DESIGN AND DEVELOPMENT OF CONTEXT-AWARE RECOMMENDATION STRATEGY FOR E-LEARNING

SHAINA RAZA1, SYED RAZA BASHIR2, MUHAMMAD TAHIR HAMEED3, MUHAMMAD JAWWAD ZAHEER4

1Faculty of Information Technology, University of Central Punjab (UCP)
2Department of Computer Science & Engineering, University of Engineering and Technology (UET), Lahore
3Department of Mathematics, Lahore Garrison University (LGU), Lahore
4Department of Computer Science & Engineering, University of Engineering and Technology (UET), Lahore

{shainaqamar, razabashir55}@hotmail.com, {Tahirhameedch1, mjawwazaheer}@gmail.com

Revised August 2015

Abstract: The practice of retrieving and recommending Learning Objects (LOs) to the learners according to their specific needs and requirements has been a very active research area in e-learning. This paper proposes the design and development of a context-aware methodology that comprises a Learning Object Repository (LOR), context-aware recommendation engine and a user-friendly interface. The existing approaches in this regard focus on learners’ ratings, history, behavior and interests, rather ignored the knowledge gain and learning outcomes by the learners. The paper contributes in the research in threefold manner. First, a comparative survey of existing research in this area is presented. Secondly, the design and development of context-aware methodology for recommending LOs to the learners is proposed. Third contribution of the research is a mapping algorithm. Finally, it provides directions for the future research in this area.

Keywords: Learning Object Repository (L.O.R); Learning Object Metadata (L.O.M); Learning Object(LO); Recommender System( RS); E-learning; Technology Enhanced Learning (TEL)

1. Introduction. E-learning (or electronic learning) refers to the provision of education, learning or instructional program through electronic technology [8]. In web-based learning, learners and instructors are not bound to space, time and distance; and the gap between them is bridged with the practice of online technologies. Each content item in e-learning domain is called Learning Object (LO) and more formally defined as “any entity, digital or non-digital that may be used for learning, education and training”, as per IEEE 1484.12.1-2002 LTSC definition[25]. LOs are composed of learning content and metadata (data about data). Learning Object Metadata (LOM) is a data model that consists of a hierarchy of data elements and is used to define a LO [23].

LOM enables exploring and cataloguing of LOs within shared digital repositories. Learning Object Repository (LOR) is a “digital repository that enables access to learning objects at local or distributed level” [45].

The strength of e-learning domain is reusability and interoperability. It requires assembling the chunks of learning from the open web and making them interoperable, reusable and sharable. However the issue in traditional e-learning systems is that the learners are confronted with bulk of learning material, most of which consists of irrelevant materials and therefore hinders the learning process. Personalization in e-learning is solution to this problem. Personalization allows individualization of interaction between system and the learner based on learners’ personal requirements and preferences and helps them shape their own learning paths (Jain, 2011). Personalization can be achieved by using Recommender System (RS).

A recommender system can be defined as an information filtering system which predicts users’ preferences for a particular item [43]. RS in e-learning differs from those of social networks in that pedagogical items or LOs
are recommended to the learners (consumers) based on their characteristics and attributes. Things that interests learners may not be pedagogically appropriate for them therefore e-learning RS consider not only learners’ interests but also their assessment mechanism. For example, there is no sense of recommending a technical paper to a first semester undergraduate student or a popular magazine to a PhD student. Unlike RS in e-commerce, it is very important to keep a pedagogically ordered list of recommendations because the decision of making the best choice cannot be left at learner’s end.

The concept of context in RS is gaining popularity in E-learning initiatives and can be more formally defined as “any information that features the situation of an entity” [1]. Therefore, a new RS i.e. Context-Aware Recommender Systems (CARS) has evolved recently to demonstrate the potential contextual recommendations in web-based learning [55]. This paper presents a context-aware recommendation strategy to make intelligent recommendations to the learners based on their characteristics with particular focus on assessment parameter.

