CHAPTER 6

Active, context-dependent, data-centered techniques for e-learning: a case study of a research paper recommender system

T. Tang\textsuperscript{1,2} & G. McCalla\textsuperscript{2}

\textsuperscript{1}Department of Computing, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. \textsuperscript{2}ARIES Laboratory, Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Abstract

In this chapter we discuss an e-learning system that recommends research papers to students wishing to learn an area of research. Recommender systems in the e-learning domain have specific requirements not present in other domains, most importantly the need to take into account pedagogical aspects of the learner and the system (such as the learning goals of each) and the need to recommend sequences of items in a pedagogically effective order. First, the architecture and some of the basic methodologies of the system are presented. Then, two studies are presented showing some of the ‘pedagogical’ characteristics affecting recommendations and comparing two recommendation algorithms: one that is content-based and the other that employs hybrid content-collaborative filtering and clustering techniques. We then generalize from the recommender system to discuss a general approach to the design of an e-learning application called the ecological approach, which is centered on finding patterns in learners’ interactions with learning objects and using these to actively compute information relevant to the application that is sensitive to the current end use context, especially to characteristics and goals of the learner. The ecological approach holds the promise of overcoming many problems in e-learning that have made systems controlling and unresponsive.
1 Introduction

The general problem of managing repositories of learning objects in order to support students who wish to learn about a domain is now being actively studied in artificial intelligence in education (AIED) and other e-learning research areas [1, 2]. We are working on a specialized instance of the general problem: recommending papers from a research paper repository to learners who wish to learn about a research area. The target audience for our experiments into the paper recommender system is graduate students and senior undergraduate students who wish to learn about the domain of data mining, although the approach is not restricted to this domain.

Most research in recommender systems have not been in e-learning domains, but instead in domains such as e-commerce or the most studied domain: movie recommendations [3–6]. There is, however, some specific research in tracking and recommending research papers. Basu et al. [7] define the paper recommendation problem as: ‘Given a representation of my interests, find me relevant papers.’ Their goal is to assign conference paper submissions to reviewing committee members. Reviewers do not need to key in their research interests as they usually do; instead, a novel autonomous procedure is incorporated in order to collect reviewer interest information from the web. Bollacker et al. [8] refine CiteSeer, through an automatic personalized paper-tracking module that retrieves each user’s interests from well-maintained heterogeneous user profiles. Woodruff et al. [9] discuss an enhanced digital book with a spreading-activation mechanism to make customized recommendations for readers with different types of background and knowledge. McNee et al. [10] investigate the adoption of collaborative filtering techniques to recommend papers for researchers. They do not address the issue of how to recommend a research paper, but rather, how to recommend additional references for a target research paper. In the context of an e-learning system, additional readings in an area cannot be recommended purely through an analysis of the citation matrix of the target paper.

Our research has a different focus from these other projects. We assume there is a pre-existing repository of relevant papers, rather than the open web (although we would eventually like to automatically retrieve papers from web sources). We also assume the goal is to put together a whole sequence of papers, in a broader e-learning context. In such an e-learning context:

- User interest is not the only relevant metric. Items liked by learners might not be pedagogically suitable for them, and vice versa. For example, a learner without prior background knowledge of the techniques of web mining may only be interested in knowing the state-of-the-art of web mining in e-commerce. Then, it should be recommended that he/she read some review papers, for example, an editorial article by two of the leading researchers in this area [11], although there are also many high quality technical papers related to his/her interest. On the other hand, for the learner coming from industry with some prior knowledge who wants to know how web mining can be utilized to solve e-commerce problems, should be recommended, because the paper is the KDD-Cup 2000 organizers’
report on how web mining can support business decision making for a real-life e-commerce vendor, and it points out challenges as well as lessons learned from the competition, which can benefit both researchers and industry practitioners. In other domains, recommendations are made based purely on users’ interests.

- More than one item must be returned and it is important to organize these items in a sequence specifically tailored to the learner [12]. This is because there are pedagogically relevant dependencies (e.g. prerequisite relationships) among learning objects that mean one object should be ‘consumed’ by the learner before another, for example an introductory paper on data mining before more technical papers outlining various techniques. In contrast, in the movie domain normally only one movie is recommended at a time, and in e-commerce domains it is usually preferred to leave the list of recommended items unordered to avoid leaving the impression that a specific recommendation is the best choice [13].

Thus, in making recommendations to learners two things are of paramount importance: individualization to particular learners and the incorporation of pedagogical aspects into the recommendation algorithms.

