Automatic Evaluation for Engineering Pathway Premier Award Winners

Wei Yu[1], Yunlu Zhang[2,] and Lin Gan[1]*

[1] Computer School, Wuhan University, 430072 Wuhan, China
[2] Snopsys Inc.(Wuhan), Wuhan, 430072 Wuhan, China

Abstract: Educating the K-Gray engineering community in today's digital world requires straightforward yet flexible access to high-quality educational resources. Inspires by this, we propose an automatic evaluation system for learning resources ranking in a real world digital library, Engineering Pathway (EP). The Engineering Pathway is a portal to high-quality teaching and learning resources in engineering, applied science and math, computer science/information technology, and engineering technology, which is designed for use by K-12 and university educators and students. We model the best and most popular leaning resource objects from Premier Award Winners to recognize high-quality and non-commercial courseware designed to enhance the engineering education. We adopt the D-S evidence theory to model our problem. After giving effective definition of the mass function, the model can be transferred into multinomial regression model. We try three different models: linear regression, quadratic regression and sextic regression to get the most practicable model. With the help of this model, it will be much more simple and precise to help our domain experts to select the most valuable learning resources in our EP digital library. Experiments show that out proposed model performs well through training and optimization.

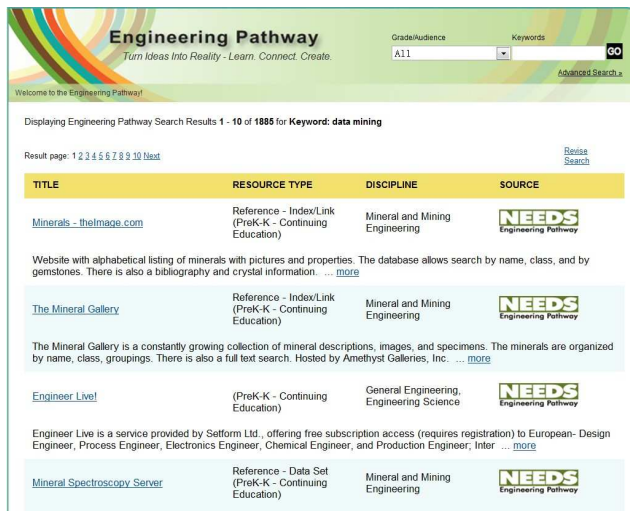**Keywords:** D-S, User behavior,learning resource,premier award

1 Introduction

Educating the K-Gray engineering community in today's digital world requires straightforward yet flexible access to high-quality educational resources. The goal of this project is to create and steward the K-Gray Engineering Pathway (EP) [1] a premier portal to comprehensive engineering and computing education resources within the greater National Science Digital Library (NSDL), by combining NEEDS (National Engineering Education Digital-library System) expertise in higher education and lifelong learning with TE's (TeachEngineering) expertise and experience in K-12 engineering education, as shown in Figure 1. The Engineering Pathway has gone well beyond its initial goals in the growth of resources since its inception in 2005. About 10% are K-12 resources and 90% are for higher education audiences.We already have more than 9,000 registered users up to now, so it is very important to research the effective way to use these users' behavior information, and it will be really helpful to improve our EP designing and to guide our users.

Fig. 1: Screen image of the Engineering Pathway

For now, users can get search results by inputting query keywords in our EP homepage, as shown in Figure 2, we input "data mining" as search keyword, and we can get 26 learning resource objects. From the results page,

* Corresponding author e-mail: Yunlu.Zhang@synopsys.com

Fig. 2: results of Search " data mining "

Fig. 3: Summary page about the learning resource

we can get a summary of each item, such as: title, resource, discipline, and the source where they come from. If we click the link under the first learning resource title, that is, *"Automated Software Engineering Research Group"*, we get more details about it, as shown in Figure 3. We can see *"Comments (0) and Reviews (0)"* on the right of this page, after user logging in, he/she can add comment about it. On the left navigation menu of our homepage, there is one part for "highlighted resources" as shown in Figure 4. The number of comments and reviews have increased dramatically during the award period, starting with 1,000 on October 1, 2005, reaching 12,259 on October 1, 2011.At the end of 2011, the cumulative number of downloads for our catalog records are approximately 860, 00. As connecting users to educational resources is EP's primary goal, we are interested in the views of the catalog records for these resources as well as how many of these views lead to downloading the resource. A" download"is defined as linking to the actual resource or downloading a document, file or executable.

The "100 Most Commented" is what we used in this paper. By clicking the link, you can get the top 100 most commented learning resource objects by usage statistics on an individual discipline or all of the disciplines in our entire site, as shown in Figure 5. We can get the most commented one on the entire site with title *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* by National Academies Press, and with 231 comments. The contents of comment should contain rating, comment title, comment description, author and post time, as shown in Figure 6. How to use these dataset to answer the questions like *"Which learning resource object should we choose"* or *"Which one is the best search result and which one is exactly what I want"* ? That is the reason

why we focus on using these comments to evaluate the learning resources, and we can achieve this by selecting the best alternative that matches all of the digital library user's criteria.

EP offers annual awards for *Premier Award for Excellence in Engineering Education Courseware* [2], announced at the Frontiers In Education Conference, which is an instrumental tool for our users to select the best quality learning resource objects. The Premier Award competition is open to a wide range of submissions of high-quality, engaging, non-commercial learning innovations designed to enhance engineering education.

During the process of selecting the winners of the Premier Award, the judges start off by reviewing the criteria for excellence and adjusting these criteria based on new engineering education research as well as new innovations in the enabling technologies. Updated criteria for excellence are published on the EP site as part of the Premier Award pages [3]. Evaluation studies show that our strength is a consistent interface with strong usability features. We are valued for our quality content that is well-defined in terms of engineering and computing education. Our users appreciate "precision"in a search over "recall", hence our emphasis on evaluation criteria and review processes. That's the reason why we focused on developing evaluation criteria, pruning existing resources and adding quality content.

All of these works are outstanding with the help of domain experts; however from another point of view, also are intensive. So in this paper, we focus on how to implement an automation evaluation system, and the Premier Award could use it as a reference. We have focused our impact analyzes on the Premier Award courseware and our most downloaded resources.

Based on our previous collaborative research [4] [5] [6], we can learn lots of lessons by mining the

Fig. 4: Highlighted Resources

Fig. 5: Top 100 Most commented Resources

Fig. 6: Comment Page of the most commented item

comprehensive information from EP, and to date, EP has 26,421 catalog records as December 31, 2012. Collectively, all of the EP collections average over 1 million page views a month. Over 60% of our records have at least one comment or review. So in this phase, we focus on how to effectively use users' behavior to help us improve our educational digital library services. One of these implement methods is to find to most helpful and popular learning resource. According to experimental results, we discover that we can use the most downloaded as a measure of quality, and at the same time consider the properties of comments of these learning resources. The most downloads are the 100 Most Popular on our homepage left menu, as shown in Figure 4, and the contents of learning resources' comments as shown in Figure 6. There are lots of uncertain properties and

missing values in the *downloads* and *comments*, so we choose D-S evidences theory as basic tool [7]. The reasons we choose this method will be given in the following sections.

