Transdisciplinary Engineering: A Paradigm Shift
C. Chen et al. (Eds.)

doi:10.3233/978-1-61499-779-5-969

Estimating Cost of New Products Using Fuzzy Case-Based Reasoning and Fuzzy Analytic Hierarchy Process

Fentahun M. KASIE[a,1], Glen BRIGHT[b] and Anthony WALKER[b]
[a] *University of KwaZulu-Natal, School of Mechanical Engineering, Durban, South Africa and Hawassa University,Institute of Technology, Hawassa, Ethiopia*
[b] *University of KwaZulu-Natal, School of Mechanical Engineering, Durban, South Africa*

Abstract. Cost estimation is one of the influencial requirements for the current manufacturers when new products are introduced. This paper aims to propose a decision support system (DSS) that retrieves historical cases/products, which have the most similar cost estimates to the current case. This helps users to estimate the costs of new products at early stages of product development. The proposed DSS combines case-based reasoning (CBR), the analytic hierarchy process (AHP) and fuzzy set theory. Cases are represented using an object-oriented (OO) approach to characterize them in *n*-dimensional Euclidean vector space. A numerical example is illustrated to show the applicability of the proposed DSS

Keywords. New product development, cost estimation, decision support systems, case-based reasoning, analytic hierarchy process, fuzzy set theory.

Introduction

As new products are introduced into manufacturing systems, one of the complex issues is estimating the costs of these new products. In order to compete in dynamically changing situations, an appropriate cost estimation approach is required at the early stages of product development processes. Research findings have approved that although these early stage processes contribute to only 10-15% of the total product development costs, the 70-80% of development costs are committed at these stages [1, 2]. Additionally, customers usually demand high quality products with decreased prices. In order to meet these substantial challenges, manufacturers have to predict effectively the costs of the proposed new products. This is because cost estimation determines the overall performances of product development processes [3]. Stjepandić et al. [4] stressed that product development knowledge should be systematically considered at the early stage of product development.

This paper aims to propose a DSS that retrieves prior cases, which are expected to have the most similar cost estimates to the current problem, using case similarity measures. The proposed DSS combines fuzzy CBR and fuzzy AHP in the case retrieval

[1] Corresponding Author. E-mail: fentahunmk@gmail.com or fantahunm@hu.edu.et

process. This kind of combination has not been adequately studied in the past to articulate the problems of new products cost estimation.

The remainder of this paper is organized as follows: Section 1 reviews the literature. Section 2 describes the proposed DSS. In Section 3, a numerical example is illustrated. Finally, conclusions and future works are forwarded in Section 4.

1. Review of litrature

1.1 Related works

Several product cost estimation methods were proposed in the past. They are broadly classified into qualitative and quantitative approaches [5-7]. These approaches are sub-divided in different ways. For example, in Niazi et al. [5], qualitative approaches incorporated intuitive and analogical techniques; and quantitative approaches included parametric and analytical methods. However, in Caputo and Pelagagge [6], expert judgment and heuristic rules were categorized under qualitative; and quantitative approaches incorporated statistical (parametric and neural networks), analogous and generative/analytical methods. The advantages and limitations of the proposed methods were studied in Duverlie and Castelain [1], Rush and Roy [2], Cavalieri et al. [3], Caputo and Pelagagge [6] and Layer et al. [7].

Recently, AI techniques have been widely utilized for product costs estimation. For example, artificial neural networks (ANNs) were applied in Zhang et al. [8], Smith and Mason [9], Cavalieri et al. [3] and Caputo and Pelagagge [6] in different problem domains. CBR has also been utilized in product cost estimation problems. Duverlie and Castelain [1] compared a parametric and CBR method to estimate the costs of pistons. Kim et al. [10] compared multiple regression, ANNs and CBR to estimate the costs of construction projects. In An et al. [11], a CBR cost estimation model combined with the AHP to weight the attributes of construction projects. The findings showed that CBR is a promising approach in this problem domain.

1.2 Combining CBR, AHP and fuzzy set theory

CBR is an analogical reasoning approach, which draws inferences of a new problem depending upon experiences learned from previously solved problems [12]. Problem solving by retrieving successful experiences is a powerful and frequently applied approach in human thought and decision-making process. Human reasoners usually prefer to reuse and/or adapt their past similar situations to the current problem instead of starting from scratch every time. Remembering previously solved problems can be difficult to human users however computers are best to do so [12]. In this aspect, CBR systems seem more consistent with the natural reasoning process of people [13]. This is the major reason that findings from cognitive psychology have approved the psychological plausibility of CBR [12, 14], [15]. Aamodt and Plaza [14] described their general CBR cycle in terms of four ‘Re’s, which is usually called R^4 model.