In this chapter we will explore these issues further. We start with an overview of our research paper recommender system architecture, and show how data clustering techniques play a significant role. In Section 3 we discuss two studies we have undertaken: (i) an empirical study of actual learners and their preferences and goals that sheds some light on important versus not so important pedagogical goals and learner characteristics that we must take into account in the design of the research paper recommender system; and (ii) a study with simulated learners that sheds light on the trade-offs between two different recommender system techniques, content-based filtering and a hybrid content-collaborative technique. Then, in Section 4 we discuss an approach to the design of e-learning systems called the ecological approach which is a synthesis of ideas that has emerged from our work on the research paper recommender system and various other investigations in the ARIES laboratory at the University of Saskatchewan. Fundamental to the ecological approach is data mining, thus making this an appropriate culmination for a chapter in this book, although there is a brief concluding section that forms our official denouement to the chapter.

2 A research paper recommender system

We can state the goal for our paper recommender system as follows:

Given a collection of papers and a model of the learner, recommend and deliver a set of pedagogically suitable papers in an appropriate sequence, so as to meet the learner’s pedagogical needs as well as to be consistent with the learner’s background and other characteristics.

Ideally, the system will maximize a learner’s utility such that the learner gains a maximum amount of knowledge and is still highly motivated in the end.
Figure 1 shows the architecture of our research paper recommender system (for more detailed descriptions of each module see [14]). The system is primarily intended as a support tool in an e-learning system for senior undergraduate or first year graduate students who want to read pedagogically useful papers that may help them either in their class projects or future studies. The darkened part in the figure is the part that we are actively exploring, i.e. the modules that make the actual recommendations. It is a longer-term goal to develop the paper maintenance module to have the ability to automatically update the paper repository with new papers through a web crawler browsing CiteSeer or other such external sources to retrieve relevant new papers; however, we have not actively explored this yet.

The paper repository is where papers related to the course are actively maintained through the paper maintenance module. Papers must currently be added and removed manually, by tutors or learners, since the crawler and garbage collector modules are not currently being explored. Any explicit metadata must also be added.
by hand at this time. Learners are responsible for giving ratings and other assessments after interacting with a paper, and these are also kept as annotations with the paper in the repository.

The sense-maker module is responsible for filtering out loosely related related papers and clustering them into their appropriate topical categories. The sense making here is adaptively performed based on the collective learning behaviors and interests of users instead of an individual learner. For instance, the majority of learners might find the paper to be highly technical which requires more extensive background knowledge. Therefore, each paper’s technical tag evolves according to the collective usage and ratings of its learners. As discussed in Section 4, in the long run we think it is promising to consider annotating the papers with full learner models of each learner after they have interacted with a learning object such as a paper, and then mining these learner models to find patterns of end use relevant to various pedagogical goals. This is the basis of the so-called ecological approach to e-learning promoted there.

It is through the recommendation module that personalized recommendations are made. The recommendation module consists of two sub-modules: the data clustering module and the focused collaborative filtering module. The data clustering module clusters learners into a sub-class according to the purpose of the recommendation, while the focused collaborative filtering module will find the closest neighbor(s) of a target learner and recommend paper(s) to him/her according to the ratings by those closest neighbor(s). Tutors are responsible for setting up the curriculum and providing the basic learning materials. Based on this information, the system can select a set of papers for a learner.

Assume a learner L wants papers with content C, arranged in an appropriate sequence. There are two paper recommendation processes: content-based filtering and collaborative filtering. In content-based filtering, the idea is to find a cluster of papers with related content to C through content annotations attached to the papers in the repository. Information about the type of similarity between two papers can be useful in deciding which papers to include in the cluster and, later, how to order the papers. For example, one paper may be a version (e.g. an update or refinement) of another paper, one paper may take a different approach to the same topic as another, one paper may explore the same technique in a different context from another paper, one paper may be of the same level of difficulty as another paper, and so on. The recommender system could stop with this content-based filtering, but it is often useful to go further and apply collaborative filtering (CF) techniques to select a subset of other learners who have read papers in this content cluster who also have similar learner models to L (a learner cluster), and then uses learner cluster’s opinions to choose which papers to recommend to L. This hybrid-CF technique is explored further in Section 3.2.

Once a set of papers has been selected through one of these techniques, the set must be ordered according to various heuristics. The similarity metrics, above, can be useful. Other ordering heuristics can also be used, for example ordering according to technical difficulty (easy to hard), length, level of abstraction, prestige of publications (most prestigious to least), or date of publication. The ordering
decisions can be affected by information in L’s learner model, for example taking into account other papers L has read or experience recorded there about the kinds of papers L has liked or disliked or found useful (drawing on techniques from the instructional planning area [15]). In this way individualization can happen in both the selection of the papers (through CF) and their sequencing.

