (Ng)</td>
<td>(3,0,4,5,0)</td>
</tr>
<tr>
<td>Middle (Md) / Middle(M)</td>
<td>(6,0,7,5,8)</td>
</tr>
<tr>
<td>High (H) / Good(G)</td>
<td>(8,5,9,0,9)</td>
</tr>
<tr>
<td>Very high(Vh) / Very good(Vg)</td>
<td>(9,5,10,0,9)</td>
</tr>
</tbody>
</table>

be “very low”, “low”, “middle”, “high” “very high”. Each lingual expression values can be defined and denoted by a TFN shown in Table 2.

C. Weighting Methods

Each criterion has its own contribution or influence to software trustworthiness evaluation. In MCDM methods, influence coefficient refers to the weights of each criterion. Up to now, there are two categories of weighting methods: subjective weighting methods and objective weighting methods.

In general, the subjective methods determine weights solely according to the preference or judgments of decision makers, which include AHP, expert investigation methods, etc. The objective methods determine weights by solving mathematical models automatically without any consideration of the decision maker’s preferences, for example, the entropy method, maximizing deviation method and multiple objective programming, etc.

However, the both weighting methods have their own strong points and defects. The subjective weighting method can explain the evaluation clearly while the objectivity one is relatively weak. The objective weighting method expresses the evaluation in data, but sometimes the weight coefficients of some indexes disagree slightly on actual significance of these criteria, and it is more difficult to explain intuitively than the subjective weighting method [45]. Therefore, in order to balance the advantages and disadvantages of subjective and objective weights, there are two-type weighting methods, specifically, the FAHP and entropy are both utilized in this study.

1) Objective Weighting Method—Entropy

In information theory, entropy is a criterion used for measuring the amount of useful information with the provided data. The larger of the difference of evaluating objects under the same criterion, the smaller the entropy value is, which illustrates that this criterion provides more useful information and the weight of this criterion should be relatively large, and vice versa [32]. The entropy theory is based on mathematical computation and ignores the evaluator’s preference, which makes it to be a comparatively objective method. The entropy weighting method consists of the following steps:

Step 1: Constructing original data matrix

Assure that the problem involves $s$ alternatives evaluated against $n$ criteria. The original data matrix can be modeled with quantitative data of each criterion and denoted as:

$$X = \begin{pmatrix} A_1 & A_2 & \ldots & A_n \end{pmatrix},$$

where $A_i(j = 1,2,\ldots,s)$ represents the candidate alternative (software product), $C_j(j = 1,2,\ldots,n)$ denotes the evaluation criterion, $x_{ij}(i = 1,2,\ldots,s; j = 1,2,\ldots,n)$ is the performance of alternative $A_i$ under criteria $C_j$.

Step 2: Normalizing the data matrix

$$X = \begin{pmatrix} A_1 & A_2 & \ldots & A_n \end{pmatrix},$$

where $A_i(j = 1,2,\ldots,s)$ represents the candidate alternative (software product), $C_j(j = 1,2,\ldots,n)$ denotes the evaluation criterion, $x_{ij}(i = 1,2,\ldots,s; j = 1,2,\ldots,n)$ is the performance of alternative $A_i$ under criteria $C_j$.
Since each criterion has its own dimension and unit, it is essential to transform the various criteria scales into a comparable scale limited between 0 and 1. The original data of evaluation criteria can be normalized as:

\[ v_q = x_{ij} / \sum_{j} x_{ij}. \]  
(5)

Therefore, the normalized matrix can be obtained and represented by \( V = [v_q]_{m \times n} \).

Step 3: Computing the entropy value for the \( j \)th criterion:

\[ N_j = -k \sum_{i=1}^{n} \ln \left( v_{ij} \right). \]  
(6)

where \( k \) is a constant and taken to be 1/\( \ln n \). If all \( x_{ij} \) are identical, \( v_{ij} = 1/s \) and \( N_j = 1 \).

