A group-decision approach for evaluating educational web sites

Gwo-Jen Hwang\textsuperscript{a,}*, Tony C. K. Huang\textsuperscript{a}, Judy C.R. Tseng\textsuperscript{b}

\textsuperscript{a}Information Management Department, National Chi Nu University, Nantou, Puli, Taiwan 545, Taiwan, ROC
\textsuperscript{b}Department of Computer Science and Information Engineering, Chung-Hua University, Hsinchu 300, Taiwan, ROC

Received 29 January 2003; accepted 15 May 2003

Abstract

With the advent of network technologies, many educational web sites have been developed to assist students in the learning of subjects on computer networks. However, without proper aid, students may have difficulty in selecting appropriate web sites, that are of benefit to them; hence, studying, evaluating and recommending educational web sites has become an important and challenging issue. In this paper, a group-decision approach is proposed for evaluating educational web sites. Several soft computing technologies have been employed in the approach, including fuzzy theory, grey system and group decision method. A computer-assisted web site evaluation system, EWSE (Educational Web Site Evaluator), has been developed, based on an experimental approach, which is capable of selecting the proper criteria for an individual web site and achieves greater accuracy when evaluating results.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Computer-assisted learning; Web site evaluation; Group decision support; Educational web site; e-Learning

1. Introduction

Since 1990, the trajectory of Internet technology has lead to the rapid growth of web sites. Investigative inspection of the Computer Industry Almanac in 1999 shows that the global expansion of the Internet has exceeded 259 million people, the implication being, the Internet world has now become an enormous virtual society. According to interNIC statistics, over 800

* This study is supported in part by the National Science Council of the Republic of China under contract number NSC-91-2520-S-260-002.
* Corresponding author. Tel.: +886-915-396558; fax: +886-49-291-5205.
E-mail addresses: gihwang@ncnu.edu.tw (G.-Jen Hwang), judycrt@chu.edu.tw (J.C.R. Tseng).
million web sites have been built since December 1999. Therefore, the “information overloading” problem becomes a barrier for Internet users when seeking information. Without sufficient aid, Internet users may spend excess time browsing uninformative web sites thus, increasing the seeking costs and the bandwidth loading.

One of the most important and popular branches of network applications is the development of educational web sites, which assist students in learning subjects via computer networks (Wong et al., 1998; Antao et al., 2000; Yoshikawa et al., 2000). In 1989, Johnson, Neste, and Duncan proposed a software design and development research program called Microcomputer Intelligence for Technical Training (MITT). They presented MITT Writer, an authoring environment for building intelligent tutoring systems for computer courses, which depicted a practical application of artificial intelligence (AI) within the technical training component. Vasandani and Govindaraj (1994, 1995), proposed an intelligent tutoring system that helps to organize system knowledge and operational information which enhances operator performance. Gonzalez and Ingraham (1994) developed an intelligent tutoring system, which is capable of determining exercise progression and remediation automatically during a training session, according to the students’ past performance. Harp, Samad, and Villano (1995) employed the technique of neural networks to model the behavior of students, within the context of an intelligent tutoring system. They employed self-organizing feature maps to capture the possible states of student knowledge from an existing test database.

In Taiwan, the CORAL (Cooperative Remotely Accessible Learning) project (Sun and Chou, 1996) was initiated by a research group at National Chaio Tung University, which consists of eight sub-tasks: (1) The study of network-based tutoring systems, including pattern recording, remote data retrieval and access control on networks. (2) The study of a network-based learning environment, including real-time monitoring and tutoring process control. (3) The study of wide area networked CAL’s, including feasibility, scalability and architecture. (4) Testing and evaluation of network-based CAL’s, including motivation and cognition analysis. (5) The study of interface design for network-based CAL’s, including screen layout, icons/windows design and knowledge visualization. (6) The study of student modeling for network-based CAL’s, including analysis of hypertext navigation and communication patterns. (7) The study of a knowledge-based system to support cooperative tutoring process control, including knowledge representation for student characteristics, real-time analysis of student behaviors and the dynamic arrangement of tutoring schedule. (8) The study of interaction pattern analysis, including analysis of social context.

One of the branches of CORAL is the ITED (Intelligent Testing, Evaluation and Diagnosis) system, which can detect the on-line status of the students and provide learning advice (Hwang, 1998, 2002a,b).

In 1998, Rowe and Galvin employed planning methods, consistency enforcement objects and structured menu tools, for constructing intelligent simulation-based tutors using procedural skills. It can be seen that the study of developing intelligent tutoring systems and learning environments has become an important issue for both the fields of computer science and education in recent years.

As it is difficult to judge the quality of a web site, the student may require assistance in selecting proper web sites that are of greater assistance. A well-designed web site evaluation procedure not only helps the end users (i.e. students and teachers who are looking for suitable educational resources) to find best-fit educational web sites, but also resolves the information-overloading problem.