2. Background. E-learning initiatives invested much in terms of budget, technology, software certificates, content development, content delivery, setup, teaching strategies, training, practices and maintenances to provide a replacement to traditional classroom system [8]. Despite all these efforts, the dropout ratio of e-learning is much higher as compared to traditional system. The major loopholes in typical e-learning systems are as:
   i. LORs in typical e-learning systems are static i.e. learning content has to be pre-loaded in them .
   ii. Attempting to load LORs dynamically from open web will create overwhelmingly huge storage with limitless choices
   iii. The system suffers from the inability to satisfy heterogeneous pedagogical needs of the learners. One of the possible solutions to aforementioned problems is to make use of RS. A survey on types of recommender systems is summarized as in table 1:

<table>
<thead>
<tr>
<th>RS</th>
<th>Definition</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative filtering (CF) (Goldberg,1992)</td>
<td>make assumptions based on collaborative past ratings of similar users</td>
<td>No need to analyze content to make suggestions. Recommendations can be pre-computed in advance</td>
<td>Cold start (requires abundant data to recommend) Sparsity (majority of items are never rated) Ramp up problem: needs sufficient number of users and ratings.</td>
</tr>
<tr>
<td>Content based (CBF) (Pazzani &amp; Bilsus 1997, Basu,1998)</td>
<td>Assumptions are based on the similar content properties between an item and a user’s profile</td>
<td>Independent of other users No first-rater problem</td>
<td>Content analysis is limited New user problem</td>
</tr>
<tr>
<td>Context-aware (Adomavicius ,2011)</td>
<td>Recommendations are based on the contextual information (time, space, mood, environment, system etc.)</td>
<td>Can deploy other RS approaches along with context</td>
<td>Incorporating context is not a simple thing</td>
</tr>
<tr>
<td>Knowledge based (Trewin &amp; Shari ,2000)</td>
<td>Based on explicit knowledge of user’s information, product features and other contextual parameters to generate reasoning to find out which item meets user’s needs</td>
<td>No ramp-up problem Don’t depend on large statistics of data to make suggestions</td>
<td>Needs prior knowledge of items and users’s needs</td>
</tr>
<tr>
<td>Demographic (Krulwich 1997)</td>
<td>Recommendations are made on the personal data (e.g. social statistics of general population) and items are classified into user classes.</td>
<td>No need of history of ratings or prior knowledge</td>
<td>Limited accuracy Needs textual reviews to improve accuracy.</td>
</tr>
<tr>
<td>Types of Recommender Systems</td>
<td>Method applied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constraint-based (Felfernig, 2008)</td>
<td>Recommendations are bound to certain conditions. Successful where items are complex, infrequently purchased and many features are not known to customers. Often neglect users’ preferences.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybrid (Wang, 2003)</td>
<td>Combines a number of techniques most popularly CBF and CF. Covers the short-comings of one or more techniques. Complex to design.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tag-aware (Zhang, 2011)</td>
<td>Recommends user tags to assist in annotating items. Works well where Timeliness is an issue. 3-dimensional recommendation function needs to be reduced to 2-Dimensional.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-aware (Campos, 2014)</td>
<td>Includes time dimension in user-modelling and recommendation function. Time when user rated the item is useful. User-item rating computational is costly.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Federated RS (L. Zhou, 2012)</td>
<td>Rich set of recommendations that depends on users’ interaction with distributed environment and resources. Provides centralized access point for resources. Needs complex algorithmic approach to efficiently combine different search results.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location-aware RS (Levandoski, 2012, Symeonidis, Ntempos, &amp; Manolopoulos, 2014)</td>
<td>Uses spatial ratings to produce recommendation function. These techniques are scalable and efficient. Dependent on location. Therefore future predictions are not so effective.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-based Reasoning(CBR) (Derek, 2005)</td>
<td>Make recommendations by reusing the statistics of similar and already suggested products or sample cases. Ability to match users’ preferences in the recommendation cycle. Recycle the process if the user is not satisfied. Which case is to be selected is often unclear and undefined.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic RS (Peis, 2008)</td>
<td>Make use of knowledge base that is usually defined as ontology. Promising in terms of short-term and mid-term results. Needs support of a number of supporting RS techniques.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Types of Recommender Systems