2 Related work

D-S evidence theory is a method broadly applied in data fusion or information fusion for decision-making, and it can also used for evaluation, building trust model and so on so forth.   As for data fusion, Khazaee et al. [8] proposed a representative data fusion approach which

exploits Support Vector Machine (SVM) and Artificial Neural Network (ANN) classifiers and Dempster-Shafer (D-S) evidence theory for classifier fusion, it was used to recognize crucial faults of mechanical systems with high confidence, indubitably decision level fusion. by using the D-S theory rules for classifier fusion, the accuracy was 14.5% and 20.2% higher than SVM and ANN respectively. Du et al. [9] also use D-S theory to fuse and classify satellite remote sensing image for monitoring, it was adopted to combine the outputs of three member classifiers to generate the final classification map with higher accuracy than that by any individual classifier. Zhang et.al [10] modified the fusion method of evidence information after considering context's reliability, time-efficiency, and relativity, to improve the classical fusion rule and to ensure the QoS of Web-based mobile application. There are many researches in this area, some of them can be seen in [11] and [12]. In our paper, we use a combination of D-S theory and multiple linear regression to deal with different application scenarios.

As for trust model, Merigo et al. [13] obtained various belief structures (BS), by using aggregation operators within a D-S framework. They assessed the available information with interval numbers such as the uncertain weighted average (UWA), the uncertain ordered weighted average (UOWA), the uncertain generalized weighted average (UGWA) and the uncertain generalized ordered weighted average (UGOWA). Jiang et al. [14] proposed a trust model for ensuring the security of interactions in open distributed systems, in this situation the entity's trust evaluation depends on both interaction experience of its own and recommendation information from other entities, so the Jiang et al. improved D-S evidence theory by introducing a time efficiency factor calculation function, multi-layer evidences reasoning and an improved fusion approach for conflict evidence.

As for evaluation, Yanget al. [15] modified D-S to aggregate the different evaluation information by considering multiple experts' evaluation opinions, failure modes and three risk factors respectively, and used it in analysis of aircraft turbine rotor blades. As for the problems of subjective evaluation and uncertain knowledge, Xiao et al. [16] proposed the concept of D-S generalized fuzzy soft sets by combining D-S theory of evidence and generalized fuzzy soft, and then applied it into a medical diagnosis problem.

There are many researches on how to improve the evaluation system by D-S theory, such as [17] and [18], the former one focus on examining proposals for decision making with D-S belief functions from the perspectives of requirements for rational decision under ignorance and sequential consistency, the latter one extended of D-S evidence theory to get probabilities of antecedents and conclusion of probability decision rules for synthesis evaluation systems.

D-S theory also could be applied in some particular areas, such as Kisku et al. [19] used it to face recognition, Pichon et al. [20] used it to reinterpret the relevance and truthfulness in information correction and fusion, and Dong et.al [21] modified it to automatically combine multiple matchers and to solve high conflicts among different matchers for deep web interface designing.

As for AHP method [22],it can be used in many applications besides our web data mining area, such as in ecosystem [23], emergency management [24], plant control [25], and sustainable groundwater resources [26] and so on so forth. From those research works, it is clear that AHP has been widely applied in many of the data management and decision making areas.Some researchers choose a combination of fuzzy theory and AHP, such as Tao et al. [27] , Uzoka et al. [28] and Feng et al. [29]. Tao proposed a decision model by the application of AFS (axiomatic fuzzy set) theory and AHP method to get the ranking order. Besides considering the preferences from decision-makers which can make the decision results more reasonable, they also provided the definitely semantic interpretations for the decision results by their theory. Uzoka used fuzzy logic and AHP for medical decision support systems. It is interpreting idea to introduce the semantic information proposed by Tao, thus we try to add them in our EP comments mining this time. However it causes high computational complexity, we will try this in our future work in this aspect.

3 Modeling Premier Award Winner for EP

3.1 D-S theory in Premier Award Winner

The Dempster-Shafer decision theory is considered a generalized Bayesian theory which is traditional method to deal with statistical problems, and it is a mathematical theory of evidence based on belief functions and plausible reasoning, which is used to combine separate pieces of information (or evidence) to calculate the probability of an event, and the nature of evidence in our EP learning resources objects are shown as following:

–Not reliable evidence : the comment is wrong sometimes and right sometimes
–Uncertain evidence:
The length of comment contents are changing all the time; The type of comment author: Organization, trusted users, or anonymous
–Incomplete evidence: some properties of some comments may be as NULL
–Contradictory evidence :High rating with low content appraisal.

Definition 1. Mass function M $\{m_1, \ldots, m_6\}$.
Give a recognition frame of learning recourses, named Θ, the mass function of Basic Probability Assignment (BPA) is a function that $2^{\Theta} \to [0, 1]$, and it requires:

$$m(\emptyset) = 0 \text{ and } \sum_{PAW \subseteq \Theta} m(PAW) = 1$$

Definition 2. Premier award Winner Evidences, denoted by E: $\{E_1,\ldots,E_6\}$.
There are six independent evidences in our situation: E_1_D (Downloads), E_2_CR (Comment rating), E_3_CT (Comment Title), E_4_CC (Comment Content), E_5_CAT (Comment Author Type) and E_6_CPT (Comment Posting Time).

So in this paper, Dempster's combinational rule should be shown as following.

Definition 3. Premier Award Winner combinational function Rule $\forall(PAW)\subseteq\Theta$, based on the six probability mass function on Θ:m_1,m_2,m_3,m_4,m_5,m_6, we can get the combinational function of these functions and evidences as:

$$m_1\bigoplus m_2\bigoplus m_3\bigoplus m_4\bigoplus m_5\bigoplus m_6 = \frac{1}{K}\sum_\nabla \varpi$$

And K is the normalizing factor, the value of it should be:

$$\text{K} =\sum_{\nabla\neq\emptyset}\varpi= 1\text{-}\sum_{\nabla=\emptyset}\varpi$$

where $\nabla = D\bigcap CR\bigcap CT\bigcap CC\bigcap CA\bigcap CPT$
and
$\varpi = m_1(D)m_2(CR)m_3(CT)m_4(CC)m_5(CAT)m_6(CPT)$

Definition 4. Premier Award Winner Belief Confidence, which accounts all evidences E_k that support the given proposition "PAW", and it's the lower bound of the confidence interval.

$$Belif_i(PAW) = \sum_{E_k\subseteq PAW} m_i(E_k)$$

Definition 5. Premier Award Winner Plausibility Confidence, which accounts all the observations that do not rule out the given proposition. It's the upper bound of the confidence interval.

$$Plausibilityi_i(PAW) = 1 - \sum_{E_k\bigcap PAW=\emptyset} m_i(E_k)$$

For there are 26,513 learning resource objects in our EP website, the amount of computation amongst those resources using mass function will be mount to astronomical figures: $2^{(26,513)}$, if we considering each mass value of the dataset's subsets. So it's very important to define our mass functions. We will explain it in the following section.

3.2 Optimizing D-S PAW Model

There are three steps to optimize our model.

–Defining mass function
–Simplifying the combinational rules
–Analyzing the model

For there are six evidences in our model, so we can get six of them with different parameters, as n_i, α_i, β_i, θ_i,S_i, m_i $i = 1,\cdots,6$. To calculate the comprehensive scores for learning resource x, we need to synthetic compute on the above six mass functions, take m_1 and m_2 as an example. There are two steps to achieve combinational rules simplification.