1. *Retrieve* the most similar prior case to the current problem.
2. *Reuse* the knowledge in the retrieved case.
3. *Revise* the retrieved prior case in order adapt to the new case.
4. *Retain* the final solution as the learned case for future retrieval.

In real situations, knowledge can be reasonably expressed in terms of fuzzy sets whose descriptions are imprecise and vague. In such uncertain situations, fuzzy set theory is useful to grade the degree of membership of objects within [0,1] [16]. A case is an object that can be represented in terms of its several features. A case is said to be fuzzy if at least one of its features is described using fuzzy linguistic terms [17]. Usually, some features of an object can be suitably represented using fuzzy linguistic terms rather than crisp values. In addition, the weights of case features can be rated in terms of linguistic terms instead of using sharp numerical values.

Evaluating the weight of case features is one of the crucial challenges in CBR. It requires domain knowledge elicitation to make effective the reasoning process. In this study, fuzzy AHP is proposed to weight case features. The AHP is a systematic approach to elicit and represent experts' domain knowledge for prioritizing case features [18, 19]. Using pairwise comparison, the preference of one attribute over the other is expressed in terms of linguistic terms like "equally preferred", "moderately preferred", "strongly preferred", etc. These terms are purely subjective to define their boundaries due to human judgement. The conventional AHP was extended into fuzzy AHP to articulate this substantial problem in real-life decision-making [20, 21].

2. Proposed system

Figure 1 presents the flow diagram of the proposed DSS. In the proposed system, it is assumed that similar products are anticipated to have similar cost estimates.

2.1 Case feature selection and case representation

Identifying important case features, which can influence the costs of products, is the primary crucial work at an early stage of product development. In this study, machining rotating shafts is taken into account. First, three major product features are identified, namely, workpiece related, finished product quality and types of operations. Then these primary features are hierarchically branched into sub-features as presented in Section 3.

The identified case features are expressed in terms of numerical values, nominal values and fuzzy linguistic terms. The combination of these features is used to represent the cases in *n*-dimensional vector space. The cases are represented with the help of an OO approach using the Java programming language. The OO case representation approach is popular and widely accepted by software developers because of its comprehensiveness, flexibility, reusability and easiness to understand.

2.2 Weighting case features

The weights of features are evaluated using fuzzy AHP as stated before. Table 1 presents the relationships among the fuzzy AHP-based linguistic terms, their equivalent fuzzy numbers and their standard forms within [0, 1]. The fuzzy numbers and their reciprocals are converted into their corresponding standard fuzzy numbers by dividing them with the maximum value of the universe of discourse, which is 10 in this case [22]. The standardized fuzzy numbers are transformed into their corresponding crisp values by adopting a fuzzy ranking approach recently proposed by Chen and Chen [22]. Equation (1) is applied to defuzzify the required fuzzy numbers. This approach is simple; it avoids the limitations of other methods; and prefers the most precise fuzzy numbers when different fuzzy numbers have an identical mean value. After determining the crisp score of any trapezoidal/triangular fuzzy number, A_{cs}, the classical AHP approach is applied for prioritizing case features.

$$A_{cs} = \frac{A_{mean}}{1+A_{std}} \tag{1}$$

Where A_{mean} and A_{std} are the mean and standard deviation of a standardized fuzzy number respectively.

