Recommenders System for Effective ICT Based Learning


[1][2][3][4][5] Department of Computer Science, SRM University, Chennai
TamilNadu, India

[1] akshaya11.cse@gmail.com [2] reddysupriya386@gmail.com

Abstract: Recommended Systems for Information & communication technology (ICT-RS) provide personalized services for recommended system. It provides learning objects for teachers and students. User profiling mechanisms are used for recommended system. This paper proposes ICT-RS which targets to support users in selecting Objects from existing Object Repositories. Automatically constructing their ICT Competence Profiles based on their actions within these ORs. In technology enhanced learning (Tel) major topic is based on user learning profile but not on student learning profile. So in our proposed system teachers and students have equal priority in selecting a Learning objects.

Index Items: Information & communication Technology, Recommended systems, Learning Objects, Technology enhanced learning.

I. INTRODUCTION

Recommender systems are a subclass of information filtering system that seek to predict the ‘rating’ or ‘preference’ that a user would give to an item. Recommender systems have become extremely common in recent years, and are applied in a variety of applications. The most popular ones are probably movies, music, news, books, research articles, search queries, social tags, and products in general[1]. Recommender systems typically produce a list of recommendations in one of two ways—through collaborative or content-based filtering[2].

Collaborative filtering approaches building a model from a user's past behaviour (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in[3]. Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties[4]. A hybrid approach, combining collaborative filtering and content-based filtering could be more effective in some cases. Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model[5].

In the context of Tel, RS is utilized for the recommendation of different types of Learning Objects (LO) based on individual teachers and students profile. The role of teachers and students should be given equal importance. In this paper we propose object based recommendation system which helps the teachers in selecting resources of their own interest by considering the users current status. It needs algorithms namely fuzzy logic and feedback algorithm. Fuzzy logic algorithm allows the system to show the most appropriate content based on his/her profile. Fuzzy logic theory contains the intelligence to provide the suitable presentation of learning objects to the analysed criteria. Some of the inference rules are used for the students to use which learning object[6]. And Feedback algorithm is used to rate the learning object. Feedback is inferred from existing groups, such as noting which one is most suitable[7]. There are two types of rating-implicit and explicit[8]. Explicit feedback system normally prompts the user through the system interface to provide ratings for items in order to construct and improve model. The accuracy of recommendation depends on the quantity of ratings provided by the user. The only shortcoming of this method is, it requires effort from the users and also, users are not always ready to supply enough information[9]. Implicit feedback system automatically infers the user’s preferences by monitoring the different actions of users such as the history of purchases, navigation history, and time spent on some web pages, links followed by the user, content of e-mail and button clicks among others. Implicit feedback reduces the burden on users by inferring their user’s preferences from their behaviour with the system[10].

All Rights Reserved © 2016 IJERCSE 193
II. LITERATURE SURVEY

Recommender systems have developed in parallel with the web. They were initially based on demographic, content-based and collaborative filtering. Currently, these systems are incorporating social information. It has to concentrate on fields, such as on (1) proper combination of existing recommendation methods that use different types of available information, (2) to get the maximum use of the individual potential of various sensors and devices on the Internet of things, (3) acquisition and integration of trends related to the habits, consumption and tastes of individual users in the recommendation process, (4) data mining from RS databases for non-recommendation uses (e.g., market research, general trends, visualization of differential characteristics of demo-graphic groups), (5) enabling security and privacy for recommender systems processes, (6) new evaluation measures and developing a standard for non-standardized evaluation measures, and (7) designing flexible frameworks for automated analysis of heterogeneous data[1].

A metric to measure similarity between users, which is applicable in collaborative filtering processes carried out in recommender systems. The proposed metric is formulated via a simple linear combination of values and weights. Values are calculated for each pair of users between which the similarity is obtained, where weights are only calculated once, making use of a prior stage in which a genetic algorithm extracts weightings from the recommender system which depend on the specific nature of the data from each recommender system. The results obtained present significant improvements in prediction quality, recommendation quality and performance[2].