Step 4: Calculating the objective weight vector \( w^{(1)} \) based on the entropies and the element can be calculated as:

\[ w_{ij} = \left( 1 - N_j \right) \left( \frac{1}{\sum_{i=1}^{n} \left( 1 - N_i \right)} \right), \]  
(7)
where \( w_{ij} \) is the objective weight of the \( j \)th criterion.

Subjective Weighting Method

2) Subjective Weighting Method

AHP is developed by Saaty [46] and addresses how to determine the relative importance of a set of criteria in a multi-criteria decision problem. In conventional formulation of AHP, human judgment is described by crisp real numbers. However, in many real world applications, human pair-wise judgment is highly ambiguous and uncertain. In order to effectively handle the problems with imprecise and incomplete information, an approach combined the fuzzy set theory with AHP was first introduced by Van Laar-hoven and Pedrcyz [16], which lead to a novel decision analysis method Fuzzy Analytic Hierarchy Process (FAHP). FAHP utilizes fuzzy numbers in place of exact numbers for the pair-wise comparisons.

In this study, the FAHP initiated by Buckley [17] is employed. Assume that an evaluation group includes \( m \) members. The procedure of determining the subjective weights of evaluation criteria through FAHP can be outlined as follows:

Step 1: Build fuzzy pair-wise comparison matrices among all the criteria by using of (8),

\[ A^j = \begin{bmatrix} 1 & \hat{a}_{12}^j & \cdots & \hat{a}_{1n}^j \\ \hat{a}_{21}^j & 1 & \cdots & \hat{a}_{2n}^j \\ \vdots & \vdots & \ddots & \vdots \\ \hat{a}_{n1}^j & \hat{a}_{n2}^j & \cdots & 1 \end{bmatrix}, \]  
(8)

where \( \hat{a}_{ik}^j = (\hat{a}_{ik})^j, i \neq j, k = 1, \ldots, m \). \( A^j \) represents the fuzzy pair-wise comparison matrix of criteria given by \( j \)th evaluator, and \( \hat{a}_{ik}^j \) denotes the relative importance of two criteria \( C_i \) and \( C_j \).

Step 2: Aggregate the subjective judgment of \( m \) members.

This study employs the geometric mean method suggested by Buckley to compute the element of synthetic pair-wise comparison matrix \( A \) of different evaluators, that is:

\[ \hat{a}_{ik} = \left( \hat{a}_{ik}^1 \otimes \hat{a}_{ik}^2 \otimes \cdots \otimes \hat{a}_{ik}^m \right)^{1/m}. \]  
(9)

Step 3: Calculate fuzzy subjective weights of each criterion.

In this step, the method of normalization of the geometric mean of the rows (NGM) [47] is adopted to obtain the fuzzy subjective weights vector \( \hat{w}^{(2)} \). To multiply the \( n \) elements in each row and take the \( n \)th root, the fuzzy eigenvector from \( A \) matrix is obtained, and then normalizes the eigenvectors via (10):

\[ \hat{w}^{(2)} = \left( \prod_{j=1}^{n} \hat{a}_{ij} \right)^{1/n} \sum_{j=1}^{n} \left( \prod_{i=1}^{n} \hat{a}_{ij} \right)^{1/n}, i,j = 1,2,\ldots,n. \]  
(10)

where \( \hat{w}^{(2)} \) is the fuzzy subjective weight of the \( j \)th criterion, expressed in a TFN \( \hat{w}^{(2)} = \left( l_{i}^{(2)}, m_{i}^{(2)}, u_{i}^{(2)} \right) \).

Step 4: Defuzzification

Defuzzification is a technique to convert the fuzzy number into crisp real numbers [29]. The above subjective weights of criteria are indicated by triangular fuzzy numbers. In order to match the objective weights of criteria represented by exact numbers, one non-fuzzy method for TFNs is important and necessary. Various defuzzification methods are available and serve this purpose, the graded mean integration representation method adopted in this paper is derived from Chen and Hsieh [48]. The defuzzification value of a TFN \( \hat{A} = (l, m, u) \) can be obtained by (11).

\[ R(\hat{A}) = \frac{l + 4m + u}{6}. \]  
(11)

Accordingly, the subjective weights vector \( \hat{w}^{(2)} \) can be obtained by using of (11).