There have been some vote-oriented web-site evaluation systems, such as e-Oscar (http://www.e-oscar.com.tw), Yahoo-Kimo (http://www.kimo.com.tw), BizRate.com (http://www.bizrate.com).
com), etc., which employ user voting mechanisms to evaluate web sites. Some evaluation schemes employ an integer ranging from 1 to 5 as the rating of an evaluated web site, e.g. e-Oscar. Some use a fixed set of criteria for evaluating web sites, e.g. BizRate.com. Such random and general-purpose rating mechanisms may give rise to several problems:

1. The users can only evaluate a web site with their inarticulate impressions, instead of considering several important criteria which may affect the use of the web sites; such as functionality, user interface, extensibility, security, user guidance, content quality, etc.
2. It is inappropriate to use a fixed set of criteria to evaluate all of the web sites with different purposes or usage; e.g., some educational web sites that provide excellent tutoring materials may not place testing or evaluation functions as their major features (Sun & Chou, 1996); alternatively, some educational web sites are designed for providing testing, evaluation and diagnosis functions (Hwang, 2002a). Therefore, it is necessary to employ different sets of criteria when evaluating educational web sites with different properties or design purposes.
3. Most users only have abilities to evaluate web sites within part of the criteria. To take the overall criteria into consideration, it is necessary to have input and participation of domain experts.
4. While taking multiple criteria into consideration, the users and the domain experts may have difficulty in rating web sites and integrating the ratings, without any professional assistance.

To cope with these problems, we propose a group-decision approach for evaluating educational web sites in this paper. Several soft computing technologies have been employed in our approach, including fuzzy theory, grey system and group decision method. To assist users and domain experts in evaluating educational web sites, an evaluation assistance system EWSE (Educational Web Site Evaluator) has been developed on the Internet. An experiment has been conducted to evaluate the performance of EWSE. From the experimental results, it can be seen that EWSE is able to provide precise and multi-dimensional information for Internet users to select desired web sites.

2. Backgrounds and relevant research

In Taiwan, the Ministry of Education and the National Science Council have tried to collect URLs of educational web sites that provide subject contents, assessment functions and educational discussion boards. The purpose of collecting those educational web sites is to recommend high quality educational resources to students and teachers. Therefore, it is very important to develop an effective and objective approach to evaluating educational web sites.

As the process of evaluating an educational web site with group-decision approach is very complicated, it is almost not possible to employ a single theory or technology in all of the evaluation steps. For example, fuzzy theory provides a more natural way for educational experts to describe the criteria as well as the corresponding importance degrees for evaluating educational web sites; while gray system theory is known to be a good way to make predictions and perform relevance analysis that are helpful to the selection of candidate criteria.
Moreover, in the group-decision evaluation process, the evaluators will require a manual interface to help them comparing and selecting the most suitable set of criteria for each evaluated website, and hence the pair-wise comparison matrix of Analytic Hierarchy Process is adopted. In the following subsections, those theories and technologies will be introduced in details.

2.1. Decision support technology

In 1960, Simon proposed intelligence activity, design activity, and choice activity as the three steps of decision-making. In 1971, Gorry and Scott Morton followed the theory of Simon and proposed a classical architecture of decision support systems (DSS). One branch of DSS, i.e. group decision support system (GDSS) has been proposed to improve group performance of various decisions (Huber, 1984).

Group decision-making methods have been developed extensively and it is widely understood that most decision problems inherently involve more than one criterion. Therefore, multiple criteria decision-making (MCDM) (Keeney & Raiffa, 1976) has been proposed, to make decisions in the presence of multiple and conflicting criteria. MCDM may be used to select or generate a ‘best alternative,’ from a finite set of existing alternatives.

2.2. Fuzzy theory

In Ibrahim (2000, 2001), assessing quality using fuzzy logic has been proposed, where the objective was to enables users with varied interests and objectives to make their own judgments based on their own criteria of the quality, rather than depending on the opinions of others.

However, for a student who is inexperienced in searching educational resources on the Internet, it would be better to take both the opinion of educational experts and the student’s own criteria of the quality into considerations. Therefore, in our approach, the opinions from domain experts or experienced Internet users have been taken into considerations, and fuzzy theory (Zadeh, 1965) has been employed to represent each participant’s opinions and to deduce the ratings of the web sites from the data collected.

Fuzzy theory has been used to cope with various evaluative problems (Feng & Xu, 1999; Klir & Yuan, 1995). It is an efficient and effective method to represent the uncertainty and vague terms in an assessment environment. A fuzzy set is a class of objects with a set of membership grades, which are normally within [0, 1] range. When the grade of membership for an object is ‘1’, it denotes that the object absolutely belongs to the class. Alternatively, ‘0’ means that the object absolutely does not belong to the class. Borderline cases are assigned to the values between ‘0’ and ‘1’. Precise membership grades cannot convey any absolute meaning as they are context-dependent and can be subjectively assessed. Echauz and Vachtsevanos (1995) have presented a fuzzy grading method that employ student and instructor performance measures, to produce a fair mark distribution. The same idea can be applied to the evaluation of educational web sites, by adopting the measures of domain experts or qualified Internet users.