Algorithm 1 Mass Function Definition m_i

1: Calculating the absolute score for all the learning resources based on property i (or evidence E_i). Take comment'rating as an example, the absolute value of resource j is given as

$$a_j^i = \frac{1}{N_j}\sum_{n=1}^{N_j} r_{j,n} + \delta^i N_j, \quad j\in\Theta,$$

where N_j is the number of comments for resource j, $r_{j,n}$ is the rating value given by comment n and δ^i is a properly selected weighting for N_j.

2: Ranking the learning resources based on their absolute values a_j^i for $j\in\Theta$.

3: Taking the top n_i resources, and saved them in S_i.

4: Normalizing the top n_i resources, as $b_j^i = \frac{a_j^i}{\sum_{k\in S_i} a_k^i}$

5: Getting $m_i(x)$=$\begin{cases} \alpha_i b_x^i + \beta_i & \text{if } x\in S_i, \\ \theta_i & \text{if } x=\Theta, \\ 0 & \text{otherwise} \end{cases}$

1.For

$$\begin{aligned} m(x) \quad &\propto m_1(x)m_2(x) + m_1(x)m_2(\Theta) \\ &+m_1(\Theta)m_2(x) + m_1(\Theta)m_2(\Theta) \\ &\propto (m_1(x)+m_1(\Theta))(m_2(x)+m_2(\Theta)) \end{aligned}$$

where $m_i(x)$=$\begin{cases} \alpha_i b_x^i + \beta_i & \text{if } x\in S_i, \\ 0 & \text{else} \end{cases}$

2.Then can get:

$$\begin{aligned} m(x) &\propto \prod_{i=1}^{6} (m_i(x)+m_i(\Omega)) \\ &\propto \prod_{i=1}^{6} ((\alpha_i b_x^i + \beta_i)I_{x\in S_i} + \theta_i) \end{aligned}$$

and where $I_{x\in S_i} = \begin{cases} 1 & \text{if } x\in S_i, \\ 0 & \text{else} \end{cases}$

From the expression of $m(x)$,we can see that, if the n is big enough as $S_i = \Theta$, this model will be a multi-variable regression model, and the the highest degree of it will be six. Otherwise, it will a piecewise polynomial model based on ranking, which means ranking the learning resources in the first step, then piecewise linear assigning them according to the ranking score, finally, multiplying their six properties values, and the highest degree still is six.

If we hypothesis $\beta_i = 0$, that means we hypothesis mass function $m_i(x)$ is a linear functions instead of an affine function when $x\in S_i$, the piecewise multinomial $m(x)$ is equivalent to a high-dimensional linear model, which is demonstrated that we can adopt the linear regression or SVM to solve our problem in this situation.

There are also two procedures to analyze our model:

1.In specific, when $\beta_i = 0$, we can get:

$$m(x) \propto \prod_{i=1}^{6} \left(\alpha_i b_x^i I_{x \in S_i} + \theta_i \right)$$

$$\propto \prod_{i=1}^{6} \left(\alpha_i \bar{b}_x^i + \theta_i \right)$$

$$\propto \sum_{U \subseteq \{1,\cdots,6\}} \left(\prod_{i \in U} \alpha_i \prod_{i \in U^c} \theta_i \right) \prod_{i \in U} \bar{b}_x^i$$

$$\propto \sum_{U \subseteq \{1,\cdots,6\}} w_U \hat{b}_x^U$$

where

$$\bar{b}_x^i = b_x^i I_{x \in S_i} = \begin{cases} b_x^i & \text{if } x \in S_i, \\ 0 & \text{else} \end{cases}$$

$U^c = \{1,\cdots,6\} \setminus U$, $\quad w_U = \prod_{i \in U} \alpha_i \prod_{i \in U^c} \theta_i$, $\hat{b}_x^U = \prod_{i \in U} \bar{b}_x^i$.

Thus $m(x)$ is an affine functions of $\hat{b}_x^U$, if $\beta_i = 0$. In this case, we need to solve the following least squares problem

$$\min_{w_U} \sum_{x \in \Theta} \left(z_x - \sum_{U \subseteq \{1,\cdots,6\}} w_U \hat{b}_x^U \right)^2$$

We can solve this optimization problem based on the linear regression model (LRM) or support vector machine (SVM) method.

2.When $\beta_i \neq 0$, we obtain the following challenging optimization problem:

$$\min_{\alpha_i, \beta_i, \theta_i} \sum_{x \in \Theta} \left(z_x - \prod_{i=1}^{6} \left((\alpha_i b_x^i + \beta_i) I_{x \in S_i} + \theta_i \right) \right)^2$$

4 PAW automatic evaluation

Analyses of our 100 most downloaded resources shows that approximately 60% are of the resource type "Teaching Resources", with "Simulation", "Case Study", "Tutorial" and "Labs" as the most popular subsets. Another 30% of the resources are cataloged as general reference resources. The last 10% are of resource type Community, which generally refers to websites or blogs. Within the top 20 most popular EP resources, 18 are Premier Award winners, and the top 10 downloaded learning objects are *MecMovies, LAS File Viewer, Engineering Graphics Tutorials and Lecture Presentations, Jeroo, MDSolids: Educational Software for Mechanics of Materials, Web-based Center for Automated Testing (Web-CAT), ARCADE: Interactive Non-linear Structural Analysis and Animation, JFLAP,Biological Information Handling: Essentials for*

Table 1: Premier Award Winner

Record Title	PAW	Downloads	Comments
M	Yes	2,261	11
LFV	No	2,183	1
EGTLP	Yes	2,035	4
J	Yes	1,809	6
M : ESMM	Yes	1,510	5
WCAT	Yes	1,349	10
ARCAD	Yes	1,275	10
JFLAP	Yes	1,080	5
BIHEE	Yes	927	8
SLMTEC	Yes	884	6

Engineers and SMET Learning Modules and Technologies for an Electronics Curriculum, as shown in Table 1, but here we only take their initial for format requirements.

To verify our method, we compare it with our "Premier Award software". The "most commented" and "most downloaded" resources are accessible on the K-12, Higher Ed and disciplinary pages. We ended 2011 with approximately 12,000 commented or reviewed records - about 74% of our records. As an example of the most highly viewed records, Table 4 shows the cumulative number of views, downloads and comments for the top 20 courseware metadata records. Note a "record download" represents the number of times the user went to either the original resource or downloaded the referenced file. All but two of these most downloaded resources won the Premier Award for Excellence in Engineering Education. Based on this, it is the very best rated by a human panel of experts and thus should be the gold standard to compare against. We have found that the Premier Award winners are usually at the top of this list. The Premier Award winners are resources in which a jury of experts has determined to be of the very highest quality. So we chose it as the gold standard against which to test any automated algorithm.

There are two dataset we want to use for optimizing our method, the first one is the Premier Award winners, and the number of it is 29. The second dataset is the top 100 most downloads learning resources.

