Personalised intelligent tutoring for digital libraries

Clodagh Moriarty Nicholas Kushmerick Barry Smyth
Smart Media Institute, Department of Computer Science, University College Dublin
{clodagh.moriarty, nick, barry.smyth}@ucd.ie

Abstract

Computer-based training is a fast-growing multi-billion dollar industry. The possibilities for systems that offer personalised training or tutoring, that dynamically adapt to the training needs of individual students, are immense. This not only means the personalisation of training content but perhaps even the personalisation of exams and student evaluations. In this paper we focus on ways of personalising multiple-choice exams and describe a technique for predicting the exam answers for individual students based on their previous exam history. We describe an evaluation of a collaborative filtering prediction system and demonstrate that accurate predictions can be achieved with limited profiling information.

1 Introduction

If we view a digital library to be a collection of services and information objects for users available via digital means, then a primary goal must be to encourage communication and learning among a wide range of communities. Online tutoring systems are an important component of any large-scale digital library project and commercially the demand for online tutoring systems drives a multi-billion dollar industry, which is growing fast. We believe that the potential benefits of personalisation in the context of online learning are staggering and the the development of fully personalised training systems has the potential to revolutionise the educational industry. Imagine an online training system that is capable of responding to the precise learning needs of an individual student? By adapting to the strengths and weaknesses of each student, such a system could dynamically plan personalised lessons, providing educational support where necessary and training challenges where appropriate, all on a case by case basis. Such an educational environment has the potential to accelerate the learning process by providing a more adaptive and enjoyable learning environment.

In this paper we focus on ways of personalising online exams, specifically, multiple-choice exams. Our central contribution is a prediction system that is capable of accurately predicting the answer that a student will provide to a specific question. We do not explore the pedagogical implications of this ability, but it is apparent that the ability to predict student performance offers significant opportunities for teachers, because they use such information to create a personalised training environment.

The key idea in our approach is to use collaborative filtering techniques to predict student performance in a multiple-choice examination. Collaborative filtering evolved from the field of information retrieval and makes predictions by correlating the experiences of similar users over a given information domain. Collaborative filtering systems collect user ratings for items in a domain, and match people who share similar tastes. These similar individuals act as recommendation partners for the target user. For example, a collaborative filtering system for recommending books (a la Amazon) would make suggestions to a target user from the preferences of similar users, where similarity is estimated by comparing the rating history of the target user to other users in the population. Collaborative filtering is a complementary recommender technique to more traditional content-based filtering methods, which make judgements on the basis of domain knowledge - for example, a content-based book recommender requires categorical information about books. All in all, the key advantage of collaborative filtering techniques is that they obviate the need for domain information, such as book descriptions, relying on shallow behavioural information, such as rating histories, that can be automatically gathered over time.

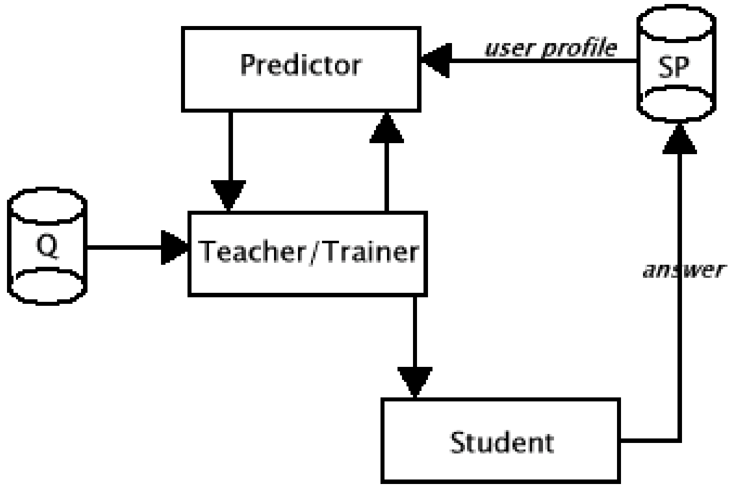

Figure 1: Our model of the online tutoring environment.

The collaborative filtering process was first automated in the GroupLens [5] project, which generated personalised predictions for Usenet news articles. Filtering was performed using a nearest-neighbours correlation algorithm. The Ringo music recommender [6] system made personalised recommendations using what they referred to as social information filtering, where the recommendations included a measure of confidence. A hybrid system that combines content-based and collaborative filtering recommendation techniques is FAB [1]. Its collection agent searched the web offline for interesting pages, and then considered recommendations based on profiles of all other users in the system. Another successful hybrid recommendation system combining these techniques is PTV [8], a personalised TV listings service.