Various applications of RS can be seen in e-commerce, web, travel, mobile phone technology, healthcare, TV, tourism, music, movies and many other fields. Majority of researchers have employed various techniques as mentioned in table 1 to propose different RSs. A number of diversified RSs are also implemented in e-learning. Tang & Gordon [52] are reported to be the first ones to propose a smart adaptive RS for e-learning environment. Verbert[36] and Manouselis [35] presented a number of surveys on various RS in e-learning during the last decade. To summarize; a survey of various RS techniques in e-learning is given in table 2:
Table 2: Recommender Systems in e-learning

<table>
<thead>
<tr>
<th>Type</th>
<th>Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content based filtering (CBF)</strong></td>
<td>- machine learning techniques [33]</td>
</tr>
<tr>
<td></td>
<td>- Vector space model[21]</td>
</tr>
<tr>
<td></td>
<td>- Vector space model[27]</td>
</tr>
<tr>
<td><strong>Semantic based</strong></td>
<td>- Ontology and knowledge base [42]</td>
</tr>
<tr>
<td></td>
<td>- Personalized relevance measure, context aware recommendation, ontology based with CF[11]</td>
</tr>
<tr>
<td></td>
<td>- Personalized hybrid RS, ontology, CF without rating[40]</td>
</tr>
<tr>
<td></td>
<td>- memory-based &amp; model-based CF , hybrid fuzzy-genetic approach[3]</td>
</tr>
<tr>
<td></td>
<td>- ontology, Singular Value Decomposition(SVD), data mining techniques[16]</td>
</tr>
<tr>
<td><strong>Hybrid filtering</strong></td>
<td>- CBF and CF, crawling and indexing techniques, Web usage mining techniques (clustering and association rule mining)[27]</td>
</tr>
<tr>
<td></td>
<td>- CBF, CF, attribute-based approach(46)[58]</td>
</tr>
<tr>
<td></td>
<td>- Relational filtering, profile-based filtering, CF techniques [34]</td>
</tr>
<tr>
<td></td>
<td>- matrix factorization techniques, user-to-user collaborative recommendation algorithm[12]</td>
</tr>
<tr>
<td><strong>Personalized/Adaptively</strong></td>
<td>- probabilistic language modeling &amp; automatic text summarization techniques [59]</td>
</tr>
<tr>
<td></td>
<td>- CF and item response theory [14]</td>
</tr>
<tr>
<td></td>
<td>- Fuzzy item response theory[15]</td>
</tr>
<tr>
<td></td>
<td>- Item response theory and artificial neural network [6]</td>
</tr>
<tr>
<td></td>
<td>- genetic algorithm and case-based reasoning [24]</td>
</tr>
<tr>
<td><strong>data mining</strong></td>
<td>- Bayesian belief network (Kumar, 2005, p. 685-687)</td>
</tr>
<tr>
<td></td>
<td>- Apriori prefix span algorithm, Nearest neighborhood algorithm and preference matrix[46]</td>
</tr>
<tr>
<td></td>
<td>- Web mining approach[44]</td>
</tr>
<tr>
<td></td>
<td>- Empirical research using data mining methods and visualization tools[48]</td>
</tr>
<tr>
<td><strong>Context aware</strong></td>
<td>- analysis of existing TEL RS and their possible dimensions[54]</td>
</tr>
<tr>
<td></td>
<td>- Context-aware method, CF and association mining[56][39]</td>
</tr>
<tr>
<td></td>
<td>- Proactive algorithms [20]</td>
</tr>
</tbody>
</table>

