Pipeline for Expediting Learning Analytics and Student Support from Data in Social Learning

Yohan Jo†, Gaurav Tomar†, Oliver Ferschke†, Carolyn Penstein Rosé†, Dragan Gašević‡
†School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
{yohanj, gtomar, ferschke, cprose}@cs.cmu.edu
‡Schools of Education and Informatics, The University of Edinburgh, Edinburgh, UK
dgasevic@acm.org

ABSTRACT

An important research problem in learning analytics is to expedite the cycle of data leading to the analysis of student progress and the improvement of student support. For this goal in the context of social learning, we propose a pipeline that includes data infrastructure, learning analytics, and intervention, along with computational models for individual components. Next, we describe an example of applying this pipeline to real data in a case study, whose goal is to investigate the positive effects that goal-setting students have on their peers, which suggests ways in which we might foster these social benefits through intervention.

Categories and Subject Descriptors

K.3.1 [**Computer Uses in Education**]

General Terms

Algorithms, Human Factors, Languages

Keywords

Learning Analytics, Social Learning

1. INTRODUCTION

More and more recent work in educational data mining and learning analytics refers to a "virtuous cycle" of data leading to insight on what students need and then improvements in support for learning [3]. An important goal is tightening this cycle. In this paper, we propose a pipeline and its component models that can achieve this goal.

In this work we are specifically interested in social learning. Social learning is based on a Vygotskian theoretical frame where learning practices begin within a social space and become internalized through social interaction. That interaction may involve observation or more intensive interaction through cycles of feedback and help exchange.

We will present a three-part pipeline for expediting data analysis and student support in social learning. The pipeline consists of the data infrastructure for a uniform interface with heterogeneous data from social interaction, a probabilistic model for analyzing learning pathways in social contexts, and recommendation technology to evaluate and support learning processes. We will then present a case study, in which we describe an example of applying our method to real data and summarize the findings from the case study.

LAK '16, April 25-29, 2016, Edinburgh, United Kingdom
ACM 978-1-4503-4190-5/16/04.
http://dx.doi.org/10.1145/2883851.2883912

2. MODELS AND CASE STUDY

This section details the three models that power the separate components of the proposed pipeline as well as a case study where we applied the pipeline to data from social learning. The goal of the case study is to investigate the positive effects that goal-setting students have on their peers, to measure the extent to which students are benefiting from these peer effects, and to offer a proof of concept of our recommendation system for increasing the extent to which these valuables connections between peers are made.

2.1 Course Context

The data used in the case study was collected from the ProSolo learning platform [2]. The course was offered on edX with the title *Data, Analytics, and Learning* (DALMOOC) from October to December 2014. In this *dual layer* MOOC, students had the option of choosing a more standard path through the course within the edX platform or to follow a more self-directed and social path in the external environment ProSolo. The ProSolo layer encouraged students (1) to set their learning goals, (2) to write posts in the discussion forum on ProSolo or link their external posts on their own blogs and Twitter, and (3) to follow other students, so that they can easily access their followees' activities and posts.

2.2 Data Infrastructure—DiscourseDB

Model: The significance of the data infrastructure proposed here, referred to as DiscourseDB[1], lies in its ability to map diverse forms of discussion into a common representation, making it easy to apply analytic tools to different types of social interactions across multiple platforms. Hence, analytic and intervention models can be applied with little change to any data once imported into DiscourseDB. The structure of DiscourseDB is represented in a relational database as an entity-relation model of connected discourse contributions organized in generic, nested discourse containers. For content information, DiscourseDB stores individual posts or messages into the `contribution` table, along with their past

[1] `http://discoursedb.github.io`

revisions and relations between them, which can be arbitrarily typed (e.g., reply-of). User information is stored in the user table, along with information about social relationships between users and users' contributions.
Case study: We import the ProSolo data into DiscourseDB. Students' individual goal notes, forum posts, blog posts, and tweets are mapped to contributions in DiscourseDB. Every contribution has the information about its source platform. The reply relations between posts are stored in discourse_relation. Students are mapped to the user table, and their follow relations are stored in the user_relation table.