First, we use linear regression method, that means all the seven variables in this model are of first degree, these seven variables are a constant term, *Downloads number*, *Comment Rating*, *Comment Posting Time*, *Comment Title Score*,*Comment Content Score*,*Comment Author Score*, so the model is:

$$Y = a + a_1 x_1 + a_2 x_2 + \ldots + a_6 x_6$$

and the values of those corresponding coefficients is $b =$ [-0.0873, 0.8996, -0.0396, -0.0199, -0.1082, 0.1927, 0.2829],the confidence intervals for them are: *bint*= [-0.4584 0.2644], [0.6925 0.9635], [-0.2764 0.3060], [-0.3181 0.4102], [-0.1700 -0.0178], [0.0164 0.2126], [0.0350 0.2963]. To validate the effectiveness of this

linear regression model, we can get s = [0.2704, 30.4522, 0, 0.0310], the parameters for s are R^2, F inspection value, threshold f, and the p that related to prominence rate, then we can discuss our model using the following three standards:

–The general rule is that: the bigger R^2 it is, the better the model is.

–To test the significance of linear regression, researchers generally use the following statistics: F_test, T_test and correlation coefficient test, by Matlab, we can get F and f, normally, the bigger F it is, the better the model is, specially, F should be over f, obviously, we have $30.4522 > 0$ here.

–The P that related to prominence rate should be able to the requirement of lesser than $\alpha(0.05)$, if not, that means there are redundant variables in this model, and should be removed from it.From this respect, our model presented here is rational.

By cross-multiply the variables from this linear regression, we can get new variables for the quadratic regression model, such as x_1x_2, x_1x_3, so the total number of it will be 22, the first parameter is constant, then others are shown as following array.

$$
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 3 & 4 & 5 & 6 & 7 \\
  &   & 8 & 9 & 10 & 11 & 12 \\
  &   &   & 13 & 14 & 15 & 16 \\
  &   &   &   & 17 & 18 & 19 \\
  &   &   &   &   & 20 & 21 \\
  &   &   &   &   &   & 22
\end{array}
$$

The figures in the first row are the variables from the linear regression: x_1, x_2, x_3, x_4, x_5 and x_6, the "2" in the second row is the x_1, the "3" means $x_1 * x_2$, the "3" means $x_1 * x_3$ and so on so forth. We can also get the b, $bint$ and s for these variables as in the linear regression model, and then can analyze them with the same three standards.By those compute results, this model is even more reasonable then the former one.

At last, we use anohter multinomial regression model to test our method out, the highest degree of its variables is six. Now, we have 64 parameters in this model, and the correspondence of those parameters ($\{y\}$) and the original parameters ($x_1, x_2, x_3, x_4, x_5, x_6$) in the first model should be like:

$$y(i_1 + i_2 * 2 + i_3 * 4 + i_4 * 8 + i_5 * 16 + i_6 * 32 + 1)$$

In here, the variables $\{i_1, \ldots, i_6\}$ mean combined variable of $x_1, x_2, x_3, x_4, x_5, x_6$, so y(1) means all the values of $\{i_1, \ldots, i_6\}$ is 0, only contains the constant; y(2) means $i_1 = 1$, so it only contains x_1, y(3) means $i_2 = 1$, so it only contains x_2; y(4) means $i_1 = i_2 = 1$,so it only contains both x_1 and x_2. We can also get b, $bint$ s as the former two models, and evaluate it with the same three standards.

In the experiment, we get top 600 learning resources as date set, the first 200 is for training data, the left
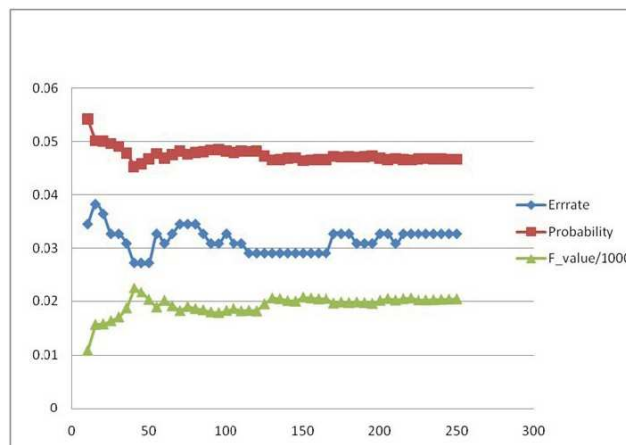

Fig. 7: liner regression model

learning resource is for testing data. During the test phase, we need to adjust the parameter N, which means select top n objects and save them into set S, as we said in section 3.2 in the Mass Function Definition algorithm. So can test those models with error rate *errate*, the probability value P and the F value, as said in these three standards mentioned before. As shown in Figure 7, Figure 9 and Figure 11, the red line is for P (Probability), the better the model is ,the more its value is; the blue line is for *errate*, and the green line is for F, it is used for describing the model's significance, the higher it is, the more reasonable the model is. It should be noticed that F is divided by 1000 to integrate all those curves into the same figures. From Figure 7, Figure 9 and Figure 11, we can see our method is reliable, all the F is high above f, and the *errate* are mostly around 3%, the best one even can low to 1.5%. Finally, combing with the comparison of these models mean value and best value, as shown in Figure 13 and Figure 14, we can draw a conclusion that it is best to select the top 40-45 learning resources as Premier Award Winner candidate, and the quadratic regression model will be the most practicable ancillary tool to help this competition. From Figure 8, Figure 10 and Figure 12 are the regression residuals plot for those three models, we can draw a similarity conclusion from these plots.