Looking at the area of intelligent tutoring systems we find a vibrant research field where extensive studies have been carried out. However, the majority of existing projects in this area have placed a lot of emphasis on developing a deep model of student ability, e.g. how students think, how they perform in given situations, etc. EUROHELP [3] was devised to provide tools and methods for developing Intelligent Help Systems. A major portion of the project involved formulating psychological models of students. InterBook [2] provided a stereotype user model, which represented levels of a user's knowledge of every domain concept. This model was modified as the user moved through the information space. Other ITS projects used certain criteria to form a model of student ability, e.g. MANIC [9], an online courseware system for transforming existing video-taped courses. A student model was used to help guide a student through course material, to prefetch relevant course material and to provide adaptive quizzes. This student model was determined by heuristics such as which slides the student has seen and quizzes he has taken.

Our work aims to eliminate the need of a deep model of student ability, or piecemeal heuristics in order to predict student ability, by taking advantage of the collaborative filtering method. In this way we can make a prediction about how a student will answer a question without any understanding of the question's semantics or its answers. All that is required is historical information indicating how other students have answered this question in the past.

Our model of the online tutoring environment is depicted in Fig. 1. We assume that the Teacher or Trainer has developed a corpus of Questions, stored in a database Q. Students are given questions from Q, and every answer is recorded in a Student Profile database SP. Our research focus concerns the Predictor element of this model. The predictor element uses information from SP to try to predict what answer a particular student s_i will give to a particular question q_j.

Note that the Teacher/Trainer controls the educational process (e.g., which questions are posed to which students), not the Predictor. Nevertheless, it seems apparent that an accurate Predictor could be useful to a Teacher/Trainer in numerous ways. For example, such information could be used to give students easier

	questions	students s_1	s_2	s_3	s_4
q_1	Which of the following is an illegal 8086 instruction: (a) ret 2 (b) push a1 (c) aDd bx, 25000 (d) and ax, dx	a	a	b	a
q_2	An OR gate generates a low output when (a) any one of its inputs is low (b) all of its inputs are high (c) all of its inputs are low (d) power fails x	b	b	c	b
q_3	The capacity of a DVD disk is around (a) 100 Mb (b) 650 Mb (c) 1.4 Gb (d) 1024 bytes	c	d	c	c
q_4	The 8-bit two's complement number 1111 1111 represents: (a) 255, (b) -255, (c) -127, (d) -1	a	d	a	?

Figure 2: A simple example illustrating our collaborative-filtering approach to personalised tutoring.

questions initially so as to build up their confidence, or to recommend additional study materials tailored to their specific weaknesses. We do not focus on these pedagogical issues, but merely provide a powerful "enabling technology" for personalised intelligent tutoring systems.

2 Predicting student answers

Our approach to predicting student performance is based on collaborative filtering (CF) techniques. Our approach thus contrasts with traditional research on intelligent tutoring which involves construct a deep psychological/pedagogical model of the students abilities or cognitive processes [3, 9, 2].

Conceptually, our Student Profile database SP is a matrix, where rows correspond to questions, columns correspond to students, and each non-empty cell consists of a student's answer for a given question.

For a given student s_i in the dataset, we wish to predict how s_i will answer questions they have not yet been presented with. The questions already answered by the student constitute the training set. The questions not yet viewed by the student constitute the test set. We use a standard nearest-neighbour approach to CF: we identify a group of students with profiles that are highly similar to s_i. We then assume that, since s_i and this neighbourhood have expressed similar knowledge in the past, it is likely that they will continue to do so in the future. We therefore use the profiles of the nearest neighbours to predict how s_i will answer a question she has not yet answered.

To complete the presentation of our CF approach to personalised tutoring, we must specify in detail how we measure student profile similarity, and how we use a set of profiles to predict an answer to a given question. We measure profile similarity as the fraction of questions for which two students gave identical answers. Having established a neighbourhood of students with the highest profile overlap to some target student, we predict the target student's answer by simply taking a weighted vote of the answers for each student in the neighbourhood, where the weights are the similarities to the target student.

We conclude this description of our approach with a simple example. Fig. 2 shows a small student-question matrix. Suppose we wish to predict how current student s_4 will answer question q_4. We give actual questions used in our experiments described in Sec. 3 just for the sake of concreteness. Like all collaborative filtering techniques, our personalised tutoring algorithm does not reason explicitly with semantic information regarding the questions to be predicted.