2.3 Learning Analytics—Sequence Model

Model: We model students' learning paths such that the building blocks of learning paths are induced from the data. This approach may find more representative units of student interests than predefined building blocks. For this purpose, we propose an extension of the previously published State Transition Topic Model (STTM) [1], in order to infer learning paths from student behavior traces in a course. STTM is a combination of a Hidden Markov Model and Latent Dirichlet Allocation, where each state is represented as a topic distribution. Therefore, STTM can learn topics students are interested in within each state, and estimate a student's state at each time point. STTM also learns the probabilities of state transitions, which reflect students' progress. However, the original published STTM model is incapable of investigating how learning paths differ depending on the student's social status, e.g., the existence or lack of connection with certain peers. Therefore, we extended STTM to learn different transition probabilities for students depending on their assigned status as well as the trend of students' engagement in course topics and media in each state.
Case study: We apply the model to the imported ProSolo data to analyze students' learning paths conditioned on their social relationships. We are specifically interested in students who set goals using goal notes and in examining whether students who follow them have different learning paths from the others. The learning paths will show how such social connections make a difference in the selection of course topics and social media the students choose. Our case study revealed that students following goal-setters show more interest in hands-on practice and subjects in the later part of the course in comparison to other students. Their transitions between states imply that they are more likely to link course materials learned across the course as well.

2.4 Intervention—Recommendation System

Model: Once we have identified patterns that distinguish successful and unsuccessful student paths, we may want to introduce interventions that we believe will increase the prevalence of successful paths. In the current work, that involves supporting students in adding connections to positive role models in their network. In many learning environments, discussions are the main means of social interaction among students. Our specific work proceeded by first assessing the extent to which students benefited from specific social connections based on analysis of goal-setting behavior, and then by proposing a social recommendation approach that would enable students to find opportunities to add such connections. For this purpose, we first investigated sensitivity on the part of students to identify effective role models to connect to naturally. The investigation was conducted through link prediction by a social recommendation system that extends the matrix factorization model developed by Yang et al.[4]. The link prediction involved predicting connections students make through post-reply actions on discussion threads. Incorporation of the goal-setting behavior of students in the model did not help this prediction task. That suggests that the students do not demonstrate a sensitivity to peer students' goal-setting behavior while making connections. In order to help students connect to positive role models in their network, we extended the matrix factorization model, which already recommends connections as per preferences made by the students in the past, by introducing a constraint which makes connections with positive role models (students having good goal-setting behavior) mandatory. The model first generates recommendations which are relevant for students and then filters out some based on the constraint imposed.
Case study: Our analytics have identified a positive effect associated with social connection with goal-setting students. Thus, we are looking forward to opportunities to introduce interventions that would introduce more of these connections into the experiences of students. For example, we might provide students with opportunities to interact more frequently with goal-setters through discussion in the forums and through other social affordances by means of social recommendation. Our model has been demonstrated to recommend learning partners that are not only relevant to a student's preferences but are also qualified role models.

3. CONCLUSION

We proposed a pipeline for expediting the process of moving from learning analytics to student support by using computational modelling approaches at every step in the pipeline. The concrete example of applying the pipeline to a case study and the result of the case study show the potential of the pipeline for various studies about social learning. We believe that social learning can benefit from technologies, such as discussion recommendation, conversational agents, and collaboration tools in MOOCs[2].

4. ACKNOWLEDGEMENT

This research was supported by the National Science Foundation under grants ACI-1443068 and IIS-1320064, and by the Naval Research Laboratory and Google.

5. REFERENCES

[1] Y. Jo and C. P. Rosé. Time Series Analysis of Nursing Notes for Mortality Prediction via a State Transition Topic Model. In *CIKM '15*, 2015.
[2] C. P. Rosé, O. Ferschke, G. Tomar, D. Yang, I. Howley, V. Aleven, G. Siemens, M. Crosslin, D. Gasevic, and R. Baker. Challenges and Opportu- nities of Dual-Layer MOOCs: Reflections from an edX Deployment Study. In *CSCL '15*, pages 848–851, 2015.
[3] C. Thille. Education Technology as a Transformational Innovation. *White House Summit on Community Colleges: Conference Papers*, pages 73–78, 2010.
[4] D. Yang, D. Adamson, and C. P. Rosé. Question recommendation with constraints for massive open online courses. In *RecSys '14*, pages 49–56, 2014.

[2] http://dance.cs.cmu.edu/resources/