Assume that \( R(\hat{A}) \) and \( R(\hat{B}) \) are the graded mean integration representations of TFNs \( \hat{A} \) and \( \hat{B} \), respectively, and define that:

\[ \hat{A} > \hat{B} \iff R(\hat{A}) > R(\hat{B}), \]
\[ \hat{A} = \hat{B} \iff R(\hat{A}) = R(\hat{B}), \]
\[ \hat{A} < \hat{B} \iff R(\hat{A}) < R(\hat{B}). \]

3) Combination Weighting Method

In order to obtain the synthesized weights, the combination weighting method is used to combine the subjective matrix and the objective entropy weighting in this study. The normalization (5) results in the normalized matrix with dimensionless and unit-free data, which reflects the difference of raw data with respect to every criterion and is suitable for computing weights based on the entropy theory. However, some evaluating criteria are benefit criteria (i.e., the higher the better) whereas others are cost criteria (i.e., the smaller the better). In order to eliminate the anomalies resulting from the difference of criterion types, the matrix \( V = [v_q]_{m \times n} \) should be further normalized by the following formula:

\[ u_q = \left( \frac{\max v_q - v_q}{\max v_q - \min v_q} \right). \]  
(12)
If \( C_j \) is a benefit criterion:
\[
    u_{ij} = (v_{ij} - m_{ij}) / (\max v_{ij} - m_{ij}).
\]
(13)

The generation equation for the combination weights vector has been proposed by Wang et al. and can be expressed as follows [14]:
\[
    W = \sum_{i=1}^{k} \lambda_i \mathbf{w}^{(i)}.
\]
(14)

where \( \mathbf{w}^{(i)} \) and \( \mathbf{w}^{(0)} \) denotes the objective weights from entropy and the subjective weights through FAHP, respectively. \( \lambda_i \) is the linear combination coefficient and is deduced based on the optimization method and Jaynes maximal entropy theory [45]. The value of \( \lambda_i \) is given as:
\[
    \hat{\lambda}_i = \frac{\exp\left(\frac{1}{q} \sum_{j=1}^{n} w_{ij} \left(1 - q_{ij}\right)\right)}{\sum_{j=1}^{n} \exp\left(\frac{1}{q} \sum_{j=1}^{n} w_{ij} \left(1 - q_{ij}\right)\right)}.
\]
(15)

where \( q \) is the balance coefficient and \( 0 < q < 1 \). The constraint condition is \( 0 \leq \lambda_i \leq 1 \), and \( \sum_{i=1}^{k} \lambda_i = 1 \).

Accordingly, the combination weights vector can be obtained and expressed as \( W = (w_1, w_2, \ldots, w_n) \).

D. Establishing the Rating of Candidate Software

1) Constructing Decision Matrix

Given \( s \) alternatives (software products) \( A_i \) \((i = 1, 2, \ldots, s)\), \( n \) criteria \( C_j \) \((j = 1, 2, \ldots, n)\) and \( m \) evaluators, the evaluators describe each alternative’s trustworthiness rating with respect to each attribute by using TFNs shown in the Table 2. The typical structure of fuzzy decision matrix can be expressed as:
\[
    C_1 \quad C_2 \quad \cdots \quad C_n
    \begin{bmatrix}
        w_{11} & w_{12} & \cdots & w_{1n} \\
        w_{21} & w_{22} & \cdots & w_{2n} \\
        \vdots & \vdots & \ddots & \vdots \\
        w_{n1} & w_{n2} & \cdots & w_{nn}
    \end{bmatrix}
\]

\( \hat{E}_i \) is a TFN and represents the fuzzy judgment rating of alternative \( A_i \) concerning the criterion \( C_j \) given by \( k \)th evaluator, and the weighting vector \( W = (w_1, w_2, \ldots, w_n) \) obtained in Section III - C is represented by real numbers.