Since student s_4 answered the first three questions, we take q_1, q_2 and q_3 as our training set, and q_4 as our test set. If we use the $k = 3$ nearest neighbours, then s_1, s_2 and s_3 constitute the neighbourhood of most similar students. We assign each of these nearest neighbours a weighted vote. In our example, student s_1 votes with a weight of 1, since s_1 and s_4 answered identically for questions that both answered. Student s_2 votes with a weight of 0.667, since she and s_1 gave the same answer for two out of three training questions. Student s_3 votes with a weight of 0.333. Answer (a) therefore gets a vote of $1 + 0.333 = 1.333$, answers (b) and (c) get no votes, and answer (d) gets a vote of 0.667. Our algorithm therefore predicts that student s_4 will answer (a) to question q_4.

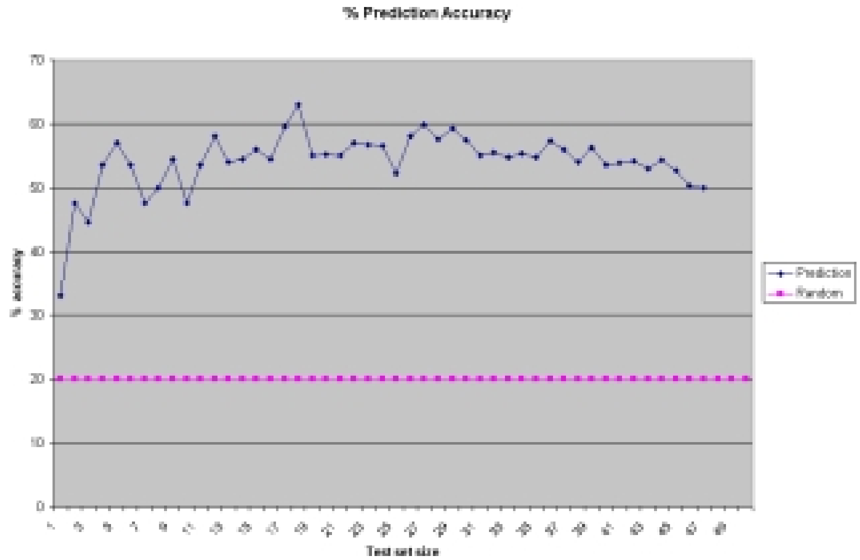

Figure 3: Prediction accuracy as a function of training set size.

We evaluated our approach on data gathered from 263 students who sat a written examination of 50 multiple-choice questions, and their respective answers. The paper was set for first year science students who took a course entitled 'Introduction to Information Technology' at University College Dublin and who sat the end-of-year examination in Summer 2000. All students sat the identical examination. Points were deducted for incorrect answers, so students had an incentive to skip questions of which they were unsure. On average, students chose to skip ten questions each.

2.1 Prediction accuracy

Our first experiment measures the accuracy of our prediction algorithm. We divide our dataset into a training set and test set of varying sizes. The training set consists of N questions, where $1 \leq N < 50$, and the remaining $50 - N$ questions constitute the test set. As described earlier, we use as student s_i's profile her answers to the training set questions, and we then find the set of k students whose profiles are most similar to s_i. We varied k from three to seven; in this paper we report results for $k = 5$.

Armed with the k similar students, their respective weighted votes, and s_i's profile, we predict how s_i would perform for all questions in the test set. For each of the five possible responses to a given question q_j (namely a multiple choice option of (a), (b), (c), (d), or the student's choice to skip the question entirely), we assign a likelihood value based on the weighted vote of each of our k similar students. The response with the highest value is the prediction for how s_i will answer question q_j.

Fig. 3 summarizes prediction accuracy obtained for our dataset, as a function of the training set size. We observe that our predictions are reasonably accurate, even when trained on only very small samples of previous answers. For example, when given just five previous answers, we can predict the student's answer 57% of the time. As a baseline, note that guessing randomly has accuracy of just 20%.

2.2 Student diversity

One of the challenges to collaborative filtering is that profile matrix is often very sparse. For example, in the MovieLens CF system, 94% of the entries are empty [7]. In contrast, only 19% of our matrix is empty. It might be argued, therefore, that we can successfully predict student answers just because our task is much easier than in most collaborative filtering settings.

To address the possibility, our second experiment measured the inherent difficulty of our task, in terms of the inherent diversity in our student population.