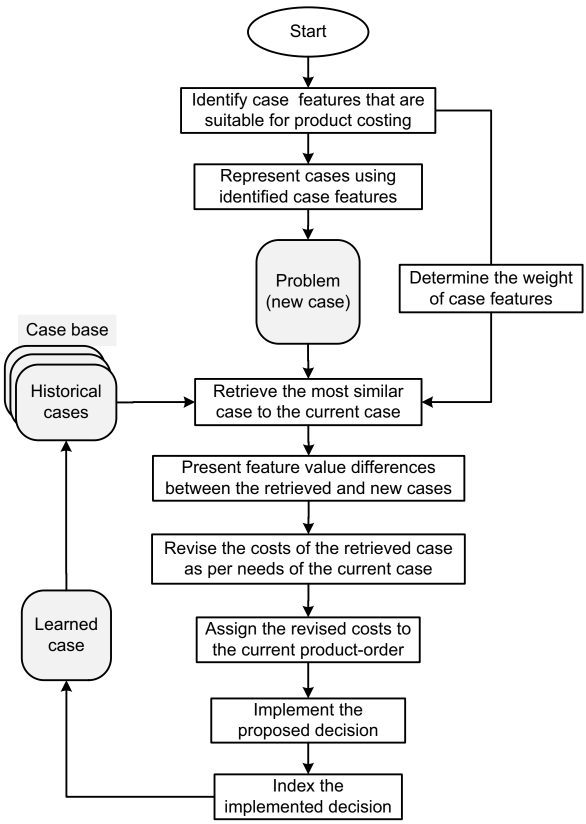

Figure 1. Flow diagram of the proposed DSS

Table 1. Linguistic terms, their equivalent fuzzy numbers and standardized fuzzy numbers.

AHP-based fuzzy linguistic terms	Equivalent		Standardized	
	Fuzzy number	Fuzzy reciprocal	Fuzzy number	Fuzzy reciprocal
Exactly equal	(1, 1, 1)	(1, 1, 1)	(0.1, 0.1, 0.1)	(1/10, 1/10, 1/10)
Equally preferred	(1, 1, 2)	(1/2, 1, 1)	(0.1, 0.1, 0.2)	(1/20, 1/10, 1/10)
Intermediate	(1, 2, 3)	(1/3, 1/2, 1)	(0.1, 0.2, 0.3)	(1/30, 1/20, 1/10)
Moderately preferred	(2, 3, 4)	(1/4, 1/3, 1/2)	(0.2, 0.3, 0.4)	(1/40, 1/30, 1/20)
Intermediate	(3, 4, 5)	(1/5, 1/4, 1/3)	(0.3, 0.4, 0.5)	(1/50, 1/40, 1/30)
Strongly preferred	(4, 5, 6)	(1/6, 1/5, 1/4)	(0.4, 0.5, 0.6)	(1/60, 1/50, 1/40)
Intermediate	(5, 6, 7)	(1/7, 1/6, 1/5)	(0.5, 0.6, 0.7)	(1/70, 1/60, 1/50)
Very strongly preferred	(6, 7, 8)	(1/8, 1/7, 1/6)	(0.6, 0.7, 0.8)	(1/80, 1/70, 1/60)
Intermediate	(7, 8, 9)	(1/9, 1/8, 1/7)	(0.7, 0.8, 0.9)	(1/90, 1/80, 1/70)
Extremely preferred	(8, 9, 10)	(1/10, 1/9, 1/8)	(0.8, 0.9, 1.0)	(1/100, 1/90, 1/80)

Additionally, fuzzy case features, which are described in terms of linguistic terms, are converted into fuzzy numbers in [0, 1] using eleven conversion scales (Figure 2) by referring to Chen and Hwang [23]. The variable x is any real number in [0, 1] and $\mu(x)$ is the degree of membership of x to the linguistic terms.

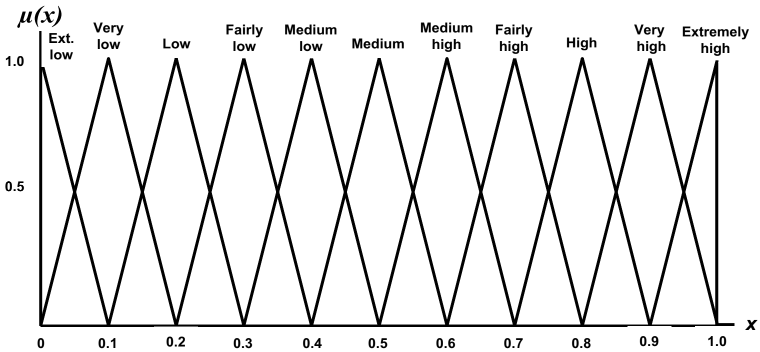

Figure 2. Conversion of linguistic case features into fuzzy numbers

2.3 Case retrieval and revision

This study uses the nearest neighbor (NN) matching function, which is popular and simple, to measure case similarities. This similarity measure is based on the inverse of the Euclidean distance between any two cases. The Euclidean distance between a new case p and a prior case q, $dist(p,q)$ can be calculated:

$$dist(p,q) = \sqrt{\sum_{i=1}^{n}[w_i \ast dist(a_i^p, a_i^q)]^2},\ dist(a_i^p, a_i^q) \in [0,1] \tag{2}$$

Where n is the number of case features; w_i is the normalized weight of the ith attribute; and a_i^p and a_i^q are the values of the ith attribute for cases p and q respectively.