The recommendation process requires considerable information in the content annotations associated with the papers. As discussed above these annotations can be explicit ‘pedagogical’ and content metadata attached by experts like tutors, or, more interestingly, information explicitly or implicitly added by the learners as they have interacted with the papers. Such learner annotations can provide an increasingly sophisticated repository of data about actual end use experience of the learners, and potentially allow ever more refined decisions about which papers are relevant to which learners in which contexts.

This is a very brief overview of the recommendation process. More technical details and a formal description can be found in [16], and a discussion of learner clustering based on learners’ browsing behavior can be found in [17]. In the next section we will look at two experiments we have undertaken, one a human subject experiment with real learners, and the other a study with simulated learners, to explore the implications of pedagogical recommender systems.

3 Two experiments in paper recommendation

3.1 What learners want: a survey

The first experiment is a human subject study to determine some of the various factors affecting learners when they select papers. Results from this experiment have helped us to understand the relevant ‘pedagogical’ features we need to incorporate into the recommender system. We provided a questionnaire (see Appendix A) to 28 people and received 26 responses. The subjects were graduate students and alumni from computer science and engineering departments in Canada, Hong Kong, and China. The following learning scenario was introduced in the questionnaire: ‘Assume that you are taking a graduate-level class where you need to read several papers (as what we usually did). For each topic taught in the class, you are required to read 2 or more items (journal paper, workshop paper, etc.) recommended by the professor or your classmates.’ Then, 10 questions were asked that can be categorized into three groups: questions about learner preferences about different items, questions about content, and questions about delivery methods.

The results of the survey are shown in Fig. 2. The upper-left chart shows the learner preferences for type of paper, assuming the learner was interested in the topic being taught, while the lower-left chart shows the preferences if the learner was not interested in the topic. The vertical axis in these charts represents the number of respondents who liked/disliked the type of item shown below the horizontal axis. The upper-right chart shows learners’ preferences regarding the presentation of the paper. The lower-right chart tracks answers to various questions in the survey, showing learner preferences for various kinds of papers and recommendations.
The results show that magazine articles were the least popular items for learners who are interested in the topic being taught, but journal papers were least popular for learners who aren’t that interested in the topic. In either case learners would rely heavily on lecture notes.

Learners preferred graphical presentations in preference to papers with formal models or algorithms only. And, as shown in the first three bars of the lower-right chart, most of them preferred papers by well-known authors, from reputable conferences, and with up-to-date results or the latest version of the research being reported (Q6 of Appendix A). On the delivery issue (bar four in the lower-right chart), respondents preferred to know more about various approaches to solving a similar problem rather than learning a single approach repetitively, which indicates their enthusiasm for in-breadth exploration. For similar papers by the same authors, most of the respondents preferred up-to-date work (third answer of Q7 in Appendix A) rather than earlier versions. Respondents who preferred the earlier version of a paper (second answer of Q7) believed that the earlier version was less ‘distilled’ and thus easier to read (even though we explicitly stated in the question that both earlier and later-versions have a similar technical level). But they would read the up-to-date version, as well, if they found the topic to be interesting enough to pursue. The majority of respondents preferred to read an interesting but unimportant paper before they proceeded to read an important but uninteresting one (see bar 6).
An even bigger majority (bar 7) of respondents were still willing to eventually read that important, yet uninteresting, paper, however. This finding substantiates our claim that uninteresting, yet pedagogically valuable papers should be recommended. Bar 8 in the lower-right chart shows a balance between learners who preferred deep review paper(s) (with many technical aspects) to shallow review paper(s) (with many interesting presentations/application descriptions). In comments made on the survey, three respondents said that they would be willing to read both papers if they were interested in the topic being reviewed and two of them stated that they would skim the shallow review paper first before going deeply into the other one. When asked about whether they would follow a link to rich resources maintained by well-known researchers or research groups, half of the learners (bar 9) preferred the recommender system to provide more specific information before they would follow the link. A solution to this problem is to provide additional annotations so as to make recommendations more personalized and specific. A more detailed analysis of this study is available in [18].

From the above analysis, a number of factors are seen to be important in making recommendations. Some of these factors are more or less independent of individual differences in learners (for example the dislike of approaches that are only formal, and the preference for papers from well known authors in reputable conferences). However, many of the recommendations are affected by individual differences among learners, most importantly the goals and motivations of the learners and their level of interest in the topic. This study thus provides further evidence that personalization is important. Since it is already well known in the AIED literature that other individual differences (e.g. in the level of knowledge of the learners) are also of great importance in helping learners access appropriate content and keeping learners engaged, this confirms the requirement to be able to incorporate individualization techniques into the paper recommender system.