A fuzzy set $F$ is characterized by a membership function $\mu_F: U \rightarrow [0, 1]$ that assigns to each object $x$ of $U$, a degree of membership $\mu_F(x)$, in the continuum $[0, 1]$. The degree of membership is also referred to as the degree of fulfillment or possibility. Thus, a fuzzy set is also a collection of objects, but with the understanding that some objects can “belong more” to it than others do.
For example, in the fuzzy set of “birds,” of the universe of “animals,” a platypus and a dove might have degrees of membership, 0.45 and 1, respectively. By blurring the boundaries of crisp set, it is possible to represent arbitrary collections of objects outside the area of mathematics. All fuzzy sets on \( U \) is denoted as \( \Gamma(U) \). The cardinality of fuzzy set \( A \), is defined by the sum of the degree of membership, i.e., \( \text{Card}(A) = \sum_{i=1}^{n} \mu_{A}(u) \).

Let \( X, Y \) be the universes and \( \Gamma(U) \) be the set of all fuzzy sets in \( Y \), \( A = \{(x, y), \mu_{A}(x, y)\} \) is called a fuzzy relation on \( XY \). \( f: X \rightarrow \Gamma(Y) \), is called a fuzzy function form \( X \) to \( Y \). The fuzzy functions and fuzzy relations are inter-related and both, can be represented by the membership matrix. The composition of two fuzzy relations \( P(X, Y) \) and \( Q(X, Y) \) can be defined in terms of an operation on the membership matrix \( P \) and \( Q \) that resembles matrix multiplication. This operation involves exactly the same combinations in the matrix multiplication are replaced with other operations. These alternative operations represent in the given context, the corresponding operations of fuzzy set intersection and union. In the Max–Min composition, for example, the multiplication and addition are replaced with the Min and Max operations respectively. The result of composition on \( P(X, Y) \) and \( Q(X, Y) \) is a fuzzy relation \( R(X, Y) \), which can be denoted as \( R = P \circ Q \).

Before performing the fuzzy computation, the input crisp values must be fuzzified first. After deriving a fuzzy result, some ‘defuzzification’ mechanisms, are needed to compose the fuzzy outputs into a crisp result. One useful defuzzification method is the centroid approach, \( Z = \frac{\sum_{j=1}^{n} \mu_{z}(w_j) w_j}{\sum_{j=1}^{n} \mu_{z}(w_j)} \), where \( n \) is the fragment counts of outputted space, \( Z \) is the crisp output, and \( w_j \), the fuzzy output, is the object of fuzzy set \( Z \) (Klirm & Yuan, 1995).

2.3. Grey system theory

Grey System (GS) is a theory that deals with uncertain, poor, or incomplete information (Deng, 1989). A system containing both known and unknown information is called a grey system. For example, the body system is a grey system, because certain characteristics of the body, such as height, weight, eyesight etc. are known and some physiological phenomena in the human body is unknown. Currently, grey system has been applied to many fields, including economics, agriculture, medicine, earthquakes and industry, etc.

The main issues of GS theory include Grey Relational Analysis (GRA), Grey Prediction, and Grey Decision. In this paper, we employ GRA [20] to determine the criteria needed for evaluating web sites. The procedure of performing GRA is illustrated in Fig. 1.

2.4. Analytic hierarchy process

In complex MCDM problems, decision makers can determine some dimensions based on the importance of each criterion. Consequentially, there must be a method that elicits the decision maker’s preference and their knowledge, which allows an understanding of the sequence of the alternatives. Analytic Hierarchy Process (AHP) was developed by Saaty in 1980, and has become a useful evaluating technique for solving MCDM problems.

Let the weights of evaluating items be \( w_1, w_2, \ldots, w_n \). In AHP method, a pair-wise comparison matrix is given as follows:
The processes of data $x_i(k) = (x(1), x(2), ..., x(k))$

Computing $\Delta_0 = ||x_i(k) - x_j(k)||$ for $j = 1, 2, ..., m, k = 1, 2, ..., n$

Computing $\Delta_{max} = \min_{x_i(k), x_j(k)} \min_{k} ||x_i(k) - x_j(k)||$ and $\Delta_{min} = \max_{x_i(k), x_j(k)} \max_{k} ||x_i(k) - x_j(k)||$

Counting $\Delta_0(k) = ||x_i(k) - x_(j(k))||$

Setting the $\epsilon$ value (depends on practical requirement) up

Counting the $\beta$ values of weights by the method of the entropy

Solving the values of grey relational coefficient:
$\gamma(x_i(k), x_j(k)) = (\Delta_{min} + \epsilon \cdot \Delta_{max}) / (\Delta_0(k) + \epsilon \cdot \Delta_{max})$ for $j = 1, 2, ..., m$

Solving the grey relational degrees:
$\gamma(x_i, x_j) = \sum_{k=1}^{n} \beta_i \gamma(x_i(k), x_j(k))$