From Equation (2), the distance between the individual features of two cases, $dist(a_i^p, a_i^q)$, can be determined as follows.

In the case of discrete and continuous numerical features:

$$dist(a_i^p, a_i^q) = \frac{|a_i^p - a_i^q|}{a_{i,max} - a_{i,min}},\ a_i^p \ \&\ a_i^q \in [a_{i,min}, a_{i,max}] \tag{3}$$

Where $a_{i,min}$ and $a_{i,max}$ are the minimum and maximum value of the ith attribute respectively.

For categorical attributes:

$$dist(a_i^p, a_i^q) = |a_i^p - a_i^q| = \begin{cases} 1\ if\ \ a_i^p \neq a_i^q \\ 0\ if\ a_i^p = a_i^q \end{cases} \tag{4}$$

In the case of fuzzy features, the authors consider trapezoidal fuzzy numbers and Equation (5) is adopted from Wei and Chen [24]. Their proposed method combines the concepts of geometric distance, the perimeter and the height of a trapezoidal fuzzy number. In this case, the value height is 1. When trapezoidal fuzzy numbers are in a standard form $a_i^p = (a_{i,1}^p, a_{i,2}^p, a_{i,3}^p, a_{i,4}^p)$ and $a_i^q = (a_{i,1}^q, a_{i,2}^q, a_{i,3}^q, a_{i,4}^q)$; and $0 \leq a_{i,1}^p \leq a_{i,2}^p \leq a_{i,3}^p \leq a_{i,4}^p \leq 1$ and $0 \leq a_{i,1}^q \leq a_{i,2}^q \leq a_{i,3}^q \leq a_{i,4}^q \leq 1$.

$$dist(a_i^p, a_i^q) = 1 - \left[\left(1 - \sum_{k=1}^{4} \frac{|a_{i,k}^p - a_{i,k}^q|}{4}\right) * \frac{min\left(per(a_i^p), per(a_i^q)\right)+1}{max\left(per(a_i^p), per(a_i^q)\right)+1}\right] \quad (5)$$

Where $per(a_i^p)$ and $per(a_i^q)$ are the perimeters of trapezoidal fuzzy attributes of cases p and q respectively.

Referring to Equation (2), the values of $dist(a_i^p, a_i^q)$ are within [0, 1]. The maximum Euclidean distance, $dist_{max}(p,q)$, is found when all the values of $dist(a_i^p, a_i^q) = 1$; and the minimum Euclidean distance, $dist_{min}(p,q)$, is found when all the values of $dist(a_i^p, a_i^q) = 0$ i.e. when $p = q$. Referring Equation (2), the $dist_{max}(p,q)$ and $dist_{min}(p,q)$ values can be determined as $dist_{max}(p,q) = \sqrt{\sum_{i=1}^{n} w_i^2}$ and $dist_{min}(p,q) = 0$

Because distance and similarity are inversely related, the similarity between two cases p and q, $sim(p,q)$, can be found as follows [25]:

$$sim(p,q) = 1 - dist(p,q) \quad (6)$$

The minimum similarity measure, $sim_{min}(p,q) = 1 - \sqrt{\sum_{i=1}^{n} w_i^2}$ and using the same approach, the maximum similarity between any two parts, $sim_{max}(p,q) = 1$, and $sim(p,q) \in [sim_{min}(p,q), 1.0]$. Using these relationships, any retrieved case with a higher similarity value to the current problem is selected for reuse and adaptation.

The proposed DSS not only measures case similarities for case retrieval but also presents the difference in each case attribute for case revision. Then the case revision can be done by human experts using any other cost estimation approaches.

3. Numerical example

This numerical example is illustrated using a lathe-machining center. Suppose the machining center produces a number rotating shafts for different purposes. To represent cases using an OO method, twelve product features are proposed. The features are structured hierarchically as indicated in Table 2 to prioritize these proposed case features. The hierarchy incorporates three major attributes: (1) workpiece related; (2) required operations types; and (3) product quality requirements. These three major features are sub-divided into their corresponding sub-features. The normalized weights of the major features and sub-features (under their preceding features) at their specific levels are evaluated using fuzzy AHP. To do this evaluation, the concepts from Table 1 and Equation (1) are applied. The normalized weights of the twelve case features are proportionally calculated as indicated in the fourth column of Table 2.