### 3.2 Evaluating pedagogy-oriented hybrid collaborative filtering

In the previous section our study reinforced the importance of individual differences in learner motivation and knowledge (among other things) in making recommendations. In Section 2 we discussed two recommendation techniques: pure content-based filtering and a hybrid content-collaborative filtering technique. In this section we describe a simulation experiment to compare these two techniques to determine how the papers each recommends may affect the knowledge and motivation of the learner after the paper is ‘consumed’. We provide only a brief overview; fuller details can be found in [16].

#### 3.2.1 Simulation setup

For the purpose of the simulation, we first generated 500 artificial learners and used 50 papers related to data mining as the main learning materials. A pure content-based recommender system then delivered recommendations of 15 papers to each learner according to each individual learner model. Each artificial learner rated these
papers according to their properties. After that, we generated 100 additional artificial learners, who became the target learners. Then, two recommendation techniques were applied for these target learners in order to evaluate their differences as well as performances. The first technique was the same pure content-based technique used with the first 500 artificial learners. The second technique used the hybrid-CF recommendation technique (content-based with collaborative filtering) described in Section 2.

3.2.2 Evaluation metrics and control variables
Normal metrics used in the recommender system community (e.g. the receiver operating characteristic (ROC), [4]) are not too relevant in recommending papers to learners. Metrics such as ROC are mainly adopted to test the ‘satisfaction’ of users in terms of item interestingness. However, in e-learning the most critical thing is to facilitate learning (not just to provide ‘interesting’ items). Therefore, we propose two new metrics:

- average learner motivation after recommendation.
- average learner knowledge after recommendation.

These should both go up if the recommended paper was appropriately chosen. In our simulation, we compared the percentage differences on these metrics between content-based recommendation and hybrid-CF recommendation.

We needed some way to simulate a variety of changes in motivation and knowledge after an artificial learner reads a paper. We also needed to capture the fact that some papers are more important (authoritative) than others. We thus created three control variables: \( x \), representing the motivation change of a learner after reading a paper; and the pair \((w_1, w_2)\), representing the knowledge gained by the learner after reading authoritative and non-authoritative papers. These were assigned differentially to different artificial learners to capture their different learning styles. There were four different levels of motivation change assigned to different artificial learners: \( x = 1 \) for fast motivation change (FMC), \( x = 0.3 \) for moderate (MMC), \( x = 0.1 \) for slow (SMC), and \( x = 0 \) for no change (NMC). Artificial learners could have one of eight pairs of knowledge change variables \((w_1, w_2)\): \((1, 0)\); \((U[0, 1], 0)\); \((U[0, 0.3], 0)\); \((1, U[0, 0.3])\); \((U[0, 1], U[0, 0.3])\); \((U[0, 0.3], U[0, 0.3])\); \((1, U[0, 1])\); \((U[0, 1], U[0, 1])\); \((1, 1)\), where \(U[0,y]\) means a random value generated from a uniform distribution between the two end points. For example, \((1, 0)\) indicates that only authoritative papers can fully increase a learner’s knowledge of the content of a paper they have ‘read’; \((1, 1)\) indicates that both authoritative and non-authoritative papers are equally weighted and can fully increase a learner’s knowledge of the content of the paper; and \((U[0, 0.3], U[0, 0.3])\) means that for both authoritative and non-authoritative papers, the knowledge increases somewhere between 0 and 0.3 after a paper is ‘read’ (randomly generated in this range for each paper). The various conditions were divided equally among the artificial learners. Each group of experiments was repeated thirty times for statistical analysis.
3.2.3 Experimental results and discussion

Table 1 shows the results of the experimentation. The value shown in each cell is the pair of numbers representing the percentage difference between content-based recommendation and the hybrid-CF technique in terms of average learner knowledge and motivation gained (or lost). A negative value indicates that the content-based recommendation technique is better (on the metric being measured) than hybrid-CF. And a positive value represents the reverse situation. For example, the pair value \((0.65; 2.93)\) represents that using hybrid-CF is 0.65% and 2.93% better than using content-based in terms of the average learner knowledge and motivation respectively. All results are checked by \(t\)-test for equal mean hypothesis (assuming different variance). A value in italics inside the table shows that the null hypothesis is not rejected (for \(\alpha = 0.05\)), or the difference between content-based and hybrid-CF is not statistically significant. If we exclude zero and italic values in Table 1, then there are 14 and 6 negative values for the difference of learner knowledge and motivation respectively, with the lowest value equal to \(-1.05\)% and \(-5.68\)% respectively. And there are 8 and 12 positive values for the difference of learner knowledge and motivation, with the highest value equals to 1.20% and 19.38%, respectively.