5 AHP Automatic Evaluation

We already have done learning object Evaluation by AHP [22] in our prior work [6]. By using AHP method to verify our PAW automatic evaluation , we choose top 100 most commented resources as testing dateset. The mainly procedure contains three steps: building the Hierarchy, ranking criteria/preferences matrices and making judgments and comparisons. There are three lays on our system by using AHP, that is, the objective layer, criteria
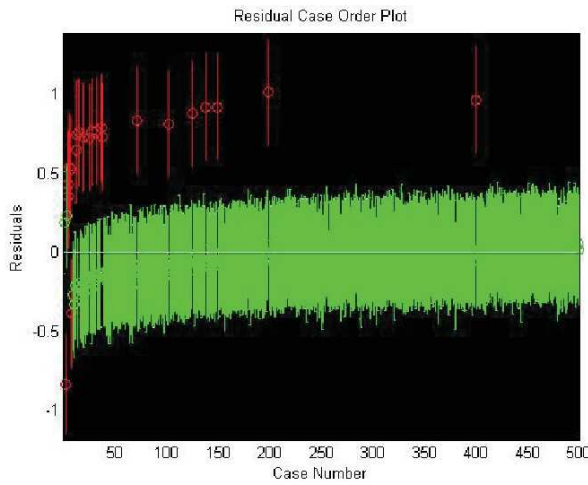
Fig. 8: liner regression residuals plot
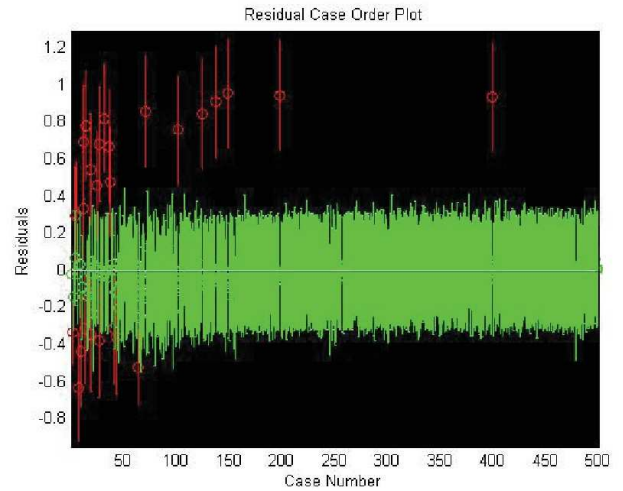
Fig. 10: Quadratic regression residuals plot
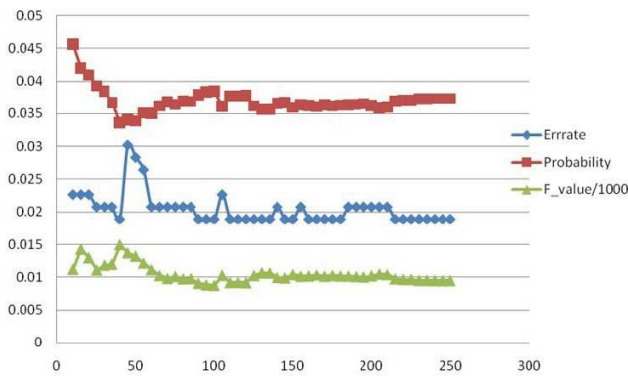
Fig. 9: Quadratic regression model
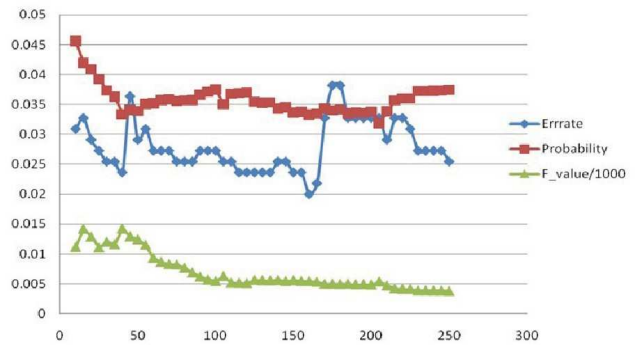
Fig. 11: Sextic regression model

layer, and alternatives layer. The framework of the AHP hierarchy and the content of each layer is as shown in Figure 15.

–Objective layer: Select the most valuable learning resource objects;
–Criteria layer: downloads number,comment rating, comment title, comment description/ comment content, comment author type (person, organization, or anonymous), posting time;
–Alternatives layer: the top 100 most commented resources in computer science discipline.

5.1 Judgement matrices

There are one judgement matrix in criteria layer and six judgement matrices in alternatives layer to be determined.

Fig. 12: Sextic regression residuals plot

Fig. 13: Comparison of three regression models' mean values

Fig. 14: Comparison of three regression models' best value

Fig. 15: AHP hierarchy for Automatic Evaluation

We will establish them based on the expert information and quantitative information respectively.

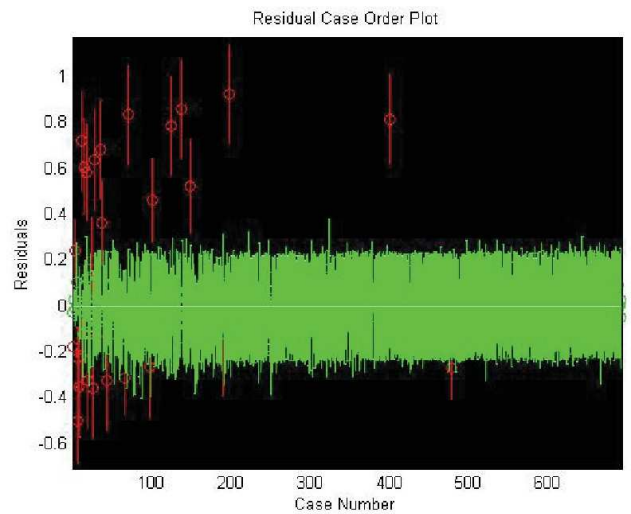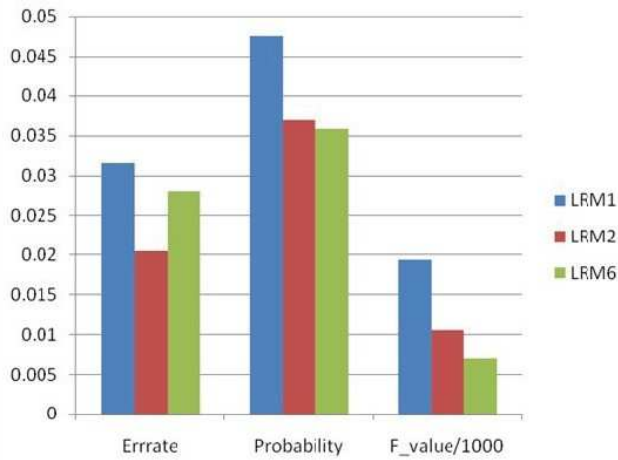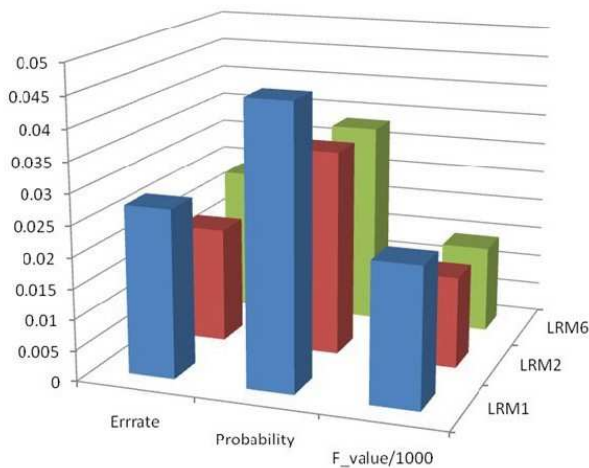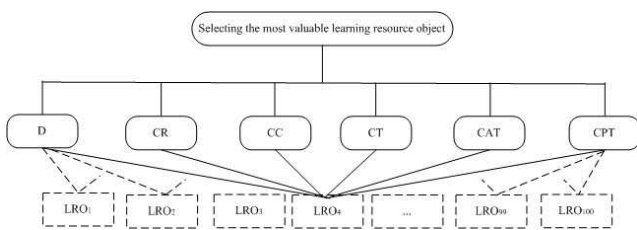*Judgement matrices in criteria layer*. Let matrix $M \in R^{6 \times 6}$ denote the judgement matrix in criteria layer and C_i, $i = 1, \ldots, 6$, denote the five criteria (factors) in the criteria layer respectively. To establish matrix M, a questionnaire is used to make the pairwise comparison on importance between these five factors. In matrix M, its element $M_{i,j}$ means the quantitative relative importance judgement of pairwise factors C_i and C_j. The standard relative importance scale is employed,

Judgement matrices in alternative layer. Let matrices $M^i \in R^{100 \times 100}$, $i = 1, \ldots, 6$, denote the judgement matrix for criteria C_i in the alternative layer and A_i, $i = 1, \ldots, 100$, denote the top 100 most commented resources respectively. To establish each matrix M^i, we use the following two steps:

–Calculate the Absolute Importance (AI^i)
   Based on the related information, such as the rating value and number of comments for rating C^1, calculate the absolute importance value for each alternative under criterion C_i;
–Scale and calculate the Relative Importance (RI^i)
   Scale the absolute importance value to 1 to 9, and calculate the relative importance to get the matrix M^i;

So, we can determine the judgement matrices for each attributes in the alternative layer, taking Rating matrix as an examples.