Since each evaluator has different individual knowledge and experience, the perception toward the software trustworthiness varies among evaluators. Assume each evaluator has the same importance, this study employs the method of average values to integrate the fuzzy judgment value \( \tilde{e}_{ij} \), that is,
\[
    \bar{e}_{ij} = \frac{1}{m} \otimes \left( \tilde{e}_{ij} \oplus \tilde{e}_{ij} \oplus \cdots \oplus \tilde{e}_{ij} \right).
\]
(16)

\[
    \bar{v}_j = \frac{1}{l_{g_j}} \sum_{k=1}^{l_{g_j}} \tilde{v}_k;
    m_{g_j} = \frac{\sum_{k=1}^{l_{g_j}} v_k}{l_{g_j}};
    u_{g_j} = \frac{\sum_{k=1}^{l_{g_j}} \tilde{v}_k}{l_{g_j}}.
\]
(17)

where \( \bar{e}_{ij} \) denotes the average fuzzy scale of alternative \( A_i \) regarding the evaluation criteria \( C_j \) for \( m \) evaluators. After defuzzifying the integrated fuzzy decision matrix \( \bar{R} = (\bar{e}_{ij})_{mn} \) in (11), and the decision matrix is obtained and denoted as \( F = (f_{ij})_{mn} \), where \( f_{ij} \) is a real number.

2) Improved TOPSIS Method

TOPSIS was originally proposed by Hwang and Yoon [19], which is a practical and useful technique for evaluation of a number of possible alternatives through measuring Euclidean distances. The underlying logic of TOPSIS is that the optimal alternative should have the shortest distance from the positive idea solution (PIS), i.e., the solution that maximizes the benefit criteria and minimizes the cost criteria; and the longest distance from the negative ideal solution (NIS), i.e., the solution that maximizes the cost criteria and minimizes the benefit criteria [11].

However, traditional TOPSIS based on the Euclidean distance has an obvious defect, that is, the alternative that has the shortest Euclidean distance from the positive idea solution (PIS) may also have the closest Euclidean distance from the negative ideal solution (NIS) [49,50,51]. Therefore, the ranking provided by traditional TOPSIS can not accurately reflect the difference of alternatives. In order to avoid it, this study proposes to use an improved TOPSIS [52] based on vertical projection distance which has an advantage that the alternative nearest to the PIS is farthest from the NIS.

The improved TOPSIS method is summarized as follows.

Step 1: Calculate weighted normalized decision matrix \( G_g = \{g_{ij}\}_{mn} \), and \( g_{ij} \) is defined as:
\[
    g_{ij} = w_{ij} \times f_{ij} \quad i = 1, 2, \ldots, s; \quad j = 1, 2, \ldots, n.
\]
(18)

Step 2: Determine positive idea solution (PIS) \( S_j^+ \) defined as follows:
\[
    S_j^+ = \begin{cases} \max \{g_{ij}\} & \text{if } C_j \in \Omega_1 \\ \min \{g_{ij}\} & \text{if } C_j \in \Omega_2 \end{cases} \quad j = 1, 2, \ldots, n.
\]
(19)

where \( \Omega_1 \) and \( \Omega_2 \) indicate the benefit criteria set and cost criteria set, respectively.

Step 3: Transfer the origin of coordinates to the positive ideal point for simplifying calculation. The elements of removed matrix \( T \) are given as:
\[
    t_{ij} = g_{ij} - S_j^+ \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n.
\]
(20)

After translation, the PIS becomes \( \{0, \ldots, 0\} \) and the NIS is denoted as \( H^- = \{H^-_j\} \quad j = 1, 2, \ldots, n \).

Step 4: Determine negative idea solution \( H^- \) as follows:
\[
    H^-_j = t_{ij} \quad |H^-_j| \geq |H^+_j| \quad 1 \leq k \leq l_s, \quad i = 1, 2, \ldots, s; \quad j = 1, 2, \ldots, n.
\]
(21)

Step 5: The vertical projection distance of each alternative to PIS can be calculated by:
where $T_i$ is the $i$th row vector of $T$ matrix. The smaller the $P_i$, the better the trustworthiness of the software is.