Ranking the sequences of each factor

Fig. 1. Procedure of Grey Relational Analysis.

$$A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix} = \begin{bmatrix}
    \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\
    \frac{w_2}{w_1} & \frac{w_2}{w_2} & \cdots & \frac{w_2}{w_n} \\
    \vdots & \vdots & \ddots & \vdots \\
    \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & \frac{w_n}{w_n}
\end{bmatrix}$$

One can treat the elements of eigenvector as weights of the item. When a pair-wise comparison matrix is constructed, decision makers are offered questions such as, “What is the degree of importance for item $X$ in comparing with item $Y$?” In our application, item $X$ and item $Y$ are two criteria for evaluating web sites. The answers to these questions are various numerical values, corresponding to linguistic measures: as shown in Table 1. When a numerical value is placed in $a_{ij}$, its reciprocal is placed in $a_{ji}$. That is, $a_{ij} = 1$, $a_{ji} = 1/a_{ij}$, and all elements for this matrix have positive values, which guarantee the matrix has positive eigenvalues, and all eigenvector elements with maximum eigenvalues are positive.

If $a_{ij} = w_i/w_j$, then $A'$ coincides with $A$, and $a_{ij}a_{jk} = \left(\frac{w_i}{w_j}\right) \times \left(\frac{w_j}{w_k}\right) = a_{ik}$. In this ideal case, one can lead an eigenvector that gives the relative weights of items. AHP also provides a measure to check if the answers of the decision makers follow the reciprocal property. Saaty defines this measure as
the consistency ratio $C.R. = \frac{(\lambda_{\text{max}} - n)}{[(n - 1)\rho]}$, where $\lambda_{\text{max}}$ is the largest eigenvalue of the matrix, and $\rho$ is a random index (Saaty, 1980).

It can be seen that AHP provides a helpful tool for the evaluators in selecting proper criteria by making a series of pair-wise comparisons.

2.5. Internet experience model

In Lu (2000), an Internet Experience Model was proposed, which made some suggestions about how to establish a popular web site:

1. Gather experiences: Find “what are interesting to people,” is the target of this stage. If we want to search a keyword, “Information Technology”, by the Google search engine, more than one million records will be found. Among those records, the information of a successful web site must be able to attract the Internet users. If a user visits the same web site for the second time, it implies that the content of the web site is interesting to the user, which is an experience to be gathered.

2. Learning from experiences: Establishing relationships among users and web sites is the target of this stage. If a user likes visiting a web site, a relationship between the user and the web site can be established; furthermore, the user may become a member who will be a good assistant, promoting what are important or interesting to users. Therefore, it is important to establish such relationships between each user and the web-site administrator. For example, building a virtual community of education, providing personal services and establishing friendship, are several alternative approaches for building such relationships.

3. Enhancing the belief degree with the experiences: To create “belief,” is the target of this stage. The uppermost principle of Internet marketing is, “the marketing is not your job, but
is your follower”. The power of the virtual community is huge. The six keys that need to be noted are: security technology, merchant legitimacy, robust order fulfillment, tone and ambiance, users’ control, and collaboration.

3. Group-decision model of EWSE

As the number of educational web sites rapidly increases, it is important to propose a method for the users to choose appropriate educational web sites. In this section, an evaluation assistance system, EWSE, is proposed. EWSE employs a fuzzy group-decision approach to assist users and domain experts to evaluate educational web sites. The whole concept of the approach is depicted in Fig. 2. In the following, we shall present each phase in details.

3.1. Phase 1: defining the evaluation standards

Before starting the task of web site evaluation, it is important to define a set of criteria for each evaluation item. It is difficult for a domain expert to specifically express his/her criteria sets for web site evaluation, and the criteria sets proposed by different experts might not be the same. Therefore, we employ the grey system to integrate the subjective opinions from several experts and generate the most acceptable criteria set. Tables 2–4 depict an example of candidate criteria for three evaluation items: “the design of user interface”, “the quality of instructional contents” and “the assessment functions”.

As the experts propose various criteria for evaluating web sites, each expert is asked to evaluate each criterion based on a 1–5 ranking mechanism. An illustrative example is given in Table 5,
where 1, 2, 3, 4 and 5 represent very negative, negative, average, positive, and very positive, respectively. For example, in the first column concerning the faithfulness of applying the criterion, 1 denotes very unfaithful, 2 denotes unfaithful, 3 denotes average, 4 denotes faithful and 5 denotes very faithful.

