An Adaptive Learning System to Increase STEM Interest

Abstract
According to a study from Rutgers University, America's most talented students and professionals are losing interest in science, technology, engineering, and mathematics (STEM). The study revealed that for students who scored in the top 20% on the SAT or ACT math test, recruitment rates to college STEM programs fell sharply from 28% in the 1990s to 14% in the 2000s[1]. If this continued today, then the recruitment rate would be below 5%. We have launched a joint project between our university and the local Children's Services Council (CSC) on increasing middle school students' interest in STEM programs. The project is based on the concept of an adaptive learning system as outlined in a recent briefing from the U.S. Department of Education on enhancing teaching and learning through educational data mining and learning analytics[2]. We have developed a plan for a learning system, which can measure and track middle school students' interest in STEM careers as they participate in an after-school course on Android App development. The intent is to determine which part of the App will be used more by the students and their social community, how often, and how that might lead to a richer experience as measured by the number of modifications and sophistication of the resulting App. Both qualitative and quantitative metrics will be measured. We will explore the following hypotheses in a case-control study: Students in the case (the intervention) group will (1) show higher propensity for STEM careers; (2) spend more time in personalizing their App and in forming larger social group; and (3) use more STEM-oriented and pro-STEM keywords in their blogs.

Background
Tai et al. found that young adolescents who anticipated having a profession in science were more likely to graduate with a science degree, highlighting the significance of early inspiration[3]. The researchers investigated if eight-grade students who stated that they anticipated joining a science-related profession by age 30 received baccalaureate degrees in science-related careers at greater proportions than students who did not have this anticipation. They studied students in the U.S. from 1988 to 2000 and measured for variations in academic accomplishment, academic traits and students' and their parents' demographics. The most noteworthy result was that approximately half of the students pursued their anticipated profession. On the other hand, a third of the students that stated a non-science profession changed to science.

Urness, T., & Manley, E. D. found that many STEM subjects are presented to students in scheduled courses throughout their middle school and high school careers[4]. Students start to gain interests in these fields early on via normal coursework. Computer science is not usually included in such a methodical way, so fostering interest at an early age in the subject is a greater hurdle. Some concentrated summer camps, such as Alice camps for middle school students, have track records for promoting increased interest in computer science[5].

Our extended group has taught a smart phone App development course to high achieving high school students. Students were able to take the course tuition free because of their high school GPA of 3.0 or above, and rising junior/senior status. Students have come from 9 different high schools in local counties, and different backgrounds. Students formed groups of three, with selfidentified roles in project management, arts, and engineering, and developed smart phone apps during an intense three-week summer course. Two professors, one each from engineering and multimedia/digital art jointly taught the students. A group of senior undergraduate students helped the high school students develop the apps. A group of volunteer judges comprised of engineering professionals, multi-media artists, and local educational professionals have evaluated their work in team presentations at the end of this period. The goal here was initially to build fun and game apps with a social and/or educational message[6]. However, due to feedback from the volunteer judges, a more nuanced objective arose out of that: a focus on community-oriented apps (such as for urban planning and the science museum), and now for early empowerment of middle school students (as detailed here).

A middle school teacher, who is also a STEM coordinator for CSC, and her middle school students worked together to define a set of Apps for empowerment. A group of our undergraduate students undertook to develop 7 apps drawn from this set of Apps. These undergraduate students met with middle school students a few times and identified the specifics of these Apps that would be appealing and useful to them. The final Apps used a combination of two programming languages ('Processing' for software elements that can be personalized by middle school students, and 'Java' for more hard-core programming). Two forms of this app will be made available during summer 2015, one each to the case-control groups. The former group will receive an App that can be personalized via programming, not only for aesthetics but also for functionality, while the other will receive a non-customizable, but fully functional, version. We have a Github open source site that documents fully these empowerment apps. More will be incorporated in the final paper. We have chosen the highest ranked app, 'City Ville', for further exploration in our pilot research effort. Our ultimate intent is to improve these apps during this summer and make a larger portfolio of Apps available for use by middle school students in the longer run. The City Ville app will be used for initial exploration and build of the infrastructure.

Method

We will recruit approximately 60 middle school students and form intervention (case) and control groups to use the CityVille app either as a customizable or non-customizable App, respectively. This app was developed for empowerment of middle school students by undergrad students at our university in collaboration with a group of middle school students. Some of the features the CityVille app has are interactive Google Maps view, ability to post city events, and report neighborhood safety information[6]. The intervention group will personalize this programmable social app, while the control group will use a nonprogrammable version of the same app. We will collect usage (click) data and keyword analysis from their blog postings regarding their experience during the usage of the CityVille app. We will also conduct pre- and post- surveys to identify any changes in their interest in STEM.