**Step 6:** The evaluation team defines the range for the linguistic variable of software trustworthiness within a scale of 0-1. Thus, according to the corresponding value of $P_i$, the trustworthy rating of the $i$th candidate software can be determined.

**IV. A CASE OF EVALUATION**

With the development of enterprise information, such large manufacturing industry as automobile, airplane and aerospace, have called for effective implementations of Product Lifecycle Management (PLM) software to enhance enterprise’s level of informationization and core competence. In this section, the proposed method is applied to evaluate the candidate PLM software’s trustworthiness for an aircraft equipment manufacturer in China, and supports decision makers of the enterprise to select the trustworthy PLM software. This evaluation is supported by results that are gained from a series of questionnaires and widespread investigation. After the preliminary screening, three PLM software, $A_1$, $A_2$, $A_3$, have remained in the candidate list. Four experts, $D_1$, $D_2$, $D_3$, $D_4$, selected from different subsystems of the organization, form an evaluator team (for each evaluator with the same importance).

In order to reach consensus on evaluation criteria of PLM software trustworthiness for the enterprise, we conduct a widespread investigation and consultation with several experts, including three professors in information system fields, two evaluation experts from third part and several experts, as well as consulting the software system. After brainstorming sessions, six determinant criteria such as maintainability, functionality, portability, operability, learnability and co-existence are considered to be essential for evaluating the trustworthiness of PLM system. In this evaluation, functionality is regarded as qualitative criteria, whereas the remaining criteria are regarded as quantitative criteria. Trustworthy criteria structure and their measurement are listed in Fig. 2. The following is the illustration of each quantitative criterion and their measurable objectives.

**Portability** means the behavior or ability of software during measurement of porting the software system to another, which can be measured by $ET/ER$, where $ET$ is the resource environment, and $ER$ denotes the resource measurement of creating this system in the target environment. The lower the ratio, the better the portability is.

**Co-existence** is the performance of software system harmonized with the other software in a common environment sharing common resources. In the evaluation, co-existence is measured by estimating of effort (person-months) in redeveloping candidate software in order to fulfill good integration with diverse independent software.

**Maintainability** means the software performance including both defect repairs and enhancements in response to new requirements. The maintenance activity can be regarded as the certain alteration, and the speed of finishing the alteration is a key to measure maintenance. Therefore, this evaluation adopts the Mean Time to Repair (MTTR) to measure the maintenance, which is the average time of finishing the alteration and returning to the normal functioning.

Training 20 new employees, **learnability** can be measured by noting the speed of learning, concretely, the average hours of cultivating before single operation, and **operability** is measured by recording the average numbers of making mistakes during training. The method focuses on the workload of learning and operating the software system.

We gather the related measured data and information by taking a variety of ways, including vendors’ demo introduction, building test platform in enterprise and try-outs for the candidate software, as well as consulting the

![Figure 2. Trustworthy criteria hierarchy of evaluating PLM software](image-url)
other aircraft equipment manufacturers that have applied
the candidate PLM systems. The results construct the or-
iginal matrix of criteria for each candidate PLM system and are
shown in TABLE 3.

The trustworthiness level of PLM software can be as-
essed by aggregating the measurements for corresponding
criteria, and the implementing procedure of evaluation is
summarized as follows.

**Step 1:** Computing the objective weights of criteria.

TABLE 3 provides the original information collected
from inside and outside of the enterprise. Notably, the crite-
ron $C_1$ is qualitative criteria. The information under $C_1$
is the experts perception described by linguistic terms with
TFNs (shown in TABLE 2) after trying out the three can-
didate software. Therefore, before computing the objective
weights, the defuzzification of TFNs is performed by (11).
The trustworthiness level of PLM software can be as-
sessed by aggregating the measurements for corresponding
criteria, and the implementing procedure of evaluation is
summarized as follows.