These results are not that conclusive, but we can say that using hybrid-CF overall results in a lower performance in terms of learner average knowledge. However, since hybrid-CF usually needs lower computational cost than content-based recommendation and the performance loss is not big, hybrid-CF is still promising in an e-learning system. The results also shed some light on individual differences, so that if we know something about the motivation and knowledge changes likely for a particular real learner, we may be able to choose the recommendation technique most appropriate for him or her.

A final note on the usefulness of computer simulation in these circumstances is in order. Computer simulation and artificial learners have long been valuable assets in shedding light on issues in intelligent tutoring systems [17, 19, 20]. Although a simulation can only model aspects of an environment with real learners, it can

<table>
<thead>
<tr>
<th>((w_1, w_2))</th>
<th>FMC</th>
<th>MMC</th>
<th>SMC</th>
<th>NMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>((1, 0))</td>
<td>(0.59; 2.77)</td>
<td>(-0.70; -0.06)</td>
<td>(-0.77; -0.42)</td>
<td>(-0.43; 0.00)</td>
</tr>
<tr>
<td>((U[0, 1], 0))</td>
<td>(0.98; 7.97)</td>
<td>(-0.28; 3.85)</td>
<td>(0.21; -0.32)</td>
<td>(0.54; 0.00)</td>
</tr>
<tr>
<td>((U[0, 0.3], 0))</td>
<td>(-0.47; 15.15)</td>
<td>(-0.52; 0.75)</td>
<td>(0.33; -5.42)</td>
<td>(1.09; 0.00)</td>
</tr>
<tr>
<td>((1, U[0, 0.3]))</td>
<td>(-0.57; 1.61)</td>
<td>(-1.05; -1.05)</td>
<td>(-0.76; -0.90)</td>
<td>(-0.29; 0.00)</td>
</tr>
<tr>
<td>((U[0, 1], U[0, 0.3]))</td>
<td>(0.30; 8.09)</td>
<td>(-0.44; 3.41)</td>
<td>(0.22; -0.01)</td>
<td>(0.69; 0.00)</td>
</tr>
<tr>
<td>((U[0, 0.3], U[0, 0.3]))</td>
<td>(-0.85; 19.38)</td>
<td>(-0.69; -0.19)</td>
<td>(0.06; -5.68)</td>
<td>(1.20; 0.00)</td>
</tr>
<tr>
<td>((1, U[0, 1]))</td>
<td>(-0.52; 1.13)</td>
<td>(-0.96; -0.8)</td>
<td>(-0.82; -0.84)</td>
<td>(-0.27; 0.00)</td>
</tr>
<tr>
<td>((U[0, 1], U[0, 1]))</td>
<td>(0.96; 7.36)</td>
<td>(-0.15; 4.68)</td>
<td>(0.16; -0.06)</td>
<td>(0.88; 0.00)</td>
</tr>
<tr>
<td>((1, 1))</td>
<td>(-0.34; 1.47)</td>
<td>(-0.69; -1.31)</td>
<td>(-0.47; -0.81)</td>
<td>(-0.43; 0.00)</td>
</tr>
</tbody>
</table>
provide a powerful tool for gaining insights in complex settings. Therefore, the simulation discussed here can provide guidance to our future work. In fact, we have designed a human subject study, which looks at some of the same issues we have studied here, focused on human rather than artificial learners. Details of this study can be found in [21]. Further studies are planned.

4 The ecological approach

The recommender system architecture and empirical studies we have discussed in this chapter illustrate the need to be sensitive to the unique characteristics of an e-learning domain, in particular the need to take into account pedagogical aspects, such as the learner’s knowledge state, learning objectives, and motivational level, and the need to provide an appropriate sequencing of the papers recommended to the learner. Strong learner modeling is necessary, including the modeling of learner characteristics and goals, as well as the capture and interpretation of learner feedback about the papers they have read. It is also important to understand the relationship of the papers to each other and to determine how they relate to these pedagogical issues. Various techniques have been discussed in this chapter including annotating papers with explicit metadata provided by experts like tutors, attaching to papers opinions and feedback provided by learners, and using clustering techniques to find clusters of like-minded learners or similar papers. Data mining techniques also have a strong role to play: usage mining to find patterns in learner interactions, text mining to help categorize the papers according to actual contents, and possibly even web mining if the paper recommender system were embedded in a larger e-learning environment that included links to web resources.