Our initial intent was to collect their blog posts at Parse.com and Google which also provide good library support for data analytics. However, we realized that students may prefer more mainstream entities (such as Twitter, Facebook, and Instagram) for their communication and collaboration. We will narrow down to one or more social networking sites, as determined from students' pre-survey answers regarding their choice of social networking sites. We will then integrate the CityVille app with these social networking sites to make it easier for students to post their blogs. We will use semantic web and data mining technologies to analyze the blog posts, to help measure the students' progress and the overall effectiveness of the program. In a semester duration project, utilizing a collaborative system, students can produce a big quantity of data. This data has useful information that could assist in evaluating student effectiveness and in discovering issues ahead of time to make appropriate changes. Still, the raw data cannot be utilized as is by teachers and students because of sheer amount and specifics[7]. This is where machine learning and other data analytics techniques are needed.

Perera et al.[7, 8] found clustering to be suitable, exposing noteworthy patterns characterizing the activities of the groups and specific students. Clustering lets one utilize several attributes to classify alike groups in an unsupervised manner. Furthermore, it gives one the option to analyze the data for a single student and subsequently to study the work of a group. As the primary clustering algorithm, we selected the k-means. This algorithm is widely preferred, applicable and reasonably proficient[9, 10]. The most difficult issue in using the algorithm is choosing the attribute. The efficiency of clustering algorithms is very responsive to the sophistication of the attributes[7, 8]. We will be looking for attributes that indicate the development of STEM skills such as leadership, innovation, creative problem solving and the ability to conduct research independently. Since part of the purpose of this study is to build the students' confidence in their skills, we will also analyze their blog post for attributes indicating self-esteem and social anxiety issues. Most search tools will list every occurrence of any of the terms that we specify, but a semantic search engine has the intelligence to locate attributes that are relevant to our need. Our semantic search tool will examine student blog posts to not only look for the terms that we specify, but also for related terms. The tool is therefore able to discern which data are most pertinent to the subject that we are interested in. It can even suggest additional attributes for us to explore. We have already created and demonstrated a semantic search tool for finding research data on the treatment of diabetes[6]. Our tools will help us retrieve particular attributes from our program participants' blog posts.

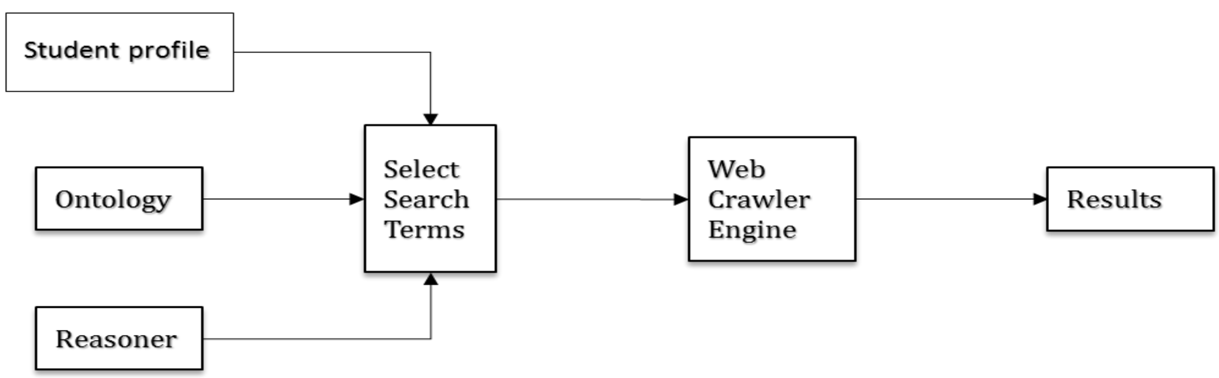

Figure 1 - Elements of search methodology

Our search methodology is depicted in Figure 1. Ontology is a graph of the relationships between terms that are relevant to a particular topic. A reasoner is a program, which interprets these relationships to find terms related to a specified term. We will create ontology for the CityVille app, and STEM-oriented and pro-STEM keywords. We will select search terms from the ontology. The reasoner will add whatever terms it may find in the ontology that are related to the terms we selected. The web crawler engine will then use the terms to search exhaustively through students' blog posts. The web crawler engine will generate search results, and the reasoner sorts them by relevance to our search needs. Thus, our methodology can obtain valuable results that might otherwise be missed, discard results, which are not relevant to our needs, and rank the results in order of interest to us[6].