3.2. Phase 2: adjusting the weight of each evaluator

In a group decision process, the final decision must be able to represent the opinions of most participants. A weight value, \( w_p \) [0, 1], is used to represent the performance of evaluator \( P \). Initially, \( w_p = 1 \) for each participant. If an evaluator often proposes opposite opinions to the group decisions, he (or she) is considered to be indecisive, (furthermore, this person’s opinion cannot be used to signify that of the group), and correspondingly the weight will be decreased. On the contrary, for the evaluators who always propose opinions close to the group decisions, the corresponding weights are increased. The increment and decrease are computed by the following function:

<table>
<thead>
<tr>
<th>No.</th>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Quality of web-page design</td>
<td>Are the objects (text, images or icons) of each web page located in the suitable position? Does the system provide an easy-to-use interface for the learner to answer a question, or open a new window, one which may become a burden for the learner to locate the answer window?</td>
</tr>
<tr>
<td>1-2</td>
<td>Suitableness of web-link design</td>
<td>When browsing a web page, can the learner always find a link to access relevant web pages or return to main menu?</td>
</tr>
<tr>
<td>1-3</td>
<td>Usability</td>
<td>Can a student learn how to operate the user interface quickly? Or are there some functions that confuse the learner?</td>
</tr>
<tr>
<td>1-4</td>
<td>Response time of the user interface</td>
<td>Are the texts, images or other media files too large such that the users have to wait an excessive amount of time when downloading the teaching materials?</td>
</tr>
<tr>
<td>1-5</td>
<td>Quality of media presentation</td>
<td>Is the instructional content presented in an attractive and acceptable way?</td>
</tr>
<tr>
<td>1-6</td>
<td>Maintainability and extendibility</td>
<td>Is the object-oriented concept applied to the development of the user interface, such that the components of teaching materials can be easily extended and maintained in the future?</td>
</tr>
<tr>
<td>1-7</td>
<td>Quality of security mechanism</td>
<td>Does the system protect personal data of the learners and provide access-control functions?</td>
</tr>
<tr>
<td>1-8</td>
<td>Quality of learning guidance and operational support</td>
<td>Does the system provide proper guidance or operational support, thus helping the learners cope with learning barriers or operational problems?</td>
</tr>
</tbody>
</table>
\[ d_{ij} = \frac{\left( \sum_{k=1}^{m} |x_{ik} - x_{jk}| / m \right)}{\zeta} \]

where \( x_{ik} \) and \( x_{jk} \) represents the decision of \( k \)-th criterion by \( i \)-th and \( j \)-th evaluators for \( k = 1, 2, \ldots, m \); \( \zeta \) are the maximum tolerable times for a participant to make inappropriate decisions. For example, assume that the group decision for \( k \)-th criterion \( x_{gk} = B > D > A > E > C \) and \( \zeta = 10 \). If the corresponding decision of evaluator \( P \) is \( x_{pk} = A > E > C > B > D \), we have \( d_{gp} = (|0–3| + |1–4| + |2–0| + |3–1| + |4–2|) / 100 = 0.24 \), and hence \( w_p = w_p - d_{gp} \). If \( w_p \) becomes zero, evaluator \( P \) is removed from the decision group.

### 3.3. Phase 3: rating the educational web site

Assume that evaluation item \( A \) is evaluated according to \( n \) criteria. EWSE will assist evaluator \( P \) to propose a vector \( (m_{p1}, m_{p2}, \ldots, m_{pn}) \), where \( m_{pj} \) is the rating of \( j \)-th criterion for web site \( S \) rated by evaluator \( P \). An evaluation matrix \( M_A(S) \) is used to represent the ratings of the \( n \) criteria of evaluation item \( A \) for web site \( S \), rated by the \( t \) evaluators:
Let the weight of evaluator $P$ be $w_p$, $W = (w_1, w_2, \ldots, w_t)$ is a vector that represents the weights of the $t$ evaluators. By grouping the ratings of the evaluators with their corresponding weights for each criterion, we have the following:

$$G_A(S) = W \ast M_A(S) = (w_1, w_2, \ldots, w_t) \ast \begin{bmatrix} m_{11}, & m_{12}, & \ldots, & m_{1n} \\ m_{21}, & m_{22}, & \ldots, & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{t1}, & m_{t2}, & \ldots, & m_{tn} \end{bmatrix} = (g_1, g_2, \ldots, g_n)$$

where $g_i = w_1 \ast m_{1i} + w_2 \ast m_{2i} + \ldots + w_t \ast m_{ti}$

where $g_i$ represents the group rating for criteria $i$. EWSE then maps each $g_i$ value to a set of predefined membership functions.
For the example given in Fig. 3, there are four fuzzy grades for evaluating web sites: $\mu_E$, $\mu_G$, $\mu_S$ and $\mu_P$ that represent the membership functions for “Excellent”, “Good”, “Satisfactory” and “Poor”, respectively. Let $d$ represent the number of fuzzy grades ($d=4$ in the above example) and $e_{ij}$ stand for the degree of membership for the grouped rating of criterion $i$ being mapped to fuzzy grade $j$, i.e., $\mu_{E}(g_i)$, a matrix is constructed as follows:

$$EA(S) = \begin{bmatrix} e_{11}, & e_{12}, & \ldots, & e_{1d} \\ e_{21}, & e_{22}, & \ldots, & e_{2d} \\ \ldots & \ldots & \ldots & \ldots \\ e_{n1}, & e_{n2}, & \ldots, & e_{nd} \end{bmatrix} = [e_{ij}].$$

for $i=1, 2, \ldots, n$ and $j=1, 2, \ldots, d$.