The research paper recommender system has been one of the core inspirations for the ecological approach, an architecture recently proposed by Mc Calla [22] for the design of learner centered, adaptive, reactive e-learning systems. Generalizing from the goal of recommending papers for one learner from a paper repository, in the ecological approach the basic e-learning environment is assumed to possibly have many learners, many tutors, a large number of learning objects (papers, web pages, online exercises, quizzes, etc.), and a number of different applications that support learners including learning object recommenders (similar to the paper recommender), help or helper finders, tutoring systems, and so on.

In the ecological approach the e-learning system keeps a learner model for each learner, tracking characteristics of the learner and information about the learner’s interactions with the learning objects each encounters. After a learner has interacted with a learning object, the learning object is ‘tagged’ with an instance of the learner model. The information in such a learner model instance can include:

- information about the learner, including cognitive, affective, and social characteristics and most importantly their goal(s) in accessing the content;
- information about the learner’s perspectives on the content itself, including the learner’s feedback on the content, the learner’s knowledge of the content
(as determined, for example, by a test administered during the learner’s inter-
actions with the learning object);
- information about how the learner interacted with the content, including observed
metrics such as dwell time, number of learner keystrokes, patterns of access;
- information about the technical context of use, including characteristics of the
learner’s software and hardware environment;
- information about the social context of use, including links to the learner model
instances attached to learning objects previously encountered by the learner.

Over time, each learning object thus slowly accumulates learner model instances
that collectively form a record of the experiences of all sorts of learners as they have
interacted with the learning object. The collected learner model instances can then
be ‘mined’ for patterns about how learners interacted with the learning object, for
example that learners whose knowledge has been evaluated as weak did not have
long dwell times, or that learners with certain cognitive characteristics did well.
The sequence of learner model instances for a particular learner forms a ‘learning
trail’ through the learning object repository, and this trail can also reveal interesting
patterns of success and failure for the learner.

To illustrate this diagrammatically, Fig. 3 shows a learning object repository
with six learning objects and four learners who have been interacting with the
learning objects. As can be seen, each learning object is annotated with the learning
model instances of these various learners, sometimes two different instances for a
particular learner attached to the same object, representing two different times that
this learner has interacted with the same learning object. Within these instances, of
course, would be all the information the e-learning system had about the learner
and their interactions with the learning object. The sequence of learning object
instances for a particular learner would also be captured.

\[\text{Figure 3: Learning object repository annotated with learner model instances.}\]
There are, of course, a potentially huge number of patterns that can be found when mining learner behavior in these learner model instances. The key to finding meaningful patterns is the purpose (in the sense of [23]) for which the patterns are sought. Each such purpose places its own particular constraints on what patterns are meaningful, how to look for these patterns, and how to use what these patterns reveal in order to achieve the purpose. Thus, if the purpose is to recommend a specific learning object to a particular learner, this may require comparing this learner to other learners on important characteristics and then looking at how similar learners have evaluated (or been evaluated on) the content (and, moreover, the characteristics considered to be important are themselves determined by the learner’s own goals). This is similar to the hybrid-CF technique that seems preferred for the paper recommender system discussed in Sections 2 and 3. On the other hand, if the purpose is to determine whether a learning object is now obsolete, this may require an examination of all learners’ evaluations of the content, trying to extract temporal patterns in the evaluations that show how recent learners like or dislike the content. The key point is that it is the purpose that determines what information to use and how it is to be used. An ideal for a real time e-learning system is that this determination be made actively (in the sense of [24]) at the time the purpose is invoked, so that no a priori interpretation needs to be given to the information; however, time constraints on executing the data mining algorithms may mitigate against such real time computation in many circumstances.

In sun, then the ecological approach promotes the notion that information gradually accumulates about learning objects, the information is about the use of the learning object by real learners, and this information is interpreted only in the context of end use. The approach is ecological because over time the system is populated with more and more information, and something like natural selection based on purposes determines what information is useful and what is not. The ecological approach, with its emphasis on end-use context thus has a focus on pragmatics, rather than semantics. Dealing with pragmatics issues is a key to the usability of any system, and pragmatics is likely to be a major focus of research as semantic web research starts to move beyond the current focus on content towards usability and understanding the social context.

