

Peer-Based Intelligent Tutoring Systems: A Corpus-Oriented Approach

John Champaign

David R. Cheriton, School of Computer Science
University of Waterloo, Waterloo, ON, Canada
jchampai@uwaterloo.ca

Abstract. Our work takes as a starting point McCalla's proposed ecological approach for the design of peer-based intelligent tutoring systems and proposes three distinct directions for research. The first is to develop an algorithm for selecting appropriate content (learning objects) to present to a student, based on previous learning experiences of like-minded students. The second is to build on this research by also having students leaving explicit annotations on learning objects to convey refinements of their understanding to subsequent students; the challenge is to intelligently match students to those annotations that will be most beneficial for their tutoring. The third is to develop methods for intelligently extracting learning objects from a repository of knowledge, in a manner that may be customized to the needs of specific students. In order to develop our research we are exploring the specific application of assisting health care workers via peer-based intelligent tutoring.

1 Content Sequencing

Two central challenges in the design of intelligent tutoring systems are compiling the material for the lessons and determining the best methods to use, for the actual teaching of those lessons. We observe in particular that it is desirable to provide a framework for determining the material to be taught that does not rely on experts hand-coding all the lessons. Indeed, that particular approach presents considerable challenges in time and effort.

We are interested in techniques for bootstrapping the system in order to initiate peer-based learning and in developing robust methods for validating the models that are presented (including the technique of employing simulated students). Once the content is in place, our efforts will be aimed at refining our model in order to enable students to benefit the most from the learning that their peers are undergoing.

We have currently developed an algorithm for reasoning about the sequencing of content for students in a peer-based intelligent tutoring system inspired by McCalla's ecological approach[1]. We record with each learning object those students who experienced the object, together with their initial and final states of knowledge, and then use these interactions to reason about the most effective

lessons to show future students based on their similarity to previous students¹. As a result we are proposing a novel approach for peer-to-peer intelligent tutoring from repositories of learning objects.

We used simulated students to validate our content sequencing approach. Our motivation for performing this simulation was to validate that, in the experimental context, our approach leads to a higher average learning by the group of students than competing approaches. We added a modeling of the knowledge that each object is aimed at addressing (for example, an object in a first year computer science course may be aimed at addressing the knowledge of recursion). By abstracting all details from the intelligent tutoring system and the student, we defined a formula to simulate learning (Equation 1).

$$\Delta UK[j,k] = \frac{I[l,k]}{1 + (UK[j,k] - LOK[l,k])^2} \quad (1)$$

where UK is the user j 's understanding of knowledge k , I is the educational benefit (how much it increases or decreases a student's knowledge) of learning object l on knowledge k and LOK is the learning object l 's target level of instruction for knowledge k .

When running our algorithm in the simulation, each student would be presented with the learning object that was expected to bring the greatest increase in learning, determined by extracting those learning objects that had resulted in the greatest benefit for previous students considered to be at a similar level of understanding as the current students.²

Simulated students and content allowed us to avoid the expense of implementing and experimenting with an ITS and human students to see the impact of our approach in contrast with alternative approaches. In particular we contrasted our method with a baseline of randomly assigning students to learning objects and to a "look ahead" greedy approach where the learning was precalculated and used to make the best possible match. One variant we considered was a "simulated annealing" inspired approach, where greater randomness was used during the initial, exploratory phase of the algorithm, then less randomness was used once more information about learning objects had been obtained. We discovered that our approach showed a clear improvement over competing approaches and approached the ideal.

2 Annotation

To extend the basic evolutionary approach, we are particularly interested in exploring the use of student annotations, which would fit naturally with our proposed corpus-based design for the lesson base. Student annotations on learning objects would involve allowing students to leave short comments on lessons

¹ This is motivated by techniques from collaborative filtering recommenders.