User profiling is a technique aimed at capturing and exploiting significant characteristics of the users towards the provision of personalized services within adaptive systems such as Recommender Systems (RS). In the context of Technology enhanced Learning (TeL), from a teachers’ perspective, the unique ICT competence characteristics of individuals have not been considered when providing Learning Object (LO) recommendations, despite their vital contribution to the level of ICT uptake of teachers. Moreover, there is a lack of mechanisms for automatically eliciting and updating such personal characteristics within Learning Object Repositories (LOR) in order to exploit them for enhanced LO recommendations. Towards tackling this issue, we propose a teacher ICT Competence elicitation mechanism utilizing fuzzy logic for inferring teachers’ ICT Competences based on their usage patterns within LOR and presents the results of its preliminary accuracy evaluation. The results indicate that the proposed approach provides high accuracy and can, therefore, construct reliable depictions of the teachers’ ICT


db

Fig 1 Architecture diagram of proposed system

Competence Profiles. It has to concentrate firstly on, from a user based perspective, related to user-based evaluation of the method by real teachers towards its further refinement. Secondly, from an added value perspective which will focus on the exploitation of the proposed approach within other educational datasets for providing focused LO recommendations to teachers [7].

Recommender systems have the effect of guiding users in a personalized way to interesting objects in a large space of possible options. Content-based recommendation systems try to recommend items similar to those a given user has liked in the past. The basic process performed by a content-based recommender consists in matching up the attributes of a user profile in which preferences and interests are stored, with the attributes of a content object (item), in order to recommend to the user new interesting items. The nature of this kind of systems, can only recommended items that score highly against a user’s profile, thus the user is limited to being recommended items similar to those already rated.

III. PROPOSED SYSTEM

Recommended Systems for Information & communication technology (ICT-RS) provide personalized services for recommended system. It provides learning objects for teachers and students. User profiling mechanisms are used for recommended system. The new system proposes ICT-RS which targets to support users in selecting Objects from existing Object Repositories, Firstly new user will create a profile which contains his/her name and professional details. User will write a test based on their area of interest and the profile is updated after writing the test. If he/she is an existing user then user has to be logged in. There are two types of users-teachers and students. If user is a teacher they will be grouped according the marks obtained after taking test and information is stored in the teacher database. By using feedback algorithm each and
every learning object is rated by teacher and stored in the LO database. If the user is a student, the Los are recommended based on marks obtained. This is done by using fuzzy logic algorithm and students details are stored in student database which is handled by the administrator.

IV. FUZZY LOGIC

A fuzzy logic system (FLS) is defined as the nonlinear mapping of an input data set to a scalar output data. A FLS consists of four main parts: Fuzzifier, rules, inference engine, and Defuzzifier.

Firstly, a crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as Fuzzification. Afterwards, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step.

a. Algorithm
1. Define the linguistic variables and terms (initialization)
2. Construct the membership functions (initialization)
3. Construct the rule base (initialization)
4. Convert crisp input data to fuzzy values using the membership functions (fuzzification)
5. Evaluate the rules in the rule base (inference)
6. Combine the results of each rule (inference)
7. Convert the output data to non-fuzzy values (defuzzification)

b. Linguistic Variables
Linguistic variables are the input or output variables of the system whose values are words or sentences from a natural language, instead of numerical values. A linguistic variable is generally decomposed into a set of linguistic terms.

c. Membership Functions
Membership functions are used in the fuzzification and Defuzzification steps of a FLS, to map the non-fuzzy input values to fuzzy linguistic terms and vice versa. A membership function is used to quantify a linguistic term.

d. Fuzzy Rules
In a FLS, a rule base is constructed to control the output variable. A fuzzy rule is a simple IF-THEN rule with a condition and a conclusion.

e. Inputs to the system:
Student profile (Area of interest, Category(Excellent, Average, Below average)

f. Rules to the interface:
While (area of interest==X)
Select X
If category == Excellent
Show advanceLO’s
Else if category == Average
Show intermediate LO’s
Else
Show Beginner LO’s

g. Outputs of the system:
LO’s (Advance Level, Intermediate Level, Beginners Level)

h. Fuzzy Set Operations
The evaluations of the fuzzy rules and the combination of the results of the individual rules are performed using fuzzy set operations. The operations on fuzzy sets are different than the operations on non-fuzzy sets.