The twelve product features are represented using numerical, nominal and fuzzy data (see Table 3). Length (L) and diameter (D) are represented using numerical values in millimeter. Material type (Ma) is represented using linguistic terms to describe the expensiveness of construction materials. Similarly, tolerance limit (Tl), surface smoothness (Ss) and durability (Du) are described in terms of linguistic terms and the terms are converted into fuzzy numbers referring to Figure 2. Machining operation types such as turning (Tu), facing (F), thread-cutting (Th), drilling (Dr), boring (B) and tapping (Ta) are expressed using nominal values of {0,1}. Additionally Table 3 indicates three product-orders (P1-P3) as new cases and two training samples (T1 and T2) as prior cases. Assume CC1 and CC2 are costs assigned initially to T1 and T2 respectively and CC3 is the revision of CC2 as P1 is retrieved as a prior case (Table 4).

Table 2. Hierarchy of case features and their normalized weights

Major feature	Middle feature	End feature	Normalized weight calculation	Normalized weight (w_i)
Workpiece (0.475)	-	Length (0.278)	0.475x0.278	0.132
		Diameter (0.248)	0.475x0.248	0.118
		Material (0.475)	0.475x0.475	0.226
Operation types (0.277)	External (0.541)	Turning (0.519)	0.277x0.541x0.519	0.078
		Facing (0.176)	0.277x0.541x0.176	0.026
		Threading (0.306)	0.277x0.541x0.306	0.046
	Internal (0.459)	Drilling (0.475)	0.277x0.459x0.475	0.060
		Boring (0.277)	0.277x0.459x0.277	0.035
		Tapping (0.248)	0.277x0.459x0.248	0.031
Quality (0.248)	-	Tolerance (0.439)	0.248x0.439	0.109
		Surface (0.296)	0.248x0.296	0.073
		Durability (0.265)	0.248x0.265	0.066

Table 3. Structured features of cases/products

P	L	D	Ma	Tl	Ss	Du	Tu	F	Th	Dr	B	T
P1	500	180	0.7,0.8,0.9	0.4,0.5,0.6	0.8,0.9,1.0	0.7,0.8,0.9	1	0	1	1	1	0
P2	820	330	0.3,0.4,0.5	0.6,0.7,0.8	0.4,0.5,0.6	0.6,0.7,0.8	1	1	0	1	0	1
P3	520	200	0.7,0.8,0.9	0.4,0.5,0.6	0.7,0.8,0.9	0.8,0.9,1.0	1	0	0	1	1	1
T1	850	350	0.3,0.4,0.5	0.7,0.8,0.9	0.4,0.5,0.6	0.5,0.6,0.7	1	0	1	1	0	1
T2	450	150	0.7,0.8,0.9	0.3,0.4,0.5	0.8,0.9,1.0	0.7,0.8,0.9	1	1	1	1	0	0

The similarity between the new cases and training samples is calculated using Equations (2) - (6). The results are compiled in Table 4.

Table 4. Summarized results of the proposed DSS

Product	Most similar case	Similarity-value	Cost to be revised	No. of cases in case-base
P1	T2	0.952	CC2	2
P2	T1	0.944	CC1	3
P3	P1 (new case)	0.939	CC3 (revised CC2)	4

4. Conclusions and future works

In the past, a combination of fuzzy CBR and AHP was not applied to estimate the costs of new products. In this study, a novel and promising DSS is proposed to articulate this problem situation. The DSS is capable to retrieve the most similar prior cases to the current case and update its case base as new cases enter to the system (Table 4). The retrieved case is expected to have similar cost estimates to the new order. This DSS recommends the case revision to be done by human experts using other cost estimation methods. This is because the aim of the DSS is not to replace the tasks of experts; however, it should be to assist them in complex situations. Searching the most similar previous cost estimate is the most cumbersome work for human reasoners at the early stages of new product development processes.

In the proposed DSS, fuzzy cases have been represented using the combination of numerical values, nominal values and fuzzy linguistic terms. This kind of unified representation emulates human thought to process imprecise and vague information in the real world. Although the numerical example has been illustrated using a few new cases and training samples, the proposed DSS can address any number of product-orders scheduled, training samples and case features. Because in real manufacturing situations, products mix variation is very high.

In the future, the proposed DSS will be tested using realistic historical data in order to validate its accuracy.

References

[1] P. Duverlie, P. and J.M. Castelain, Cost estimation during design step: Parametric method versus case based reasoning method, *The International Journal of Advanced Manufacturing Technology*, vol. 15, 1999. pp. 895-906.