One of the difficulties in the ecological approach is getting information about learners to stock the learner models. Fortunately, experience in all sorts of learning environments (electronic and otherwise) have proven learners to be more willing to provide information to systems that will help them to learn than they have been to be so open with standard application systems, aimed at commercial profit for others, especially if they think that such information will make the e-learning system more effective and responsive to their own needs. They are also more likely to be willing to be monitored and evaluated, including to allow the testing of their knowledge and the diagnosis of their misconceptions. A learner’s learning goals can be explicitly known because he or she will tell the e-learning system directly or they can be implicitly determined from the very fact that the learner is interacting with this particular e-learning system. For example a learner may want to learn about some subject, to find content relevant to a particular issue, to get help to
overcome an impasse, and so on. E-learning systems themselves can also have explicit pedagogical goals that can be known and used in doing the purpose-based computations.

There are many possible applications for the ecological approach in e-learning. In addition to learning object recommender systems, such as the paper recommender discussed in this chapter or the educational recommender system discussed in [25], the ecological approach could underlie the design of:

- a group formation tool, to suggest to the learner a group of other learners relevant to solving a particular task or learning about a particular subject;
- a help seeker, to find another learner who can help the learner solve a problem he or she has encountered (as in I-Help [26]);
- a reminder system, to keep a learner updated with new information, say from the web, that is relevant to the learner’s goals;
- an evaluation and assessment tool, to allow learners’ interactions with an e-learning system to be studied by instructional and cognitive scientists, in particular to look at the experiences of all learners or particular types of learners with some educational content;
- an end-use learning object annotation system, to automatically derive educational content annotations from pre-established ontologies based on the experiences of the actual users of the content, and that can be parameterized by end use variables such as type of learner, success/failure of the educational content for each type of learner. A variant of this is possibility is the ability to refine, modify, or change pre-assigned metadata based on inferences from end use;
- an ‘intelligent’ garbage collection system, to determine the on-going relevance of educational content and, if necessary, to suggest modifications or even that it be deleted as no longer being useful to learners (e.g. as discussed in [27]).

Most of these applications would involve a large amount of purpose-based data mining of the learner model instances attached to the learning objects in the learning object repository.

The ecological approach is still more of a research proposal than a proven architecture, and it stimulates the need to resolve many deep and interesting educational and computer science issues. In the ARIES Laboratory we are exploring some of these issues. Following up on the initial inspiration provided by the paper recommender system discussed in this chapter, we have a number of other projects underway that fit the ecological framework. In one of these projects we have developed the MUMS middleware [28] that allows the designers of the e-learning system to tailor modeling processes to look for particular kinds of patterns in raw data collected about learner interactions with learning objects in the repository. These patterns can then be fed into the higher-level applications that need information about the learners to make their decisions. We have collected detailed raw data about learner interactions with two fully online computer science courses (and the associated I-Help open peer forum used by learners to find help as they took the courses) over a couple of offerings of the courses. In one experiment with MUMS we then
created MUMS modelers to mine this data so as to inform a proof of concept ‘open modeling’ prototype system that displays information about learners’ progress to their instructors at several levels of time granularity [29].

We have also explored in [30] how to capture learner goals in ‘purpose hierarchies’ that provide ‘purpose clichés’ for doing active learner modeling. Such an approach to representing learner goals is highly useful, since the purpose cliché is not just a label but is a full-fledged procedure for actually using the goal information appropriately in the application. When this procedure is executed, all of the relevant information about the end-use context will be available and can be factored appropriately into the processing, and the purpose itself can make tough decisions about what is relevant or not to its goals.

Perhaps our most ambitious exploration of the ecological approach is within our University of Saskatchewan theme 3 of the LORNET project, a multi-pronged, multi-situated Canadian research project to develop sophisticated techniques for supporting learning object repositories [2]. In LORNET theme 3 we are exploring the application of artificial intelligence techniques to the design of e-learning tools to support learners interacting with a learning object repository. In particular we are looking at making learning objects into agents, and also adding into the learning object repository personal agents representing learners and tutors. These extended learning object repositories are thus distributed systems where the learners and tutors (through their proxy agents) and learning objects (now agents) all co-exist and carry out interactions with each other in support of the e-learning application and the learners and tutors using the application. Such large-scale distributed computation will allow the e-learning applications to be more reactive to the end-use context (since the agents will carry out their computations ‘just in time’ within this context). Such context-sensitivity is exactly the goal of the ecological approach and the active modeling perspective. Work is still relatively new on this vision for LORNET theme 3, although the MUMS system is an early spin-off.