After evaluating the result of each rule, these results should be combined to obtain a final result. This process is called inference. The results of individual rules can be combined in different ways.

i. Defuzzification
After the inference step, the overall result is a fuzzy value. This result should be defuzzied to obtain a crisp output. This is the purpose of the defuzzifier component of a FLS. Defuzzification is performed according to the membership function of the output variable.

V. FEEDBACK ALGORITHM

a. Average rating:
This is the model we use as a baseline measure. The estimated quality of an object is the average rating it has received from all authors in the training data.

b. Median rating:
This is set up identically to average rating, except for the statistic used. Here the estimated quality of an object is the median rating the object has received from all reviews for the object.

c. Lower bound on normal confidence interval:
Some Learning objects have more consistent ratings than others. For example, we would like to give a higher score to a Learning object that has received 100 5-star reviews than to a Learning object that has received a single 5-star review, even though the average rating model would give these the same score. We approximate ri ~ N(qi, σ2i) that is, a rating for a Learning object falls in a distribution around its true quality, with some variance. We then use the lower bound for the quality score. More precisely, qi = ri - za/2σi√|ri|, where the constant za/2 = 1.96, for a 95% confidence.

d. Lower bound on binomial confidence interval:
Such a normal approximation may not be accurate. However, we could instead simplify the star ratings into positive/negative for instance, every rating of 4 stars or above is positive and then take the lower bound of the confidence interval of the percentage of positive reviews. Also known as the Wilson Score, it is calculated in the following manner: First, obtain p̂, the proportion of “positive” ratings for a given object oi. We also define n=|ri| the number of reviews for an object. Next, the statistic is

\[ \hat{p}_i = \hat{p} + \frac{z^2_α}{2n} - \frac{z_α}{n}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]

This score is calculated in a few steps:
1. Aggregate reviews by source, and for each object, calculate the average rating ri from all recommenders j that are reviewers from that source.
2. Sort all objects oi for each source, based on that average.
3. From each sorted list, assign a percentile score for a source-object pair.
4. For each object, take qi to be the average percentile it receives for all its sources.

VI. VARIOUS DISTRIBUTIONS

Uniformly distributed ratings
First consider the item that receives an equal number of ratings at each star level. That is, let n=n1=...=nK

Also let a=1=...=aK=n+1.

To specialize the variance formula for an equal number of ratings at each level, we substitute ak=\alpha
and aK=Ka into (1):

\[ \text{Var}(f(p)) = \frac{1}{K+1} \left(\sum_k = 1^{Ks} k^2 aK - K^2 aK \right) \]

Then simplify:

\[ \text{Var}(f(p)) = \frac{1}{K+1} \left(\sum_k = 1^{Ks} k^2 - K^2 \right) \]

The above formula applies to any rating system (that is, arbitrary values of sk). For a star-rating system, substitute sk=k:

\[ \text{Var}(f(p)) = \frac{1}{K+1} \left(\sum_k = 1^{Ks} k^2 - K^2 \right) \]

Then get rid of the second summation with an identity and simplify:

\[ \text{Var}(f(p)) = \frac{1}{K+1} \left(\sum_k = 1^{Ks} k(2k(N+1)) - K^2 \right) \]

In terms of the total number of ratings N:

\[ \text{Var}(f(p)) = \frac{1}{N+K+1} \left(\sum_k = 1^{Ks} k^2 - K^2 \right) \]

Incidentally, the formula is easily inverted to determine the number of ratings required to produce a given amount of variance:

\[ N = 1 \text{Var}(f(p)) \left(\sum_k = 1^{Ks} k^2 - K^2 \right) \]