² Each student in the simulation is modeled to have a current level of understanding for each possible knowledge area, a value from [0,1] reflecting an overall grade from 0 to 100.

they are interacting with (more than tags, this could be a question or commentary about what they're learning). Subsequent students would identify which annotations they found useful, which would then be intelligently shown to similar students. The idea behind this is allowing students to "collaborate" with one another but not in real time (or, at least, to allow the interactions of the student in the past to inform the interaction with the current student, which honours the ecological approach[1]). There will be a decision theoretic reasoning element to this, when "low quality" annotations should be shown as part of a dialogue involving high quality annotations, and some trust modeling in addition to student modeling similar to what we are advocating for Content Sequencing.

To date we have only developed preliminary steps towards an overall algorithm for reasoning about annotations. We have not yet explored how best to validate our approach.

3 Corpus-Based

We are interested in exploring the construction of the lesson base that forms the centrepiece of a peer-based intelligent tutoring system, and are concerned with facilitating the authoring of such a lesson base, through the mining of existing repositories of information. This stands in contrast to McCalla's ecological work[1], which assumes that learning objects are already created and available to the system. This would be especially useful for applications where large repositories of information already exist, possibly employing varied forms of media, that could be leveraged for the creation of an ITS. Towards this end we have been exploring scenarios applicable to peer-based home healthcare assistance for caregivers or patients (e.g. facilitating a healthcare decision by presenting a learning object that has been useful to others in the past). Working in conjunction with health care workers affiliated with our hSITE (Healthcare Support through Information Technology Enhancements) project (an NSERC Strategic Research Network), our aim is to design a system that will be of some benefit in actual healthcare environments.

For future work, we are interested in refining learning objects created for a variety of purposes (e.g. book chapters, instructional videos or research papers) by combining the original learning object with a student model. The aim is to separate the most important information in the object, potentially breaking it into multiple, more targeted learning objects based on what the student would find most relevant. A second path for future work is to identify learning objects currently missing from an existing ITS which could be deployed for pedagogical benefit.

4 Plan for Research

Our plan for research is to first of all explore in greater depth the problem of presenting annotations of learning objects to new students. The primary challenge is determining which annotations should be presented to each student. A

reasonable first step would be to make use of our existing algorithm for content sequencing which proposes specific learning objects to students based on the benefits derived from these learning objects by students at a similar level of knowledge. We could then troubleshoot this algorithm, to see where the requirements for delivering effective annotations to students differ from simply presenting the most appropriate learning objects. One element that would be missing is an interpretation of the value of the annotations, recorded by previous students. As such, experiences of students who were not as similar to the current student would still be of use; likewise, annotations offered by similar students that were considered to be less valuable need not be included. Our aim would be to develop a rich user modeling framework to result in the most effective system for presenting annotations.

Once we have made some progress on the topic of annotations, we would return to our current framework for content sequencing, to determine whether additional improvements could be introduced to this model. In particular, we may have some insights into how to model the similarity between students in a somewhat richer manner.

The third problem of extracting appropriate learning objects from a corpus would be our next focus of study. We expect to receive valuable feedback from the health researchers with whom we have been corresponding, in order to work with a series of motivating examples, towards algorithms for delivering a set of refined learning objects from an existing repository of knowledge. In so doing, we will have made some modest first steps as well for providing an intelligent tutoring system of use as a health informatics application.

A final thread for our research will be to investigate additional methods for verifying the value of our models. One direction will be to develop more effective methods for simulating students. Another will be to consider user studies with real users; we would explore in particular the value of qualitative approaches to experimentation. We expect this research to be an important contribution to the subfield of peer-based intelligent tutoring systems (e.g. [3]). We would be providing a concrete realization of a framework for intelligent tutoring that honours McCalla's ecological approach[1]. We would also offer some contributions to other researchers interested in exploring the use of simulated students as a replacement for real students (e.g. [2]) as a method for verifying the value of an ITS.

References

1. McCalla, G.: The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information About Learners. *Journal of Interactive Media in Education: Special Issue on the Educational Semantic Web* 7, 1–23 (2004)
2. VanLehn, K., Ohlsson, S., Nason, R.: Applications of simulated students: An exploration. *Journal of Artificial Intelligence in Education* 5, 135–175 (1996)
3. Vassileva, J.: Toward social learning environments. *IEEE Transactions on Learning Technology* 1(4), 199–214 (2008)