An example of $E_A(S)$ is given as follows:

<table>
<thead>
<tr>
<th>Criterion no.</th>
<th>Faithfulness of applying the criterion</th>
<th>Suitableness of applying the criterion to evaluate the web sites</th>
<th>Reasonableness of the definition of the criterion</th>
<th>Applicable range of the criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–1</td>
<td>Faithful</td>
<td>Suitable</td>
<td>Very reasonable</td>
<td>Applicable</td>
</tr>
<tr>
<td>1–2</td>
<td>Average</td>
<td>Unsuitable</td>
<td>Average</td>
<td>Inapplicable</td>
</tr>
<tr>
<td>1–3</td>
<td>Faithful</td>
<td>Very suitable</td>
<td>Very reasonable</td>
<td>Applicable</td>
</tr>
<tr>
<td>1–4</td>
<td>Faithful</td>
<td>Suitable</td>
<td>Very reasonable</td>
<td>Average</td>
</tr>
<tr>
<td>1–5</td>
<td>Very faithful</td>
<td>Suitable</td>
<td>Reasonable</td>
<td>Average</td>
</tr>
<tr>
<td>1–6</td>
<td>Very faithful</td>
<td>Suitable</td>
<td>Very reasonable</td>
<td>Very applicable</td>
</tr>
<tr>
<td>1–7</td>
<td>Unfaithful</td>
<td>Unsuitable</td>
<td>Average</td>
<td>Average</td>
</tr>
<tr>
<td>1–8</td>
<td>Average</td>
<td>Suitable</td>
<td>Average</td>
<td>Average</td>
</tr>
</tbody>
</table>

$$\mu_E(x) = \begin{cases} (x - 85)/5 & 0 \leq x \leq 85 \\ 1 & 85 < x \leq 90 \end{cases}$$

$$\mu_G(x) = \begin{cases} (90 - x)/5 & 90 < x \leq 90 \\ 1 & 90 < x \leq 100 \end{cases}$$

$$\mu_S(x) = \begin{cases} (x - 60)/5 & 0 \leq x \leq 60 \\ 1 & 60 < x \leq 75 \end{cases}$$

$$\mu_P(x) = \begin{cases} (65 - x)/5 & 60 < x \leq 65 \\ 0 & 65 < x \leq 80 \end{cases}$$
A vector \( \mathbf{R} = (r_1, r_2, \ldots, r_n) \) represents the ratio of importance of each criterion, where \( \sum_{i=1}^{n} r_i = 1 \). A vector \( Y_A(S) \) is derived by composing \( \mathbf{R} \) and \( E_A(S) \) to find the membership degree of each grade:

\[
Y_A(S) = \mathbf{R} \circ E_A(S) = (r_1, r_2, \ldots, r_n) \circ \begin{bmatrix}
  e_{11} & e_{12} & \cdots & e_{1d} \\
  e_{21} & e_{22} & \cdots & e_{2d} \\
  \vdots & \vdots & \ddots & \vdots \\
  e_{n1} & e_{n1} & \cdots & e_{nd}
\end{bmatrix} = (y_1, y_2, \ldots, y_d)
\]

where \( y_j = (w_1 e_{1j}) \oplus \cdots \oplus (w_n e_{nj}) \).

\[
A \cdot B = \{(x, \mu_c(x)) | \mu_C(x) = \mu_A(x) \cdot \mu_B(x) \} \quad \forall x \in U
\]

and

\[
A \oplus B = C = \{(x, \mu_c(x)) | \mu_C(x) = \min\{1, \mu_A(x) + \mu_B(x, \mu_c(x)) \} \} \mu_C(x) = \mu_A(x) \cdot \mu_B(x) \} \quad \forall x \in U
\]
By assuming that \( R = (0.1, 0.2, 0.3, 0.2, 0.2) \), we have the following:

\[
Y_A(S) = (0.1, 0.2, 0.3, 0.2, 0.2) \otimes \begin{bmatrix}
0 & 0 & 0.4 & 0.6 \\
0 & 0 & 0.2 & 0.8 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0.4 & 0.6 & 0
\end{bmatrix}
= (0, 0.08, 0.7, 0.22)
\]

Based on values of \( Y_A(S) \), we conclude that by considering evaluation item \( A \) for site \( S \), the degree of being “Poor” is 0, the degree of being “Satisfactory” is 0.08, the degree of being “Good” is 0.7 and the degree of being “Excellent” is 0.22. Accordingly, (“Good”, 0.7) is taken as the evaluation result of evaluation item \( A \).