Rating matrix M^1. For the criterion of rating, the average rating value, varying from 1 to 6, and the number of comments are used to determine the absolute importance of the rating for each resource $j = 1, \ldots, 100$ as follows:

$$AI_j^1 = \frac{1}{N_j} \sum_{n=1}^{N_j} r_{j,n} + \alpha^1 N_j$$

where N_j is the number of comments for resource j, $r_{j,n}$ is the rating value given by comment n and α^1 is a properly selected weighting for N_j. By this method, we can also get Comment title matrix M^2, Comment description matrix M^3, Author type matrix M^4, Posting time matrix M^5 and Download number matrix M^6.

$$AI_j^i = \frac{1}{N_j} \sum_{n=1}^{N_j} LT_{j,n} + \alpha^i N_j (i = 1, ..., 6)$$

Suppose AI_{max}^i and AI_{min}^i are the maximum and minimum absolute importance of the top 100 most commented resources respectively, thus the relative importance value of resource j is given by the following linear scaling method:

$$RI_j^i = \frac{8}{AI_{max}^i - AI_{min}^i} (AI_j^i - AI_{min}^i) + 1$$

We assume that $AI_{max}^i > AI_{min}^i$. The element in matrix M^i can be given by

$$M_{k,j}^i = \frac{RI_k^i}{RI_j^i}$$

Note that we will use the same linear scaling method and definition of RI_j^i for the following matrices, $i = 2, \ldots, 6$, and omit them.

5.2 Weighting

After getting the judgement matrix M, we obtain the weightings of the five criteria in the second layer by calculating its normalized eigenvector w_{max} corresponding to the maximum eigenvalue λ_{max}, that is,

$$Mw_{max} = \lambda_{max}w_{max}, \quad \sum_{i=1}^{6} w_{max,i} = 1.$$

Similarly, we get weightings of different alternative under criteria C_i, $i = 1, \ldots, 6$, as the normalized eigenvectors w_{max}^i of M^i as follows:

$$M^i w_{max}^i = \lambda_{max}^i w_{max}^i, \quad \sum_{j=1}^{100} w_{max,j}^i = 1.$$

Note that for M^i, $i = 1, \ldots, 6$, they are 100×100 dimension matrices, thus it is time-consuming job to exactly calculate the eigenvalues and eigenvectors. To obtain a high quality estimation in short time, we use the following approach (take M for example):

1. Normalize each column in M, that is, $M'_{i,j} = \frac{M_{i,j}}{\sum_{i=1}^{n} M_{i,j}}$;
2. Sum $M'_{i,j}$ in each row, that is, $M'_i = \sum_{j=1}^{n} M'_{i,j}$;
3. Estimate the desired eigenvector w by normalize M^i, that is, $w = \{w_j\}_{j=1,\ldots,n}$, where $w_j \approx \frac{M'_j}{\sum_{i=1}^{n} M'_i}$;
4. The maximum eigenvalue λ can be estimated by $\frac{1}{n}\sum_{i=1}^{n} \frac{(Mw)_i}{w_i}$.

Note that, a well-defined judgement matrix should be transitive, here we can check its consistence by Saaty Rule[22]. It suggests to calculating the following Consistency Ratio (CR):

$$CR_M = \frac{CI_M}{RI_n}$$

where CI_M is so-called Consistence Index of matrix M, defined as,

$$CI_M = \frac{\lambda_{max} - n}{n - 1}$$

and RI_n is the so-called Random Index for $n \times n$ matrix. **If $CR_M < 0.1$, then matrix M is consistent; otherwise, we need to adjust M until the condition satisfies**. For matrices $M^i, i = 1, \ldots, 6$, we assert that they are always consistent because

$$M_{k,j}^i M_{j,l}^i = \frac{RI_k^i}{RI_j^i} \frac{RI_j^i}{RI_l^i} = M_{k,l}^i$$

thus $\lambda_{max}^i = n$ and $CI_{M^i} = 0$. Furthermore, the total consistency is also holds because

$$\frac{w_{max,1}CI_{M^1} + w_{max,2}CI_{M^2} + \cdots + w_{max,6}CI_{M^6}}{w_{max,1}RI_{100} + w_{max,2}RI_{100} + \cdots + w_{max,6}RI_{100}} = 0 < 0.1.$$

5.3 Experimental Set up

To summarize the implementation of AHP for our problem, we need to:

1. Criteria Evaluation
   Get judgment matrix M based on expert information, calculate its maximum eigenvalue and corresponding eigenvector w_{max}, and check its consistence by Saaty Rule (adjust it if necessary);

2. Alternative Weightings Calculation:
   For each criteria C_i, $i = 1, \ldots, 6$, calculate its judgement matrix, and further calculate its maximum eigenvalue w_{max}^i and corresponding eigenvector;

3. Alternative Evaluation:
   The final evaluation v_j for alternative j is given as

$$v_j = \sum_{i=1}^{6} w_{max,i} w_{max,j}^i.$$

Then the top 10 learning resources will be returned to users. To implement the proposed AHP method, first the following judgement matrix M in criteria layer is obtained based on the expert information from the questionnaire.

$$M = \begin{bmatrix} 1 & 3 & 1 & 4 & 4 & \frac{1}{5} \\ \frac{1}{3} & 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{4}{7} \\ 1 & 2 & 1 & 2 & 2 & \frac{1}{5} \\ \frac{1}{5} & 1 & \frac{1}{6} & 1 & \frac{1}{2} & 6 \\ \frac{1}{4} & 4 & \frac{1}{2} & 1 & 1 & \frac{1}{9} \\ 5 & 7 & 5 & 9 & 9 & 1 \end{bmatrix}$$

The maximum eigenvalue λ_{max} of M is 6.5536, the $CI = \frac{\lambda_{max} - 6}{6 - 1} = 0.1107$ and $CR = \frac{CI}{RI} = 0.0893 < 0.1$, thus it is consistent. The corresponding normalized eigenvector is $w_{max} = (0.1609, 0.04315, 0.1143, 0.0767, 0.0767, 0.5281)$, which is also the weightings of the six criteria.

Next we calculate the judgement matrices in alternative layer. To properly set parameters α^i, $i = 1, \ldots, 6$, we calculate the average number of comments, rating, comment title length, comment length, author value, posting time value and download number $(4.31, 3.51, 38.3, 481.4, 2.51, 0.24, 1951)$. Note that α^i can be viewed as a scale factor for the number of comments, so we set α^i according to the above average value, as shown in Table. 2.

