RESOLVING LEARNING PATHWAYS USING TRANSITIVE ITEM NETWORKS ONLINE

Charles Lang¹, Erik Choi², Karolina Wójcik², Tomasz Sienkiewicz²

¹Teachers College, Columbia University (UNITED STATES)
²Brainly.com (POLAND / UNITED STATES)

Abstract

In this study, we demonstrate the utility of item-learner and transitive-item networks for generating personalized item suggestions to learners. We use data from the Brainly social learning network in which learners can interact by asking and answering each other’s questions. We first build an item-learner network and then weight this network using data on whether or not answers to the items received a nomination as a ‘best answer’ and the sequences of best answer items. We then use these two networks to generate suggested next questions using nearest neighbours and the PageRank algorithm. We discuss patterns in the subject areas that the suggested next items come from.

Keywords: Learning pathways, item networks, learning progression, stochastic network analysis, online education.

1 INTRODUCTION

Understanding the heterogeneous pathways that students take through content to gain mastery of a topic or skill is a longstanding goal of the educational enterprise. Many solutions to the problem have been suggested using a wide range of methods including expert opinion [1], error analysis [2], skill analysis [3], sequential pattern analysis [4] and Q-matrices [5]. Learning pathways are an essential tool in standards-based or competency-based education and are a necessary component of any automated personalization system such as intelligent tutoring systems. A known pathway allows student progress to be estimated, students to be differentiated based on progression and specific interventions to be designed to aid students at different stages or branches of the path. This mapping enterprise has been greatly aided by the advent of the internet and mobile computing. The ease with which data can be collected within educational platforms has ballooned allowing for greater granularity in mapping learning paths. For example, large-scale mapping efforts using online games, such as those of Andersen et al.[6], have provided insight into differences across nations and education systems. Of particular interest to us is the use of mapping techniques to provide personalized guidance to learners. An informative map of domain areas can be used as a structure to direct learners to the next area of study that they should tackle, based on success of that sequence for previous learners.

Brainly is a social learning network in which leaners can ask, and get answers, to questions about school-level content such as mathematics, chemistry or English [7]. The network incentivizes engagement with the site content through motivational points that aim to encourage question asking and answering. Understanding how successful learners navigate this semi-structured format might allow the development of adaptive processes that can aid less successful learners on the website. For example, suggesting the next question a learner should tackle based on their previous sequence of questions. There are many features that we could optimize these suggestions on, for example engagement or correctness, but here we test using the structure of the network of questions (items), connected by which students answered which questions as a means of suggesting the next item a learner should tackle.

Building on the mapping techniques of both Q-matrices and error analysis the following research classifies items according to whether or not answers to the item receive a “best answer” nomination by the website administrators. However, in addition to considering the sequence, we also consider how different sequences impact students differentially. We have developed a method (transitive item networks) based on probabilistic longitudinal network analysis [8] to extract preferred question order from a set of items. In this way, we can make visible the structure of the underlying knowledge required to master the content. For example, if question A contains required knowledge to answer question B, then receiving question A before B will more often lead to a best answer than receiving question B before A. In this paper the theory behind the analysis is explained. It is then applied to a
data set from an online academic question and answer platform, Brainly, in which students ask and answer each others’ questions. This data is then used to simulate the proportions of questions that would be suggested to students if the network was used to suggest the next question.

2 METHODOLOGY

The data consist of 800 items randomly drawn from the Brainly.com database in lots of one hundred for each of eight subject areas: Biology, chemistry, English, geography, history, mathematics, physics and social studies. 9587 learners from the US provided answers to these items and 10% of the items received answers that were denoted “best answer” by the system administrators. Time stamps for these interactions were also recorded.

2.1 Item-learner and transitive item networks

First, an item-learner network was generated by creating an adjacency matrix of items with respect to which learners answered each item. If we think of an item as a club, and an answer as a learner being a member of an item club, the adjacency matrix represents which clubs have common members. If a learner interacts with two items then those items are considered to have a connection and a network can then be mapped of items that were commonly related by learners.