Now we can specialize the formula to a five-star system with K=5 and sk=k:

\[ \text{Var}(f(p)) = \frac{15(N+1)}{(5+1)(5)} \]

Using the normal approximation above, the width of a credible interval for the mean becomes simply:

\[ w = \sqrt{\frac{2\alpha K}{N+1}} \]

Rewriting in terms of N:

\[ N = 8z_α2w^2 - 6 \]

For example, a 90% credible interval (z=1.65) with a width of a half-star (w=0.5) requires:

\[ N = 8z_α2w^2 - 6 = 8(1.65)^2(0.5)^2 - 6 \approx 81 \]

or about 16 ratings at each of the five star levels. A 95% credible interval (z=1.96) spanning a full star (w=1) requires:

\[ N = 8z_α2w^2 - 6 = 8(1.96)^22(1.96) - 6 \approx 25 \]

or about 5 ratings at each of the five star levels.

Consensus ratings
Next we consider an item that is receiving only the highest star rating from users. That is, let n1=...=nK=1 and nK=N. and correspondingly, let a1=...=aK-1=1 and aK=N+1.

Plugging these values into the variance formula (1):

\[ \text{Var}(f(p)) = 1a0 + \left(\sum_k = 1^{Ks} k(2k(N+1)) - K^2 \right) \]

Factoring out the shared a0:

\[ \text{Var}(f(p)) = 1a0(a0+1) \left(\sum_k = 1^{Ks} k^2 - K^2 \right) \]

Plugging in the star formula sk=k and using the sequence summation identity:

\[ \text{Var}(f(p)) = 1a0(a0+1)\left(\sum_k = 1^{Ks} k^2 + 2K(N+1)1a0(\sum_k = 1^{Ks} k^2 + 2K(N+1)2 + K(N+1))2 \right) \]

Multiplying out the square in the second term:
Var([p])=1\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K(N+1)−K2\alpha_0((K−1)2+
+(K−1)(N+1)−(N+1)2)

Using the fact that \alpha_0=N+K

the equation simplifies to:
Var([p])=1\alpha_0(a+1)\Sigma_k=1K−1k2+K2\alpha_0(−1)

The above equation contains both quadratic and cubic terms for \alpha_0

, but we can use it to arrive at a convenient upper bound on the variance:
Var([p])<\Sigma K−1k=1k2α0(a+1)=\Sigma K−1k=1k2(N+K)(N+K+1)

For the five-star case, the formula specializes to:
Var([p])<30(N+5)(N+6)

The width of the credible interval is then approximately bounded by:
\text{w}<2\alpha\sigma/230(N+5)(N+6)

Which can be used to produce a conservative approximation of \( N \):

\( N=11\alpha\sigma/2w−5.5 \)

Repeating the credible interval calculations from the first example, a 90% credible interval with a width of a half-star requires at most:
\( N=11\alpha\sigma/2w−5.5=11(1.65)0.5−5.5≈31 \)

Likewise, a 95% credible interval spanning a full star requires no more than:
\( N=11\alpha\sigma/2w−5.5=11(1.96)1−5.5≈16 \)

Note that \( N \) and \( w \) have an inverse linear relationship in the consensus case, whereas they had an inverse quadratic relationship in the uniform ratings example. There is less noise when everyone agrees on the rating, and so tightening the credible interval requires fewer observations.

Polarized ratings

Finally, let’s consider an item that receives an even mix of the highest and lowest ratings. That is, let \( n1=nK=N/2 \) and \( n2=...=nK−1=0 \), and likewise let \( aK=\alpha K=\alpha K+1 \) and let:
\( \alpha K=\alpha K+1=1 \)

Once again, we plug these values into the variance formula (1) and pull out the shared \( \alpha_0 \):
Var([p])=1\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0((K−1)2+
+(K−1)(N+1)−(N+1)2)

Simplifying:
Var([p])=1\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1)

0((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1)