3.4. Phase 4: integrate the grades of all evaluation items and present final decision to the users

By applying Phase 1, 2 and 3 repeatedly, EWSE finds the (grade, degree) for each evaluation item. Assume that \( Y_B(S) = (0, 0.68, 0, 0.32) \) and \( Y_C(S) = (0, 0.04, 0.92, 0.04) \), the grades and degrees of all evaluation items are shown in Table 6. We have the integrated result as follows:

\[
\frac{(0.7 \times \text{DF(Good)} + 0.68 \times \text{DF(Satisfactory)} + 0.92 \times \text{DF(Good)})}{(0.7 + 0.68 + 0.92)} = \frac{(0.7 \times 82.5 + 0.68 \times 70 + 0.92 \times 82.5)}{2.32} = 78.8.
\]

4. Implementation and experiments

EWSE consists of a representation unit (Java Server Pages), a resource unit (MS-SQL Server) and a business and logic unit (JavaBeans component). In the business and logic unit, the techniques of GRA computation, fuzzy computation, fuzzy reasoning, defuzzification computation, and connection pooling for database are employed.

A Unified Modeling Language (UML) (Fowler & Scott, 1997) tool is used to develop EWSE system. A description of UML sequence diagram is given in Fig. 4 to illustrate the operational flow chart of EWSE.

4.1. Illustrative examples of EWSE

An illustrative example of a GRA interface is depicted in Fig. 5. Each evaluator is asked to rate the degree of suitableness for each criterion with respect to the candidate web site.

<table>
<thead>
<tr>
<th>Evaluation Item</th>
<th>Degree</th>
<th>Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: The design of user interface</td>
<td>0.7</td>
<td>Good</td>
</tr>
<tr>
<td>B: The quality of instructional contents</td>
<td>0.68</td>
<td>Satisfactory</td>
</tr>
<tr>
<td>C: The assessment functions</td>
<td>0.92</td>
<td>Good</td>
</tr>
</tbody>
</table>
Before starting the evaluation process, EWSE must collect the criteria of the evaluators, via the GRA interface. An illustrative example of the collected preferences of the evaluators (domain experts) is shown in Fig. 6.

After finishing all GRA works determining the criteria set for four web sites, evaluators can begin to evaluate them by the EWSE system, as shown in Fig. 7. The evaluation results are depicted in Fig. 8.

4.2. Experiments on educational web sites

Four well-known educational web sites have been selected as the targets of the experiment (Table 7). Two of the target web sites, 000001 and 000002, have been evaluated by e-Oscar and hence, the evaluation results can be compared with those of e-Oscar.

The evaluator group consists of undergraduate and graduate students of Department of Educational Technology—Tamkang University (Taiwan), officials from Department of Social Education—Ministry of Education (Taiwan), and the Internet users. The participants were asked to participate in the stage of determining the criteria set for each dimension and the stage of scoring and grading the target web sites.

In the first stage, 11 participants with educational background (i.e. the students and the officials of the Education Department) were asked to determine a set of criteria for each target web site, via the assistance of the GRA interface. As each educational web site has its main features or functions such as user interface (UI), teaching materials and examinations, EWSE evaluates each
Fig. 5. GRA interface of EWSE.

Fig. 6. Illustrative example of the preferences of domain experts.
Fig. 7. Evaluation interface of EWSE.

Fig. 8. Evaluation results of target web sites.
<table>
<thead>
<tr>
<th>Web site no.</th>
<th>URL</th>
<th>Name of the web site</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>000001</td>
<td><a href="http://www.literature.idv.tw/">http://www.literature.idv.tw/</a></td>
<td>The traditional Chinese literatures</td>
<td>This site has literary and artistic characteristics, traditional Chinese literature, poetry and songs of the Chinese literati.</td>
</tr>
<tr>
<td>000002</td>
<td><a href="http://www.gvm.com.tw/">http://www.gvm.com.tw/</a></td>
<td>The WWW of global views monthly</td>
<td>This site contains many articles concerning Technology, Management and Humane, written by domain experts.</td>
</tr>
<tr>
<td>000003</td>
<td><a href="http://cls.admin.yzu.edu.tw/hlm/">http://cls.admin.yzu.edu.tw/hlm/</a></td>
<td>The educational and research WWW</td>
<td>A web site of Chinese literature, Hong Lou Meng, also containing tutoring and research materials of the topic.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of Dream of the Red Chamber</td>
<td></td>
</tr>
<tr>
<td>000004</td>
<td><a href="http://www.phy.ntnu.edu.tw/demolab/">http://www.phy.ntnu.edu.tw/demolab/</a></td>
<td>The educational WWW of Department</td>
<td>This is a computer-assisted instruction web site for teaching “physics” course.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of Physics of National Taiwan Normal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td></td>
</tr>
</tbody>
</table>
site according to the adopted evaluation dimensions. The final evaluation dimensions and the selected criteria for each dimension are given in Table 8.