**Step 2:** Calculating the subjective weights of criteria.

First, the evaluators are asked to conduct their subjective
judgments on the relative importance of criteria based on
their background of experience and knowledge, and estab-
lish the fuzzy pair-wise comparison matrices. The related
linguistic variable can be indicated by TFNs shown in
TABLE 1. Next, to aggregate the fuzzy pair-wise compar-
ison matrices. The related linguistic variable can be indicated by TFNs shown in
TABLE 2. Then the objective weights can be obtained using (10) and (11), which are
shown in TABLE 7. The normalized subjective weights

$$\mathbf{w}_S = (0.4255, 0.0815, 0.1536, 0.0746, 0.2180, 0.0468)$$

**Step 3:** Determining the final combination weights
through combination weighting method.

Combining the evaluators' judgment and impersonality,
the optimal linear combination weighting method is utilized
by $q = 0.5$. According to (12) - (15), the combination
weights are calculated as:

$$\mathbf{w}_C = \lambda w^{(1)} + (1 - \lambda) w^{(2)} = 0.4762 \mathbf{w}_{(1)} + 0.5238 \mathbf{w}_{(2)}$$

$$= (0.2390, 0.0799, 0.1164, 0.1997, 0.2443, 0.1206).$$

Fig.3 shows the difference of criteria weights through en-
tropy, FAHP and combination weighting methods. It is

![Comparison of different weights](image-url)
TABLE 8.
Fuzzy Decision Matrix of Four Evaluators

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>A2</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>A3</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>A4</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

TABLE 9.
Integrated Fuzzy Decision Matrix

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>A2</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>A3</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>A4</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Step 4: Constructing decision matrix.

Evaluators are asked to evaluate the trustworthiness of software by using the linguistic terms in TABLE 2 based on the raw information of original matrix shown in TABLE 3. TABLE 8 shows the fuzzy judgment values of each evaluator. Based on the fuzzy numbers defined in TABLE 2, the linguistic scales can be transferred to the corresponding fuzzy numbers. To employ (16) and (17), the integrated decision matrix established using TFNs is obtained and presented in TABLE 9, for εi,j as an example, the trustworthiness scales of the 1st candidate software A1 with respect to criterion $C_i$ provided by four experts are H, Md, Md, L, respectively, then

$$\begin{align*}
\varepsilon_{i,j} &= \left( \frac{\sum_{k} a_{i,k} / 4 \cdot \sum_{k} m_{i,k} / 4 \cdot \sum_{k} a_{i,k} / 4}{\text{THW}} \right) \\
&= \left( \frac{(8.5 + 6.0 + 6.0 + 3.0) / 4 \cdot (9.0 + 7.5 + 7.5 + 4.5) / 4}{\text{THW}} \right) \\
&= (8.5 + 6.0 + 6.0 + 3.0) / 4 \cdot (9.0 + 7.5 + 7.5 + 4.5) / 4.
\end{align*}$$

The other matrix elements can be obtained by the same calculation procedure. Next, to defuzzi the normalized fuzzy decision matrix using (11), and the non-fuzzy decision matrix is established and shown in Table 10, where the elements of matrix are the real numbers between 0-10.

Step 5: To employ (18) to calculate the weighted decision matrix.

Table 11 gives the weighted result. In the decision matrix, the trustworthy scales of candidate software under each criterion can be regarded as benefit-related performance. In other words, the higher the trustworthy rating the better the decision making of this alternative is. Therefore, according to (19), the positive idea solution (PIS) $S^+$ can be determined and presented as:

$$
S^+ = \{2.3701, 0.7407, 1.1009, 1.8888, 2.3107, 1.1130\}.
$$

Step 6: According to (20), the removed matrix $T$ is obtained and shown in TABLE 12, from which, the negative idea solution $H^-$ can be determined in (21) and shown as follows:

$$
H^- = \{-0.5975, -0.1747, -0.2303, -0.4743, -0.5802, -0.3467\}.
$$

Step 7: The vertical projection distance of each alternative to the PIS can be calculated by using (22), and the experts group provides the range for the linguistic variable of software trustworthiness within a scale of 0-1 based on their under-standing and background knowledge. The results are shown in TABLE 13. The PLM software $A_3$ with the smallest $P_i$ value (0.1949) has the highest trustworthiness among the three candidate PLM software and becomes the most preferred one for the aircraft equipment manufacturer.
V. Sensitivity Analysis and Discussions