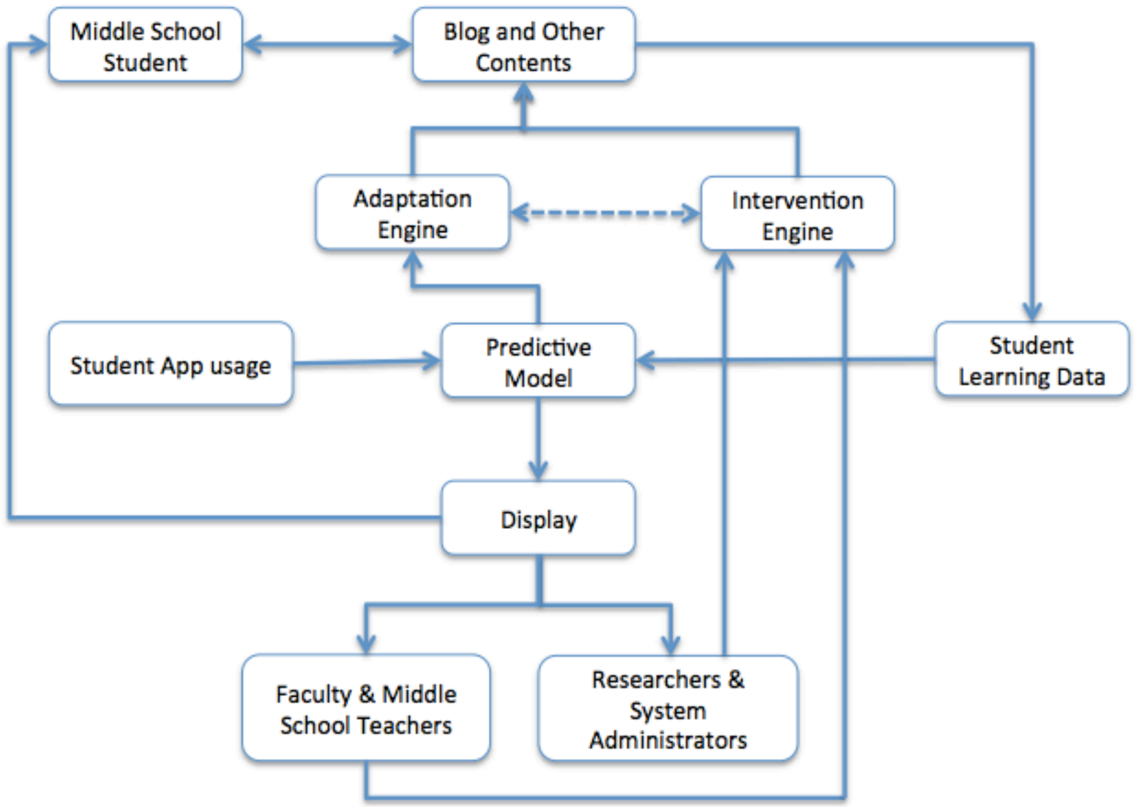

Figure 2 - U.S. Department of Education's Adaptive Learning System

Figure 2 shows the U.S. Department of Education's proposed exemplary learning system with six modules, which are content management, student learning data, predictive model, dashboard, adaptation engine, and intervention engine. Students will create inputs when working with the content management module. The inputs are time-stamped and prepared as needed and stored in the student learning database conforming to a specific form. The predictive model gets data for analysis from both the student-learning database and the student app usage. At this time, diverse data mining and analytics tools and models might be utilized contingent on the goal of the analysis. Once the analysis is finished, the results are utilized via the adaptation engine to modify what should be performed for a specific student. The content management will provide these modified lessons and teaching approaches to the student. The outcomes will be displayed to the dashboard, and, in the final step in this model, different users of the system will look at the reports for feedback and reply (via the intervention engine) in manners suitable for their responsibility[2]. Our long term objective is to integrate the flows in the two figures so it is useful for individual students. However, Figure 2 provides us guidance in addressing our exploratory research objectives.

Our project's time line is as follows: In Spring 2015, we will build our infrastructure and evaluate it with a small group of middle school students. We will measure students' use of the CityVille app, and perform data analytics on app metrics and semantic analysis of students' blog contents. In Summer 2015, we will gather usage and semantic data and perform data analytics for a new group of middle school students in a case-control study. We will collaborate with the local school county educational system/ local CSC. In Fall 2015, we will analyze the data and evaluate our hypotheses listed above. We will provide updates in the final paper. This is a long term study. This paper will be restricted to the initial build up of the technology and research infrastructure, as relevant to a doctoral dissertation in engineering.