As each domain expert (evaluator) may have different degrees of majority on different evaluation dimensions or criteria, it is necessary to assess the evaluator’s majority degree for each target website, based on the evaluation dimensions and the corresponding criteria. Initially, each expert was assigned a majority degree “1.0.” After several iterations of employing the GRA interface for evaluating the educational websites, the majority degrees were updated as follows:

\[
\text{MD (Evaluator, Majority degree)} = \left\{ (E_1, 0.925), (E_2, 0.884), (E_3, 0.915), (E_4, 0.735), (E_5, 0.811), (E_6, 0.905), (E_7, 0.871), (E_8, 0.895), (E_9, 0.908), (E_{10}, 0.831), (E_{11}, 0.854) \right\}
\]

The mean of majority degree (\( \mu \)) was 0.8667 and the standard deviation of the values (\( \sigma \)) was 0.054. Hence, there were four evaluators with majority degrees below the mean, including: (E_4, 0.735), (E_5, 0.811), (E_{10}, 0.831), (E_{11}, 0.854), which implies that those evaluators’ decisions were usually quite different from the final group decisions.

In the second stage, over 30 evaluators (large sampling) participated in the scoring process, which was performed from 27 January 2002 to 28 March 2002. Table 9 presents the evaluation results for the target websites.

By comparing the evaluation results with those of e-Oscar, it can be seen that EWSE offers much more detailed information for users in selecting appropriate educational websites. For website 000001, the rating given by e-Oscar was “four stars”, which was close to “good” in EWSE. For website 000002, e-Oscar’s rating was “three stars” (which was equivalent to “satisfactory” in EWSE), while EWSE’s rating for website 000002 was “good”. It can be seen that some evaluators of e-Oscar might employ improper evaluation dimensions or criteria to evaluate website 000002. Furthermore, the decision made by e-Oscar was completely afforded by the number of the Internet users that voted 000002, that is, professional opinions were not taken into consideration.

<table>
<thead>
<tr>
<th>Site no.</th>
<th>Dimension’s description</th>
<th>Selected criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>000001</td>
<td>1. Design of user interface</td>
<td>1-1, 1-2, 1-3, 1-4, 1-5</td>
</tr>
<tr>
<td></td>
<td>2. Quality of instructional contents</td>
<td>2-2, 2-3, 2-4, 2-6, 2-10</td>
</tr>
<tr>
<td>000002</td>
<td>1. Design of user interface</td>
<td>1-1, 1-2, 1-3, 1-4, 1-5</td>
</tr>
<tr>
<td></td>
<td>2. Quality of instructional contents</td>
<td>2-1, 2-2, 2-3, 2-4, 2-6</td>
</tr>
<tr>
<td>000003</td>
<td>1. Design of user interface</td>
<td>1-1, 1-2, 1-3, 1-4, 1-5</td>
</tr>
<tr>
<td></td>
<td>2. Quality of instructional contents</td>
<td>2-2, 2-3, 2-4, 2-5, 2-10</td>
</tr>
<tr>
<td></td>
<td>3. Assessment functions</td>
<td>3-2, 3-3, 3-4, 3-6, 3-7</td>
</tr>
<tr>
<td>000004</td>
<td>1. Design of user interface</td>
<td>1-1, 1-2, 1-3, 1-4, 1-5</td>
</tr>
<tr>
<td></td>
<td>2. Quality of instructional contents</td>
<td>2-2, 2-3, 2-4, 2-5, 2-10</td>
</tr>
<tr>
<td></td>
<td>3. Assessment functions</td>
<td>3-2, 3-3, 3-4, 3-6, 3-7</td>
</tr>
</tbody>
</table>
From the experiment results, we also found that three web sites were rated “Good” and one was rated “Excellent”, and no web site was rated “Satisfactory” or “Poor”, which is reasonable since these well-known web sites have been evaluated based on the features and functions they offer. For a user, it is more important to find an educational web site with appropriate and superior functions (or features), instead of receiving an overall rating. In evaluating a web site, the ratings given by those evaluators from different fields may significantly differ than those from other fields. For example, to evaluate an educational web site concerning “computer science” courses, the ratings given by a group of people with “computer science” backgrounds and those given by the people with “education” backgrounds may be quite different. Therefore, it would be better to show the users how and from whom the ratings were derived.

5. Conclusion

In this paper, a group-decision approach is proposed for evaluating educational web sites. Several soft computing technologies have been employed in our approach, including fuzzy theory, grey system and group decision method. The major contribution of the research is to help the students and teachers who are looking for suitable educational resources to efficiently access the desired information, based on multiple appropriate dimensions.

For group decision support issue, some problems in traditional group decision processes have been solved. The use of expert weights has balanced the decisions made by experienced evaluators, inexperienced evaluators, and evaluators with different backgrounds or credibility.

Presently, the Ministry of Education and the National Science Council in Taiwan have collected hundreds of URLs of educational web sites. The purpose of collecting and evaluating those educational web sites is to recommend high quality educational resources to students and teachers. Currently, the recommended lists can be found in the Ministry of Education web
Moreover, an educational resource registration website is going to be established. For those who develop educational websites or relevant resources, the registration website will offer an entry for registering their URLs and receiving objective evaluation for their contents. A committee supported by the Ministry of Education and the National Science Council in Taiwan will administer the evaluation process.

References