The sensitivity is measured by exchanging each criterion’s weight with another criterion weight. Hence, 15 different interchanging and calculations are formed. We use $P_{ij}$ values for each calculation and give them different names, for example $P_{11}$ means the weights of criterion 1 and criterion 4 have changed. TABLE 14 shows the 15 calculations results and Fig. 4 summarizes 15 $P_{ij}$ values of the alternatives on graph.

As can be seen in these calculations, software $A_2$ has the minimum $P_{ij}$ values in all 15 calculations and all the 15 values are less than 0.5, among which there are 10 $P_{ij}$ values in the range (0.00 ~ 0.25) with high trustworthy. By contrast, software $A_1$ has the highest $P_{ij}$ values in the first, third, fourth, sixth, seventh, eighth, ninth, 10th, 12th, 13th, and 14th calculations. And all the 15 $P_{ij}$ values of software $A_1$ are in the range of low trustworthy and untrustworthy, among which there are 11 $P_{ij}$ values in the range of untrustworthy. In addition, except that $P_{36}$ and $P_{46}$ values are in the range of untrustworthy, software $A_2$ has 13 $P_{ij}$ values belong to trustworthy and low trustworthy, among which $P_{34}$ and $P_{45}$ values are in the range of trustworthy. These results prove that the trustworthy PLM software chosen for the aircraft equipment manufacturer is the best one as we detected in the paper.

What’s more, sensitivity analysis implies that the trustworthy rating of alternative software is not obviously sensitive to the changes in the criterion weight, and illustrates the robustness of the evaluation model.

VI. Conclusions

Nowadays, software trustworthiness, a holistic property encompassing a set of trustworthiness criteria, arouses the increasing concerns in the academia and industry. In order to explore and build a new evaluation system for trustwor-

![Figure 4. The new $P_{ij}$ values of the candidate software](image)

**TABLE 14.** THE $P_{ij}$ VALUES OF 15 INTERCHANGES

| $P_{ij}$ | $P_{11}$ | $P_{12}$ | $P_{13}$ | $P_{14}$ | $P_{15}$ | $P_{21}$ | $P_{22}$ | $P_{23}$ | $P_{24}$ | $P_{25}$ | $P_{26}$ | $P_{31}$ | $P_{32}$ | $P_{33}$ | $P_{34}$ | $P_{35}$ | $P_{36}$ | $P_{41}$ | $P_{42}$ | $P_{43}$ | $P_{44}$ | $P_{45}$ | $P_{46}$ | $P_{51}$ | $P_{52}$ | $P_{53}$ | $P_{54}$ | $P_{55}$ | $P_{56}$ |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| $A_1$    | /         | /         | /         | 0.6495   | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $A_2$    | /         | 0.3215    | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $A_3$    | /         | /         | /         | 0.5754   | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $P_{41}$ | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $P_{42}$ | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $P_{43}$ | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $P_{44}$ | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $P_{45}$ | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |
| $P_{46}$ | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         | /         |

Note: An entry “-” indicates a null value.
With fuzzy MCDM methodology, this paper develops account. FAHP method is used to determine the importance of information included in numerical data are taken into account. FAHP method is used to determine the importance of degree of criteria according to the experts’ judgment. To avoid the subjective one-sidedness of weights, entropy weighting method is employed to calculate objective weights by exploiting the useful information of raw data. Then the linear combination weighting method is used to synthesize the subjective weights and objective weights. Additionally, in order to calculate the trustworthy rating of the candidate software, an improved TOPSIS based on vertical projection distance is taken to replace the conventional TOPSIS based on Euclidean distances, which offers a way to overcome the shortcoming of conventional TOPSIS.