Table 2: Parameters setting for AHP

i	1	2	3	4	5	6
α^i	0.8	8	10	0.5	0.05	30

Fig. 16: Comparison with AR
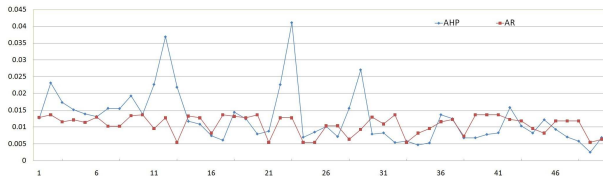

Fig. 18: Comparison with ANCL
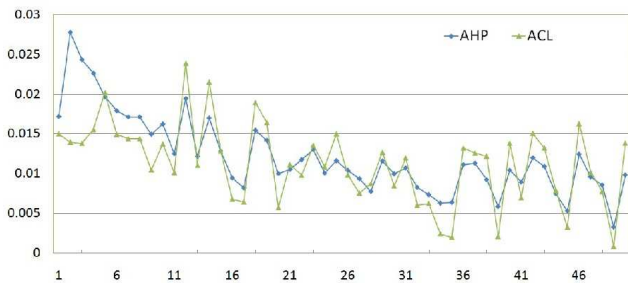

Fig. 17: Comparison with ACL

Fig. 19: Comparison with ATL

Then we can obtain the eigenvectors of $M^i, i = 1, \ldots, 6$, by estimation method, we can further get the final evaluation of the top 100 commented resources. We normalized the evaluation of the 100 resources by average rating (AR), average comments length (ACL), average comment title length(ACTL), average number of comments (ANC), average author value (AAV) and download number (DN). Figure 16 - Figure 21 show the first 50 resource evaluation value by AHP and the above average methods. From Figure 16-Figure 21, we can see that the evaluation results obtained by AHP and ACL own the highest similarity. Furthermore, AHP can be viewed as a mixed method of these average method.

To make it clear, we define the standard deviation between AHP and method $*$ as

$$SD_* = \left(\sum_{j=1}^{100} (v_j^{AHP} - v_j^*) \right)^2$$

where v_j^{AHP}, v_j^* are the normalized evaluation value for resource j obtained by AHP and method $*$. Table. 3 reports the standard deviation for different methods. ACL, AR and DN have the smaller SD values compared with other methods. It is consistent with the expert information and our proposed AHP method from the judgement
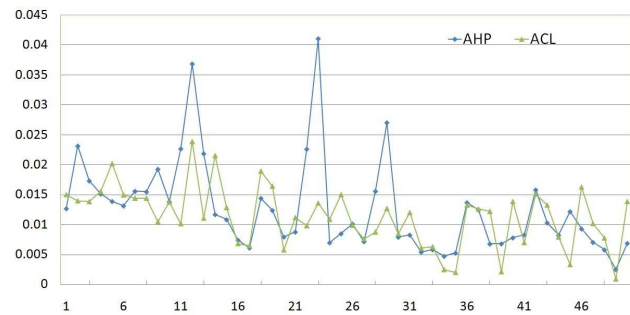

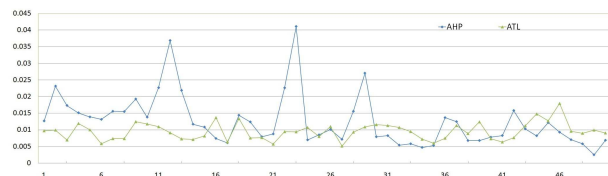**Fig. 20:** Comparison with AAV
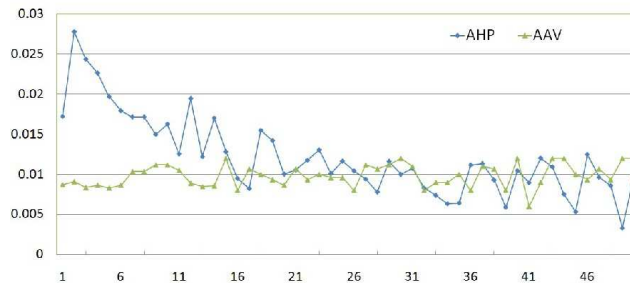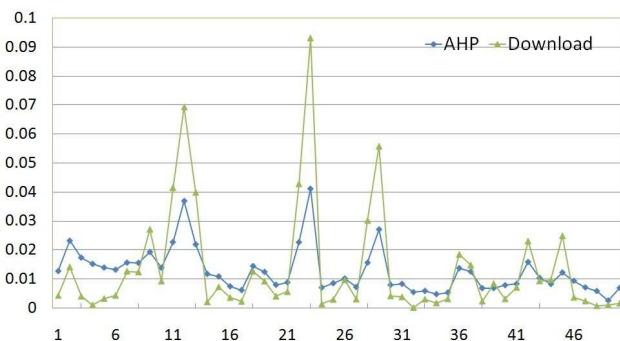

Fig. 21: Comparison with AD

Table 3: The squared deviation of different methods with AHP

	ACL	AR	AAV	ATL	APTV	DN
SD	0.034	0.041	0.050	0.057	0.059	0.006

Table 4: The top 10 recommended resources by different methods

	AHP	ACL	DN
1	Alice 2.0	The Black Swan	Saul Griffith
2	iWoz	ACM K-12	World Without Oil
3	First Barbie	Pair Programming	Darmstadt Dribblers
4	Pair Programming	Educating Engineers	Software Engineering With Java
5	The Black Swan	Computer Museum	Toy Story
6	Computing	ThinkCycle	A Threat in the Air
7	Computational Geometry	Initiatives	World Without Oil
8	Language	Media	Science and Engineering Indicators
9	Achieving Dreams	Scratch	Autonomous Flying Robots
10	ACM K-12	Children Website	Algorithm Animation at Georgia Tech

matrix M, we see that criteria rating and comment length have the bigger weightings, and DN has the biggest one.

6 Conclusion and Future work

In this paper, we proposed an automatic evaluation system for learning resources ranking in a real world digital library, Engineering Pathway (EP). We model the best and most popular leaning resource objects from Premier Award Winner, which is introduced to recognize high-quality, non-commercial courseware designed to enhance engineering education. Then we select top 600 most popular learning resource objects as the data set, first 200 of them are as training data, the rest 400 are as testing date. By using D-S evidence theory to model our problem, after we give the effective mass function definition, this model can be transferred into multinomial regression model. To test the validity of our method, we try three different models: linear regression, quadratic regression and sextic regression, by all of these tests, we can get the most practicable model. With the help of this model, it will be more much simple and precise to help our domain experts to select our most valuable learning resources in our EP digital library.Here, we compare it with baselines(linear regression, quadratic regression and sextic regression), instead of classical F-score, precision and recall. The reason for it is that we do this experiment not to verify the effectiveness of the proposed method, but to aim to choose the most important factor that could help us most in off-line training phase.

Our work is based on the assumption: "We Assume the longer the comment title is, the more important it is". However, there are many factors to effect the importance of the review, such as readability and coverage. There are some former work, such as [30] and [31]. And Another assumption is "We assume the more greater the average

posting time is, the more important the resource is". The background of this assumption is that we have lots of learning objects with high number of downloads or comments, but the reason for them is they are not latest ones, those latest ones tend to have less browsers and then less downloads and comments. So the posting time and the importance may be relevant. Another deficiency in our this phase work is our unconvincing dataset number, for there are truly too few Premier Award Winner each year, we only have 30 of them as referee, but we will confirm its veracity in the following years.

Our work opens up several interesting future directions. First, we can introduce more semantic information analysis on the comments data. Second, we can also conduct a more detailed study on how to accurately classify the comments' author type. And without doubt, we would like to extend our model to other decision making tasks. For example, we can do this analysis for metadata search, which is also an important function in our EP digital library.