**Problems with existing RS techniques in e-learning.** The first group of related works discusses CF techniques which proved to be quite useful as it generates information by considering the suggestions of other related users. However, it suffers from data sparsity and cold start problem i.e. insufficient new item’s ratings [49]. Verbet[36] applied CF algorithms to obtain learner’s preferences in e-learning environment. The experimental study highlighted the data sparsity issue using this technique. The authors also recommended a solution to this issue that makes use of a collection of adequately massive datasets to capture learner’s behavior. Data sparsity has also been pointed as a major problem in CF by Wen & Wang [56] who proposed to use cosine similarity in replacement to inner product for sparse user/item matrix. Experiments by Wen [57] shows that there is no need of auxiliary dataset as is required by Verbet & Duval[54] approach. But cosine similarity has the overhead of computing explicit score normalization [17]. However, RS in e-learning can benefit from multi-attribute CF algorithms as they are capable of parameterizing multiple rating data captured from numerous online portals[38]. But the multi-attribute CF is biased towards clustering users and resources and ignores the dissimilar characteristics of alike users which eventually lead to recommendation deviation [32].

The second group of related works is about content based RS that are implemented in e-learning. A key issue of CBF is over-specialization due to which only those items are recommended to the users that score high against user's profile. In general practice, items similar to those which have already been rated by the user should also be preferred [33]. Ghauth & Abdullah [21] proposed a new e-learning RS that uses CBF technique, good learner's rating and pre-test to post-test to get feedback from the user. However the proposed methodology is prone to 'cold start' problem and the authors suggest using better approach in future. Khribi, Jemno and Nasraoui[27] proposed an automatic RS that uses learners’ browsing histories and preferences. They used web usage mining techniques along with CFB and CF algorithms to make recommendations. The shortcoming in
their proposed methodology is that there is need to improve the learner’s model by including learners’ preferences, pedagogical level and profiling information.

The third group of related works deals with semantic based RS in TEL. Semantic RS allows dynamic contextualization and limits the ‘cold start’ problems that are typical of other RS [40] however the approach is deficient in distribution environment and a generic domain. Cantador [11] presented a RS that uses CBF and CF models with addition to semantic context aware recommendation approach. Typically the recommendation engine doesn’t take into account those news items that are related but not directly addressed by the profiles. The proposed approach makes use of ontology with additional weight concepts and the running context but rather ignored users’ preferences. Nadolski [40] contributed by creating a simulation environment for informal learning networks in which he compared different variants of recommendation algorithms in a hybrid RS. The semantic approach is also used by Al-Shamri [3] and Cobos[16] together with multiple information retrieval techniques.

In addition to this, multiple RS are being proposed in e-learning that uses hybrid approach to recommendation algorithms. Khribi [27], Salehi [46] and Wen & Fang [58] used hybrid version of CF and CBF approaches in their RS proposals but what is missing in the aforementioned models is the lack of ability to recommend appropriate learning content to the learners according to their learning outcomes. Menderia (2013) and Capuano [12] proposed integrating hybrid RS with e-learning environment however there is need to improve rating function to make better recommendations.

Personalization in e-learning involves learners’ preferences; behavioral patterns and interests but often learner’s ability and content difficulty are ignored. Chen [14][15] and Baylari [6] proposed personalization collaborated with item response theory to provide individual learning paths to the learners. Both systems emphasize on learner’s ability but the latter one supersedes the former by incorporating learner’s learning problems using artificial neural networks. MJ Huang [24] adds to the personalization research by employing genetic algorithm and case-based reasoning to provide individualized optimal learning paths. However limited research is available on such learning behaviors and learning activity patterns.

Knowledge Discovery and Data Mining (KDD) is the discipline that deals with the process of extracting knowledge from data. With the advent of e-learning, researchers begin investing various KDD methods to improve the underlying systems[44,46,48]. A hybrid approach (CBF and CF) proposed by Kumar [29] implements Markov model to evaluate qualitative LOs. The proposed approach differs from its contemporary ones in that it makes use of Bayesian Networks to compare the learning object quality reviews to overcome any deficiencies. Salehi [46] proposed a hybrid RS that implements sequential pattern mining approach and CBF methodology using Apriori and prefix span algorithms respectively. The proposed framework makes an attempt to discover latent patterns of the learning content particularly from the perspective of learners’ attributes. As an extension to their previous research, Salehi [47] make use of weighed, cascaded and hybrid methods to generate recommendations. Comparative to other traditional RS algorithms, the results outperforms in terms of precision, recall and other evaluation metrics.