5 Conclusion

In conclusion, many of the investigations underway in the ARIES Laboratory are exploring learner-centric e-learning systems, with a strong need for data mining of actual learner use of the systems. As shown in this chapter through the paper recommender system ‘case study’, this data mining must be able to make sense of learner behavior in context, preferably actively as the context evolves and changes. Knowledge of goals, both learner goals and the explicit or implicit pedagogical goals of the application doing the data mining, is critical to determining which patterns to look for. Knowledge of various pedagogical aspects of the learners, accessed through their learner models, provides further constraints on the data mining algorithms. The ecological approach outlined in the previous section may provide a unifying architecture in which to explore such active, context-dependent, data centered techniques for e-learning. But, there is much left to do before this hypothesis is proven.
Acknowledgments

We would like to acknowledge the Canadian Natural Sciences and Engineering Research Council for their financial support of this research, both through an on-going discovery grant to the second author and through their funding of the LORNET Network of Centres of Excellence. We also thank our colleagues at Hong Kong Polytechnic University and in the ARIES Laboratory at the University of Saskatchewan who have influenced our ideas substantially.

References


[29] Webster, A., Visual models of MUMS’ events. Report on Senior Course Project, Department of Computer Science, University of Saskatchewan, Saskatoon, SK, 2005.

Appendix A: Questionnaire used in the study in Section 3.1

Thank you for taking your time to participate in this short survey.

The purpose of this survey is to collect information about recommendations (research papers, tutorials, survey papers, magazine papers, resource links, etc.) for e-learning.

The scenario is as follows:

Assume that you are taking a graduate-level class where you need to read several papers (as what we usually did). For each topic taught in the class, you are required to read 2 or more reading materials (journal paper, workshop paper, etc.) recommended by the professor or your classmates.

Q1. Please number your preference in reading those materials if you are VERY interested in the topic being taught (please number with ‘1’ for the most preferred one and ‘4’ for the most disliked one):
   ( ) lecture note (power point slide) ( ) related magazine article
   ( ) conference paper (5 to 10 pages) ( ) journal paper (30 to 60 pages)

Q2. Please number your preference in reading those materials if you are NOT interested in the topic being taught (please number with ‘1’ for the most preferred one and ‘4’ for the most disliked one):
   ( ) lecture note (power point slide) ( ) related magazine article
   ( ) conference paper (5 to 10 pages) ( ) journal paper (30 to 60 pages)

Q3. Please number your preference in reading the same material with different presentations as follows (please number with ‘1’ for the most preferred one and ‘4’ for the most disliked one):
   ( ) paper with formal model only (theorem/equation)
   ( ) paper with algorithm only
   ( ) formal model + graphs/diagrams only ( ) algorithm + graphs/diagrams only

Q4. Assume there are two papers with the same topic, same technical level and same length. The first paper was written by a well-known author and the other was written by an unknown author. Which one will you read first? (please mark ‘X’ on your choice)
   ( ) Paper by well-known author ( ) Paper by unknown author

Q5. Assume there are two papers with the same topic (e.g. multi-agent planning) and the same technical level. The first paper was presented in a reputable international conference and the other was presented in an unknown local conference. Which one will you read first?
   ( ) Paper in a reputable conference ( ) Paper in an unknown conference

Q6. Assume there are two related papers written by the same author with the same topic and technical level. The first paper was presented in a conference in 1998 and the second one is the revised version with some improvements and was presented in a similar conference in 2002. Which one will you read first?
Q7. Assume your professor recommends three papers with the same technical level and the same topic. He/she asks you to read 2 papers only. Papers 1 and 2 solve the problem using method X. Paper 2 is a refined version of paper 1 with minor improvements. Paper 3 solves the problem by method Y (written by a different author and substantially different from X). Which two will you read?
(  ) Papers 1 and 2  (  ) Papers 1 and 3  (  ) Papers 2 and 3

Q8. Assume your professor recommends an important but uninteresting paper. Besides, he/she also recommends an interesting paper, whose work is based on the first one. How will you read them?
(  ) Read the interesting one and followed by the boring/important one
(  ) Read the boring/important one and then read the interesting one
(  ) Read the interesting one and then skim (read quickly) the boring/important one

Q9. Assume your professor recommends two review papers. Which one will you read?
(  ) Deep review paper with many technical aspects
(  ) Shallow review paper, yet with many interesting examples/illustrations/applications

Q10. Assume your professor recommends several important resources links (containing papers, conference presentations, tutorial slides, etc.) maintained by some 'big name' researchers in the area, which one of the following do you like?
(  ) just list those links and you want to explore them by yourself
(  ) you do not want to explore them by yourself, you hope to receive a more specific recommendation