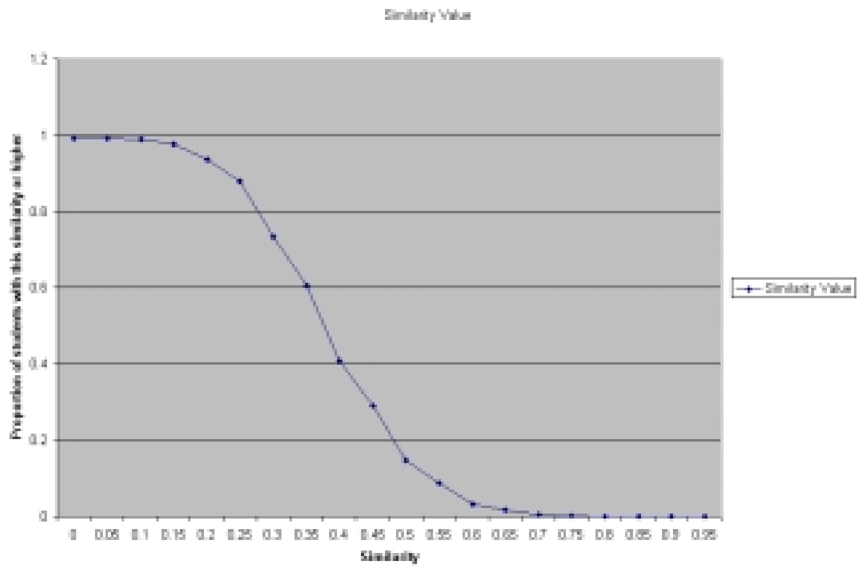

Figure 4: Student Diversity.

Fig. 4 summarizes student diversity values for our dataset. The graph plots the percentage of pairs of students with similarity greater than X, versus similarity of X. We observe that while students do not behave in a wholly unique manner, the majority of students fall within the 15-60% similarity range. Only a nominal amount (0.01%) of students have a similarity value of 70% or over.

3 Conclusions

The current success of online training and tutoring systems represents the tip of the iceberg in terms of what the future holds. Recent advances in personalisation technology now mean that it is possible to automatically adapt a given information environment for the preferences and needs of an individual user. This also holds true for online training environments. By using personalisation techniques we can adapt a training environment in line with the precise training goals, and educational history, of a given user in order to accelerate the learning process.

In this paper we have focused our attention on the personalisation of multiple-choice exams, an important component in most online tutoring systems. We have argued that the ability to accurately predict the likely answer that a user will provide to a specific question, can help teachers and trainers to formulate exams that are personalised for individual students. We have described a prediction system that, by using collaborative filtering techniques, obviates the need for deep student models or expensive domain knowledge. Instead, student performance is modelled using shallow information about a student's past exam performance, and the performance of a population of students is leveraged in order to make accurate predictions for an individual case.

Preliminary experiments show that our prediction mechanism produces accurate predictions of student performance even in situations where only limited historical information is available. Our current work is focusing on ways of improving the collaborative prediction technique. For example, we are currently evaluating a number of different user similarity models and prediction ranking techniques in the hope that further benefits will accrue. In the long term we will investigate how this prediction capability can be used within the context of a larger online tutoring system. For example, precisely how can a trainer use prediction information to formulate personalised multiple-choice exams?

References

[1] Marko Balabanovic and Yoav Shoham. Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66-72, March 1997.

[2] Peter Brusilovsky and John Eklund. A Study of User Model Based Link Annotation in Educational Hypermedia. *Journal of Universal Computer Science*. 4, 4. p.429-448. Springer Science Online.

[3] Joost Breuker, editor. *EUROHELP: Developing Intelligent Help Systems*. EC, Copenhagen, 1990. The final report.

[4] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61-70, December 1992.

[5] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3):77-87, March 1997.

[6] Upendra Shardanand and Patti Maes. Social information filtering: Algorithms for automating "word of mouth". *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, p.210-217, 1995.

[7] Badrul Sarwar, George Karypis, Joseph Konstan and John Riedl. Analysis of recommendation algorithms for e-commerce. *ACM Conference on Electronic Commerce*, p.158-167, 2000.

[8] Barry Smyth and Paul Cotter. Sky's the limit: A personalised TV listings service for the digital age. *Proceedings of the 19th SGES International Conference on Knowledge-Based and Applied Artificial Intelligence (ES 99)*, Cambridge, UK, 1999.

[9] Mia Stern, Beverly Park Woolf and James F. Kurose. Intelligence on the Web? *Proceedings of the 8th World Conference of the AIED Society*. Kobe, Japan, August, 1997.