[2] C. Rush and R. Roy, Analysis of cost estimating processes used within a concurrent engineering environment throughout a product life cycle, In: *7th ISPE International Conference on Cocurrent Engineering: Research and Applications*, Lyon, 2000, pp. 58-67.

[3] S. Cavalieri, P. Maccarrone and R. Pinto, Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry, *International Journal of Production Economics*, vol. 91, 2004, pp. 165-177.

[4] J. Stjepandić, W.J.C. Verhagen, H. Liese, P. Bermell-Garcia, Knowledge-Based Engineering, In J. Stjepandić et al. (eds): *Concurrent Engineering in the 21st Century: Foundations, Developments and Challenges*, Springer: International Publishing, Switzerland, 2015. pp. 255-286.

[5] A. Niazi, J.S. Dai, S. Balabani and L. Seneviratne, Product cost estimation: Technique classification and methodology review, *Journal of Manufacturing Science and Engineering*, vol. 128, 2006, pp. 563-575.

[6] A.C. Caputo and P.M. Pelagagge, Parametric and neural methods for cost estimation of process vessels, *International Journal of Production Economics*, vol. 112, 2008, pp. 934-954.

[7] A. Layer et al., Recent and future trends in cost estimation, *International Journal of Computer Integrated Manufacturing*, vol. 15, 2002, pp. 499-510.

[8] Y.F. Zhang, J.Y.H. Fuh and W.T. Chan, Feature-based cost estimation for packaging products using neural networks, *Computers in Industry*, vol. 32, 1996, pp. 95-113.

[9] A.E. Smith and A.K. Mason, Cost estimation predictive modeling: Regression versus neural network. The Engineering Economist, vol. 42, 1997, pp. 137-161.

[10] G.-H. Kim, S.-H. An and K.-I. Kang, Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning, *Building and Environment*, vol. 39, 2004, pp. 1235-1242.

[11] S.-H. An, G.-H. Kim and K.-I. Kang, A case-based reasoning cost estimating model using experience by analytic hierarchy process, *Building and Environment*, vol. 42, 2007, pp. 2573-2579.

[12] J. Kolodner, Improving human decision making through case-based decision aiding, *AI Magazine*, vol. 12, 1991, pp. 52-68.

[13] J. Kolodner, An introduction to case-based reasoning, *Artificial Intelligence Review*,vol.6, 1992, pp.3-34.

[14] A. Aamodt and E. Plaza, Case-based reasoning: Foundational issues, methodological variations, and system approaches, *AI Communications*, vol. 7, 1994, pp. 39-59.

[15] R.L. de Mantaras, Case-Based Reasoning, In G. Paliouras et al. (eds.): *Machine Learning and Its Applications*, Springer-Verlag, Berlin and Heidelberg, 2001, pp. 127-145.

[16] L.A. Zedah, Fuzzy sets, *Information and control*, vol. 8, 1965, pp. 338-353.

[17] H.J. Zimmermann, *Fuzzy Set Theory—and Its Applications,* Springer, New York, 2001.

[18] C.-S. Park and I. Han, A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction, *Expert Systems with Applications*, vol. 23, 2002, pp. 255-264.

[19] T.L. Saaty, How to make a decision: The analytic hierarchy process, *Interfaces*, vol.24,1994, pp. 19-43.

[20] P.J.M.Van Laarhoven and W. Pedrycz, A fuzzy extension of Saaty's priority theory, *Fuzzy Sets and Systems*, vol.11, 1983, pp. 229-241.

[21] J.J. Buckley, Fuzzy hierarchical analysis, *Fuzzy Sets and Systems,* vol. 17, 1985, pp. 233-247.

[22] S.-M. Chen and J.-H. Chen, Fuzzy risk analysis based on ranking generalized fuzzy numbers with different heights and different spreads, *Expert Systems with Applications*, vol. 36, 2009, pp. 6833-6842.

[23] S.-J. Chen and C.-L. Hwang, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, Berlin Heidelberg, 1992.

[24] S.-H. Wei and S.-M. Chen, A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *Expert Systems with Applications*, vol. 36, 2009, pp. 589-598.

[25] T.W. Liao, Z. Zhang and C.R. Mount, Similarity measures for retrieval in case-based reasoning systems, *Applied Artificial Intelligence: An International Journal*, vol. 12, 1998, pp. 267-28814.