The literature shows that there are few CARS available for e-learning. Verbert [36], Manouselis [35] presented a survey of recent CARS that are being implemented in e-learning and therefore invites researchers to drive contextualization in e-learning. Henceforth, their research outlines future directions and challenges for the deployment and validation of CARS. It emphasizes that learners’ preferences are influenced by multi-dimensional contexts [51].Berri [7] contributes to contextualization by differentiating between technical (operational devices) and learners’ (aims, goals, background, assessment levels) context elements. Gallego [20] argues that LOs recommendations should be proactively pushed to the learners according to the context without their explicit feedback. Mayeku & Betty [39] emphasized that human intelligence cannot be fully replaced by absolute machine control, therefore context-awareness can be used to support engaged learners’ experience. Tang & Gordan [52] being the first ones who proposed a RS for e-learning almost a decade ago, came with a newer proposal in the same field[51]. They argued that multi-dimensional recommendation function, though complex in nature produces better results on a diversified group of learners as compared to CF techniques. Thus, CARS seems to be most fruitful when it comes to effective personalization.

In this research, a context-aware recommendation methodology is proposed that is based on dynamic assessment mechanism and learners’ feedback. Similar to previous related work, learners’ preferences shall be considered foremost with the addition of their evaluation and assessment mechanism.

3. Proposed Methodology The proposed system consists of a database, recommendation and an interface layer. The database layer includes a LOR and their metadata record that suffices the purpose of retrieval, reusability, interoperability and sharing. Repository software can be created either by the instructional designers
or they can use third party software to avail the educational economy. Taking inspiration from Biletskiy [9] for how to create a LOR, a crawler is used for massive downloading of LOs along with their metadata. In this case, Apache Nutch [4] is used to crawl on the web and Apache Solr for providing indexing. The middle layer is the recommendation layer that uses context-aware methodology to further improve, customize and rank search results from LOR. The RS layer explores its functionality through a Graphical User Interface (GUI) that allows an easy to use and interactive environment. A general description of the proposed framework is shown in the figure 1.

Figure 1: Overall Proposed System

a. **Context Aware Methodology for Recommendation of Learning Objects**

i. **Context in RS in e-learning.** The concept of ‘context’ in RS for e-learning differs substantially from other fields. Apart from generic factors such as location, time, environment, current activity etc., it also includes individualized learner’s learning paths, collaborations, learning history, timing, moods and behavioral patterns [2]. This paper puts forward another important contextual element that is often ignored in education based RS i.e. assessment and evaluation mechanism pertaining to a particular learner.

ii. **Initial Input.** The model begins with the initial input of calculating main contextual values pertaining to the learners. Typically a recommendation function can be viewed as a prediction problem which takes as input user profile and a target item and predicts users’ rating for that item [2]. The recommendation function is defined in equation 1 as:

\[ R: \text{User} \times \text{Items} \rightarrow \text{Ratings (1)} \]

Unlike traditional models, CARS incorporates additional information to approximate users’ preferences on unseen items. In e-learning, context-awareness provides more information about a learner in order to make more accurate and personalized recommendations. This leads to a multi-dimensional recommendation [2] function as shown in equation (2):

\[ R: \text{User} \times \text{Items} \times \text{Contexts} \rightarrow \text{Ratings (2)} \]

Here ‘contexts’ represent consists of not only explicit users’ preferences but also latent factors derived from the observations of users’ actions and environment.
Representing and Modelling Context. There are three major approaches to obtain contextual information i.e. explicit, implicit or inferred from observed data [2]. Those contextual factors whose values are known at the time of recommendation are known as explicit factors (e.g. in case of purchase, when, where and with whom the purchase is made). Explicit information can be gathered either directly or through other elicitation means. Implicit contextual factors are those which are gathered from the observed data or the environment (e.g. spatial, temporal, surrounding etc.). Contextual information can also be inferred using statistical or predictive models (e.g. the frequency a user is changing the mode of his activity etc.). There exist two algorithmic paradigms to incorporate contextual information into RS [2]:

**Contextual pre-filtering (PreF):** contextual information is used in prior to filter out irrelevant ratings before computing recommendations.