In the application, the presented approach is adopted to evaluate the trustworthiness of PLM software for an aircraft equipment manufacturer in China. Finally, by exchanging each criterion’s weight with another criterion weight, the study conducts the sensitivity analysis to illustrate the robustness of the evaluation model. The calculated results of sensitivity indicate thatsoftware $A_2$ has the minimum $P_j$ values in all 15 calculations and all the 15 values belong to the rating of trustworthy.

Although the case study is related to a specific software product and industry, the underlying concepts and methods of this evaluation model can be applied to evaluate other software systems. Furthermore, it is should be pointed out that this study focused primarily on static criteria as the basis of evaluation criteria, and neglected other dynamic influencing factors of ST in different stage of software life cycle. Future research can consider constructing a dynamic criteria structure and improving evaluation approach with other methods.

ACKNOWLEDGMENT

The authors appreciate the support from the National Natural Science Foundation of China (Nos.70871032, 71071045 and 71131002), Anhui Provincial Natural Science Foundation under fund No. 11040606Q27. Special thanks are given to Prof. Jing-biao Huang, Dr. Chun-feng Liu, and other anonymous reviewers for their constructive suggestions.

REFERENCES


the AHP and fuzzy TOPSIS methods

service quality by fuzzy MCDM

initial training aircraft under a fuzzy environment

Applications

T. C. Wang and T. H. Chang,

107

Providence, Rhode Island, pp. 125-138, October

P. A. Morteza,

1103

building in Korea

a housing performance evaluation model for multi-family residential

S. S. Kima, I. O. Yang, M. S. Yeo, and K. W. Kim,

573


H. Y. Wu, G. H. Tzeng, and Y. H. Chen,

2009.

Z.  Y.  W an g ,  H .  F .  G u ,  X .  X .  Y i ,  an d  S .  R .  Z h an g ,


F. B. Schneider,


ISO 9126-1,

2010.


Li Shi received the B.Sc. degree in Electrical Engineering from Hefei University of Technology, Hefei, China, in 2002. She has been working as an engineering and researcher for an electronics company after studies. Since 2005, she is working for the school of Physical and Electronic Information at Huaibei Normal University, Huaibei, China.

She is currently pursuing the Ph.D. degree in the group of Prof. Shan-Lin Yang. Her main research interests are in the areas of multiple criteria decision analysis, decision-making science and technology and trustworthy software evaluation.

Shan-Lin Yang received the M.Sc. degree in Computer Science and Application from Hefei University of Technology, Hefei, China, in 1985. He had been a visiting scholar at Dresden University of Technology, Germany and the University of Melbourne, Australia in 1986 and 1988, respectively.

He is currently working as a Professor in the School of Management, Hefei University of Technology, Hefei, China. His main research interests include decision-making science and technology, decision support system, operation research, and information management and information system. His work has been reported in many journals such as Computer & Operations Research, European Journal Operations Research, International Journal of Production Economics, Systems Engineering—Theory & Practice(Chinese), Journal of Management Science (Chinese), etc.
Kai Li received the B.Sc. degree (1999) in Computer Science Education from Anhui Normal University, Wuhu, China, the M.Sc. degree (2005) in Computer Software and Theory and Ph.D. degree (2009) in Management Science and Engineering from Hefei University of Technology, Hefei, China.

He is currently working as an associate professor for School of Management at Hefei University of Technology. His research interests include operations research, optimization algorithm, production scheduling. He made the paper such as International Journal of Systems Science, Applied Soft Computing, Mathematical and Computer Modelling, Applied Mathematical Modelling.

Ben-gong Yu received the B.Sc. degree in Computer Science, the M.Sc. and Ph.D. degrees in Management Science and Engineering from Hefei University of Technology, Hefei, China, in 1994, 2005, and 2011, respectively. His research interests include information management and information system.