Second, a transitive item network was generated by incorporating information about “best answer” status and answers sequence into the item-learner network. Learners who provide best answers to items presumably demonstrate superior knowledge of the content in that item, but are also likely have superior content knowledge more generally. We can quantify this general domain knowledge by weighting connections between items that have common best answerers more than those that have only common answers in the item-learner network. What’s more, we can incorporate the sequence of answers by weighting connections between a sequence of items with the same best-answerer more than those for which the best answerer has only provided a single best answer. In aggregate this also provides some directionality to the network, as items that received many best answers in a sequence will become very strong, while those that do not will become weaker.

2.2 Simulating suggested items

We simulated how the transitive item network can be used to suggest questions by using a nearest neighbors approach and the PageRank algorithm. For every item in the data set we can calculate the nearest neighbors to that item in terms of connections. We can then rank those neighbors with respect to their overall connectedness within the network using the PageRank algorithm [9]. The suggested next item is can then be defined as the nearest neighbor that is most connected within the network. We can then compare the item-learner and transitive item networks with respect to which subject areas they would tend to send learners next.

All analysis was performed in R, code is available at https://github.com/charles-lang/transitive-item-network

3 RESULTS

3.1 Networks

The item-learner network contains 5538 connections, the average number of connections per item being 7 (Fig.1). Connections tend to be concentrated around physics (28%), the subject that also has the highest level of intra-subject connections. Geography and history are also well-connected subjects with 16% and 14% of all connections going to these two subjects respectively. Chemistry has the lowest number of connections (6%).

The transitive item network contains 179 connections and also emphasizes connections within physics and geography and from physics to geography, though it de-emphasizes connections toward English and chemistry.
3.2 Suggested items

The unweighted item-learner network tends to suggest learners tackle items within English next, with English making up more than 40% of the item suggestions. The next most popular is physics with 23% of suggestions (Fig. 2). Once the sequence and best answer information is incorporated into a transitive-item network suggestions toward English questions are dramatically reduced to less than 10% while physics remains stable. Suggestions toward chemistry are also reduced from 12% to 6% but geography, history and mathematics all increase.

4 CONCLUSIONS

We can break the conclusions from this study into two categories, those that are clearly sample specific and those that are concerned with the utility of this method to produce useful question suggestions to students. In the first category we can see that clearly physics is very connected subject. This is a result of it having a core group of engaged students on this platform who are working on common material. We can see this as physics seems to have many intra-subject connections suggesting that students work on a common set of problems in that domain. Unlike a subject such as English that also has a core group of students but those students do not necessarily cover the same
content. For example, the texts studied in English may vary across schools so that questions can appear unrelated in the network as they cover specific aspects of those texts (characters, themes, locations). This leads to an active but much more diffuse network among English students and students who may come across each other often within physics may never see questions and answers from each other in English. What’s more, there are relatively few connections between English and other subjects - English content may therefore not be important for success in other content areas. Whereas concepts in physics are likely shared by geography, mathematics chemistry and biology and the strong group of physics students makes it appear that transference between physics knowledge and other subjects is important. Physics has a sense of gravity in this network with students highly active in that area and this influences activity in subjects related to it.

The transitive network emphasizes connections between items that knowledgeable learners have provided answers to. It also preferences sequences of items that are connected by best answers. Given these characteristics the small number of transitive connections toward English and the drop in suggestions of English content between the two networks suggests that although there are a lot of learners doing English questions on the site, the connections between those questions do not seem to require each other – they require discreet knowledge. This is different in the case of physics in which students tend to require other knowledge, particularly from within physics, to successfully answer the questions.

If we are building a recommender system within the Brainly site and we want to drive pedagogical goals, the transitive network is likely the superior network than the item-learner network only. Although this requires in situ testing, the transitive network provides suggestions that are based on required sequences that may be useful for a subject like physics. It may be less useful for a subject like English because the heterogeneity of content knowledge is too great. It is also interesting to see the amount of inter-subject suggestions. Perhaps subject demarcations are not the best predictor of what a student should study next and restricting students to questions within subject boundaries may not make the best suggestions either.

REFERENCES