Plugging in \( sk=K \)

and eliminating the second summation:
Var([p])=1\alpha_0(a+1)\Sigma_k=1K−1k2+K2\alpha_0((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1)

Consolidating the terms inside the square, we have:

\[ \text{Var}(\hat{p}) = 1+\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1) \]

Since \( \alpha_0=N+K \), we can write this as:
\[ \text{Var}(\hat{p}) = 1+\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1) \]

A bit of algebra reveals:
\[ \text{Var}(\hat{p}) = 1+\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1) \]

Assuming \( N>K \), we can construct a lower bound for the variance, which happens to be the same as the upper bound in the consensus ratings case:
\[ \text{Var}(\hat{p}) = 1+\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1) \]

This bound confirms that, to reduce the variance of the average to a certain level, the polarized ratings will require more observations than consensus ratings, which makes sense as polarized ratings exhibit more variance than consensus ratings.

We can also plug in numbers to come up with a specialized version of the variance for the five-star case:
\[ \text{Var}(\hat{p}) = 1+\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1) \]

Which can also be written:
\[ \text{Var}(\hat{p}) = 1+\alpha_0(a+1)((N+2)\Sigma_k=1K−1k2+K2\alpha_0(−1) \]

We can see that the variance is roughly twice that of the uniform ratings case, and thus for a given \( N \), the credible intervals will be inflated by a factor of roughly \( 2\sqrt{2} \). Likewise, to achieve a credible interval of a certain width, polarized ratings will require roughly twice as many observations as uniform ratings. Since polarized ratings represent the maximum possible variance of observed ratings, a good rule of thumb for calculating the worst-case \( N \) required to achieve a given credible interval is to use the approximate formula, following:
\[ N=16\alpha\sigma/2w−6 \]

For the 90% / half-star interval, the formula produces:
\[ N=16\alpha\sigma/2w−6=16(1.65)0.5−6=168 \]

For the 95% / full-star interval, the formula gives us:
\[ N=16\alpha\sigma/2w−6=16(1.96)1−6=55 \]

Sample size table

The following table summarizes the formulas for the \( N \) to achieve various credible intervals in a five-star rating system for the three distribution assumptions.

<table>
<thead>
<tr>
<th>Uniform N</th>
<th>Consensus N</th>
<th>Polarized N</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N=8\alpha\sigma/2w−6 )</td>
<td>( N=11\alpha\sigma/2w−5.5 )</td>
<td>( N=16\alpha\sigma/2w−6 )</td>
</tr>
</tbody>
</table>

Below is a summary of example values computed from the above (approximate) formulas:

<table>
<thead>
<tr>
<th>Width (Stars)</th>
<th>Credibility Level</th>
<th>Uniform N</th>
<th>Consensus N</th>
<th>Polarized N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>80%</td>
<td>7</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>1.0</td>
<td>90%</td>
<td>16</td>
<td>13</td>
<td>38</td>
</tr>
</tbody>
</table>
The figures for Consensus and Polarized can be seen as representing the lowest and highest theoretical values of N, since they have the lowest and highest theoretical amount of variance for five-star system. (Note that because Consensus calculations use an approximation, which breaks down for low N, the Consensus N in the first row are higher than the N for Uniform.) Uniform might be seen as a kind of middle ground for a sample size calculation, and perhaps a conservative one for data that exhibits at least some amount of consensus on rated items. In most cases it will be preferable to calculate the credible interval on a per-item basis, but when that is not possible, the formulas and table above should provide some useful guidance.

VII. CONCLUSION

Research should be focused on incorporating additional contextual factors in the LO recommendation process for teachers, such as the ICT Competences of their school. The reason for this is that such contextual factors have been repeatedly reported as significant in the level of uptake of ICT in the formal learning processes. Therefore, accommodating them within the LO selection process would potentially extend the scope of the proposed RS and improve its capacity to facilitate teachers in selecting appropriate Los. Usage of Hybrid recommended system and the best use of rating algorithm that is normal distribution and Bayesian distribution to given an appropriate rating for a Learning object.

REFERENCES