**Contextual post-filtering (PoF):** contextual information is used after the application recommendation.

Given the learners’ problem, the most suitable contextual approach would be PoF, since PreF method could narrow down the exact context in the proposed architecture. Another key issue is how to produce ratings i.e. explicitly or implicitly. Seeing the nature of the problem, implicit manner of getting ratings would be a better approach. A latent predictive model is to be used by RS to estimate unknown ratings and unobservable contextual variables.

Evaluating Learning Material. The system also proposes a module that classifies LOs according to their difficulty index. The objective of this module is to find, quantify and rank document according to their complexity level. According to IEEE 1484.12.1 – 2002 standard, LOs are defined as “any reusable digital entities (audio, video, text, image etc.) that may be used for learning, education and training” (IEEE LTSC) [25]. They need to be classified according to their difficulty index so that they can be later recommended to the learners based on their evaluation. E.g. a learner that is classified as being “A” grader should be recommended those LOs whose difficulty index is highest or relatively higher. A simple approach to classify LOs for the proposed system is defined as in the subsequent sub-sections.

Annotation of LOs. Extracting relevant metadata tag e.g. “Educational” element (IEEE LTSC) from Learning Objects Metadata (LOM) profile which are embedded with XML binding in LOR. This is useful method for both textual and non-textual LOs or part of them.

Readability Index. Sometimes it becomes necessary to strictly find the exact difficulty index of a LO that majorly constitutes the textual data. The readability assessment metrics are therefore applied on such LOs. Here readability refers to the comprehensiveness and ease of the reading material. There is a number of readability metrics [18] proposed in the literature. They are being classified as being classical and modern. The traditional readability metrics includes those famous formulae: Flesch Reading Ease formula, Flesch-Kincaid Grade Level, Fog Scale, SMOG Index, Coleman-Liau Index, Automated Readability Index, Laesbarhedsindex, New Dale Chall, Powers-Sumner-Kearl Readability Formula, RIX, Kandel & Moles [62]. The modern research in readability measures includes strain index (2007), McAlpine EFLAW (2004), Fernandez-Huerta measure (2008), Hayashi / Tateisi (1992), Douma, (1996), Lexile measure (2011), Coh-metrix indices (2010), DeLite system for German (2006), Gulpease Index (1988), Degrees of Reading Power (DRP) (1981) and naïve bayes [62].

Evaluating Learners. Students are being assessed on different levels e.g. advanced, mediocre and beginners. A number of instruments are used to assess their competency e.g. Class tests, Moderated Discussion Boards (MDBs), examination, experimental exercises, analysis of videotaped interviews, tutorial problems, simulation exercises, essay questions, GPA, help-seeking and other behavior. Learners are assessed by conducting pre-tests (two-tailed) to post-tests (one-tailed) in the learning environment.

Mapping Algorithm. An important algorithm in this paper is to map clusters of LOs, categorized according to their complexity level, to the clusters of learners, grouped according to their competency levels. Therefore a mapping function is designed that has the following set of steps:
4. Conclusion. This paper presents a context-aware methodology to be implemented in the e-learning environment. A distinctive feature of this proposal is that it takes into account learners’ assessment along with other contextual attributes pertaining to the learner. Mapping learners’ competency level with the classified LOs is a key challenge to be met in this study. In short, this proposal outlines a number of key ideas that need to be explored for the purpose of sufficing the e-learning environment in real sense. In future, an extensive survey of Top-N algorithms will be studied to implement a suitable variant during the Pref phase in the recommendation function.

REFERENCES