Results

One can anticipate that students will have distinct patterns of activity different from those of others. For example, a student with leadership skill will seem more active. The lack of this pattern would allow us to postulate that the leadership skills are missing. We also anticipate that students with better STEM skills will display greater level of communication than the students with less STEM skills[7]. We will discuss with the teaching staff and take into account student's final grades in the course before characterizing student's activity patterns. Appropriate approval from IRBs (the institutional research boards) and consent from parents, students, and teachers will be sought.

We will collect the following data:

- App usage and blog data
- Ratings/grades of assignments and projects completed by students in the program
- Answers and analysis of multiple-choice surveys given to help measure their progress and the quality of the program
- Student and other consent forms, as per the IRB-approved protocol
- Interaction, evaluations and feedback from middle school students, teachers, mentors and mentees
- Psychometric evaluations of middle school students in the program

This raw data will be used to provide the following outcomes (tentative):

- Data analytics: predictive model to identify patterns of STEM interest and STEM attrition (absences, non-cooperative team, grades, timeliness, positive posts, etc.).
- Mentoring role: Rules for best practices for mentoring (number of visits, surveys, peer support, etc.) and migrating students towards STEM interest.
- Student feedback: provide alert on immediate objectives (submissions and other goals) and feedback on their progress trajectory relative to students who stayed highly interested in STEM with similar characteristics (done with clustering algorithms).

Discussion

It has taken us a significant amount of time to reach this point. Much of the infrastructure built had to be built as part of the teaching load of professors involved. Involvement of teaching administrators and coordinators has helped us find the right partners who are passionate about making a difference. We thus feel that we are finally on the right path and useful results will ensue. Briefly, we will use a socially relevant app that can be customized by middle school students. We anticipate that students will take more interest in programming, aesthetics and data collection (not for our research purposes, but for interactions of their social work) if there are tangible benefits to them and their social group. We will work with local STEM program coordinator to address the logistics of study implementation. We plan to include both students from STEM and non-STEM programs in the case and control groups, totaling at least 60 students. All these students will be provided with a smart phone by us for the duration of the project. Both groups will be informed that we are collecting usage data of their phone, but we may be allowed by IRB not to mention the reasons for data collection (to avoid any biases in their behavior). We will give a survey ahead of time to determine which social networking site will be used for blogging. In the survey, we will include questions regarding students' life style, awareness of job opportunities, and family and friends' STEM background. Also, we will utilize social networking tool to find students' key influencers. A group of professors from engineering, education, and data sciences, along with education leaders from the school county, are overseeing the project. Progress is being made in a slow but steady manner.

Conclusion

We propose to develop an adaptive learning system to improve STEM interest among middle school students. We will use engineering tools and methods to build predictive models to improve facilitating and mentoring of students for this purpose. We will analyze the research infrastructure with our hypotheses to show that it is working. Weekly usage graph will show frequency and duration of App usage. We will measure STEM interest, pre- and post-survey, based on the App usage. Educational researchers can then take this work forward to a full-scale research study with our or other Apps, and pose more meaningful questions on informal learning.

Bibliographic Information

1. Lowell et al., Steady as She Goes? Three Generations of Students through the Science and Engineering Pipeline, Annual Meetings of the Association for Public Policy Analysis and Management, Washington, D.C., November 2009.
2. Bienkowski et al., Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, U.S. Department of Education Office of Educational Technology, October 2012.
3. Tai et al., Planning Early for Careers in Science, Science, volume 312, AAAS 26 May 2006.
4. Urness, T., & Manley, E. D., Generating interest in computer science through middle-school android summer camps. *Journal of Computing Sciences in Colleges*, *28*(5), 211-217, 2013.
5. Adams, J. C., Alice, middle schoolers & the imaginary worlds camps, SIGCSE Bull, 39, (1), 307-311, 2007.
6. References to our previous work will be provided later (as this is a blind review)
7. Romero et al., Handbook of Educational Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, October 25, 2010.
8. Perera et al., Clustering and Sequential Pattern Mining of Online Collaborative Learning Data, IEEE Transactions on Knowledge and Data Engineering, 21, 759-772, 2009.
9. Tan et al., Introduction to Data Mining, Addison-Wesley; Second edition, November 13, 2014.
10. Witten et al, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann; Third edition, January 20, 2011.