Acknowledgement

References

[1] http://www.engineeringpathway.com:8080/engpath/ep/Home.

[2] http://www.engineeringpathway.com:8080/engpath/ep/premier/2012/index.jsp

[3] http://www.engineeringpathway.com:8080/engpath/ep/premier/2012/criteria.jsp

[4] Yunlu Zhang, Alice M. Agogino, Shijun Li, Lessons Learned from Developing and Evaluating a Comprehensive Digital Library for Engineering Education JCDL '12 Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries , 393-394.

[5] Yunlu Zhang, Guofu Zhou, Jingxing Zhang, Ming Xie, Wei Yu and Shijun Li, Engineering Pathway for User Personal Knowledge Recommendation.The 13th International Conference on Web-Age Information Management (WAIM 2012), 459-470.

[6] Yunlu Zhang, What Can We Get from Learning Resource Comments on Engineering Pathway.The 15th International Asia-Pacific Web Conference (APWeb 2013). Accepted.

[7] Dempster, A. P. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics,, **38**, 325-339 (1967).

[8] Khazaee, Meghdad; Ahmadi, Hojat; Omid, Mahmoud. Vibration condition monitoring of planetary gears based on decision level data fusion using Dempster-Shafer theory of evidence, JOURNAL OF VIBROENGINEERING, **14**, 838-851 (2012).

[9] Du Peijun; Yuan Linshan; Xia Junshi,Fusion and classification of Beijing-1 small satellite remote sensing image for land cover monitoring in mining area, CHINESE GEOGRAPHICAL SCIENCE, **21**, 656-665 (2011).

[10] Zhang De-gan, Zhang Xiao-dan, A New Service-Aware Computing Approach for Mobile Application with Uncertainty, APPLIED MATHEMATICS and INFORMATION SCIENCES, **6**, 9-21 (2012).

[11] Wang Peng; Zhang Yuan; Xin Jing-Lei,Research Vision-Guided Robot for Obstacle Avoidance of Information Fusion in the Unstructured Environment, SENSOR LETTERS, **9**, 2021-2024 OCT (2011).

[12] Zhang, Yong; Zhang, Xian-ming, Multi-information fusion diagnosis of lubrication oil contamination using fuzzy distance, ENERGY EDUCATION SCIENCE AND TECHNOLOGY PART A-ENERGY SCIENCE AND RESEARCH, **28**, 95-104 (2011).

[13] Merigo, Jose M.; Casanovas, Montserrat, Decision-making with uncertain aggregation operators using Dempster USING-Shafer Belief Structure, INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING INFORMATION AND CONTROL, **8**, 1037-1061 (2012).

[14] Jiang, Liming; Xu, Jian; Zhang, Kun, A new evidential trust model for open distributed systems, EXPERT SYSTEMS WITH APPLICATIONS, **39**, 3772-3782 (2012).

[15] Yang, Jianping; Huang, Hong-Zhong; He, Li-Ping,Risk evaluation in failure mode and effects analysis of aircraft turbine rotor blades using Dempster-Shafer evidence theory under uncertainty, ENGINEERING FAILURE ANALYSIS, **18**, 2084-2092 (2011).

[16] Xiao, Zhi; Yang, Xianglei; Niu, Qing, A new evaluation method based on D-S generalized fuzzy soft sets and its application in medical diagnosis problem,APPLIED MATHEMATICAL MODELLING, **36**, 4592-4604 (2012).

[17] Giang, Phan H,Decision with Dempster-Shafer belief functions: Decision under ignorance and sequential consistency,INTERNATIONAL JOURNAL OF APPROXIMATE REASONING, **53**, 38-53 (2012).

[18] Pei, Zheng; Zou, Li; Karimi, Hamid Reza,Consistency of Probability Decision Rules and Its Inference in Probability Decision Table, MATHEMATICAL PROBLEMS IN ENGINEERING, Article Number:507857, (2012).

[19] Kisku, Dakshina Ranjan; Gupta, Phalguni; Sing, Jamuna Kanta, Probabilistic approach to face recognition. JOURNAL OF THE CHINESE INSTITUTE OF ENGINEERS, **35**, 529-534 (2012).

[20] Pichon, Frederic; Dubois, Didier; Denoeux, Thierry, Relevance and truthfulness in information correction and fusion, INTERNATIONAL JOURNAL OF APPROXIMATE REASONING, **53**, 159-175 (2012).

[21] Dong Yongquan; Li Qingzhong; Ding Yanhui, ETTA-IM: A deep web query interface matching approach based on evidence theory and task assignment, EXPERT SYSTEMS WITH APPLICATIONS, **38**, 10218-10228 (2011).

[22] T.L. Saaty, Fundamentals of Decision Making and Priority Theory with the AHP, 2nd ed., RWS Publications, Pittsburgh, (2000).

[23] Koschke, Lars; Fuerst, Christine; Frank, Susanne , A multi-criteria approach for an integrated land-cover-based assessment of ecosystem services provision to support landscape planning , ECOLOGICAL INDICATORS, **21**, 54-66 (2012).

[24] Ergu, Daji; Kou, Gang; Peng, Yi, Data Consistency in Emergency Management, INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL, **7**, 450-458 (2012).

[25] Forsyth, G. G.; Le Maitre, D. C.; O'Farrell, P. J , The prioritisation of invasive alien plant control projects using a multi-criteria decision model informed by stakeholder input and spatial data, JOURNAL OF ENVIRONMENTAL MANAGEMENT, **103**, 51-57 (2012).

[26] Adiat, K. A. N.; Nawawi, M. N. M.; Abdullah, K , Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool - A case of predicting potential zones of sustainable groundwater resources, JOURNAL OF HYDROLOGY, **440**, 75, MAY 29 (2012).

[27] Tao, Lili; Chen, Yan; Liu, Xiaodong,An integrated multiple criteria decision making model applying axiomatic fuzzy set theory

[28] Uzoka, Faith-Michael Emeka; Obot, Okure; Barker, Ken. An experimental comparison of fuzzy logic and analytic hierarchy process for medical decision support systems. COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE, **103**, 10-27 (2011).

[29] Feng, Zhigang; Wang, Qi, Research on health evaluation system of liquid-propellant rocket engine ground-testing bed based on fuzzy theory , ACTA ASTRONAUTICA, **61**, 840-853 (2007).

[30] J. Otterbacher "Helpfulness" in Online Communities: A Measure of Message Quality. CHI2009.

[31] Yue Lu et al. Exploiting Social Context for Review Quality Predition. WWW2010."

Wei Yu received the Ph.D degree in Computer School of Wuhan University and is working in Computer School of Wuhan University as instructor. His research interests are in the areas of Web Data Ming and Social Networks. He has published research articles in reputed international journals of computer and engineering sciences. He is member of China Computer Federation .

Yunlu Zhang received the Ph.D degree in School of Computer at Wuhan University and is working in Snopsys Inc.(Wuhan Office).Her research interests are in the areas of Web Data Ming and Social Networks.She has published research articles in reputed international journals of computer and engineering sciences.

Lin Gan is working for her doctorate in School of Computer at Wuhan University .His research interests are in the areas of Data Ming and Time Consistency of Internet. He has published research articles in reputed international journals of computer and engineering sciences. She is member of China Computer Federation.