To meet the challenges stated above my thesis combines computational models, algorithms and empirical methodologies. It benefits computer science by designing user modeling algorithms, incentive design strategies and adaptation methods for these exciting new domains. The efficacy, robustness and scalability of my approach is evaluated in several ways: First, by comparing the performance of the modeling, incentive design and adaptation approaches to the state-of-the-art approaches, and where applicable, to a gold standard. Second, by showing the generalizability of the techniques to several e-learning and citizen science domains. Lastly, by assessing users' performance using the tools I will develop in real world scenarios.

To evaluate my proposed algorithms, I am collaborating with the designers of e-learning systems and citizen science applications with access to large-scale data sets of people's interactions with task systems, as well as the possibility to conduct intervention experiments in both types of domains. Specifically, I am cooperating both with one of the leading e-learning companies in Israel, and with Zooniverse, one of the leading citizen science platforms in the world. These collaborations have already provided me with opportunities to conduct large scale empirical evaluations, saving significant overhead of time and effort.

My preliminary results include a new algorithm for modeling students and for personalizing educational content to them that combines collaborative filtering algorithms with social choice theory. A paper based on this research [5] was presented at the the Educational Data Mining Conference 2014 and won best student paper award.

In addition I've worked to develop a general methodology for disengagement reduction using interventions in citizen science environments.This work was based on the analysis of two years of participation data in 16 citizen science projects. This methodology included: (1) Surveys to reveal the motivations that drive users' participation in the different projects; (2) Identifying cohorts based on the survey results and the participation data; (3) Designing an intervention strategy that targets specific cohorts and addresses the motivational issues revealed in the survey; (4) Analyzing the efficacy of this strategy over time, according to performance and persistence measures.

Applying the methodology revealed that disengagement was triggered by life distractions, classification anxiety, and boredom. I've identified target communities for the intervention and designed interventions in the form of emails that directly addressed underlying issues uncovered by the survey. The methodology was shown to successfully promote re-engagement of users across 16 different citizen science projects. I am now working on building computational models for interventions based on these findings.

2. THE EDURANK ALGORITHM

I have developed a novel algorithm for sequencing content in e-learning systems that directly creates a "difficulty ranking" over new questions. My approach is based on collaborative filtering [1], which generates a difficulty ranking over a set of questions for a target student by aggregating the known difficulty rankings over questions solved by other, similar students. The similarity of other students to the target student is measured by their grades on common past question, the number of retries for each question, and other features. Unlike other uses of collaborative filtering in education, this approach directly generates a difficulty ranking over the test questions, without predicting students' performance directly on these questions, which may be prone to error.[2]

The algorithm, called EduRank, weighs the contribution of these students using measures from the information retrieval literature. It allows for partial overlap between the difficulty rankings of a neighboring student and the target student, making it especially suitable for e-learning systems where students differ in which questions they solve. The algorithm extends a prior approach for ranking items in recommendation systems [4], which was not evaluated on educational data, in two ways: First, by using social choice theory to combine the difficulty rankings of similar students and produce the best difficulty ranking for the target student. Second, EduRank penalizes disagreements in high positions in the difficulty ranking more strongly than low positions, under the assumption that errors made in ranking more difficult questions are more detrimental to students than errors made in ranking of easier questions.

I evaluated EduRank on two large real world data sets containing tens of thousands of students and about a million records. I compared the performance of EduRank to a variety of personalization methods from the literature, including the prior approach mentioned above as well as other popular collaborative filtering approaches such as matrix factorization and memory-based K nearest neighbors. I also compared EduRank to a (non-personalized) ranking created by a domain expert. EduRank significantly outperformed all other approaches when comparing the outputted difficulty rankings to a gold standard.

REFERENCES

[1] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.

[2] A. Irwin. *Citizen science: A study of people, expertise and sustainable development.* Psychology Press, 1995.

[3] E. K. Kahiigi, L. Ekenberg, H. Hansson, F. Tusubira, and M. Danielson. Exploring the e-learning state of art. *Electronic Journal of e-learning*, 6(2), 2008.

[4] N. N. Liu and Q. Yang. Eigenrank: a ranking-oriented approach to collaborative filtering. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 83–90. ACM, 2008.

[5] A. Segal, Z. Katzir, K. Gal, G. Shani, and B. Shapira. Edurank: A collaborative filtering approach to personalization in e-learning. In *Educational Data Mining 2014*, 2014.

[6] A. Toscher and M. Jahrer. Collaborative filtering applied to educational data mining. *KDD Cup*, 2010.

[2]To illustrate, in the KDD cup 2010, the best preforming grade prediction algorithms exhibited prediction errors of about 28% [6]