Tree Structure of Collective Attention Network: Revisiting the Problem of Dropout

Jingjing Zhang
School of Educational Technology, Beijing Normal University
Jingjing.zhang@bnu.edu.cn

Ming Gao
Research Centre of Distance Education, Beijing Normal University
mgao519@mail.bnu.edu.cn

ABSTRACT: Treating the dropout phenomenon as a sign of an individual's choice highlights the importance of understanding how dropouts learn in MOOCs. Conventional learning analytics methods failed to make sense of limited behavior data left by dropouts. This study uses the minimum spanning tree of collective attention network to investigate how dropouts behave in a selected MOOC. It is interesting to note that assessments embedded in the MOOCs seem to play a rather important role in guiding dropouts to learn. Redefining assessment in open and flexible learning environments to construct a minimum cost network of collective attention is vital to make this online space cost-effective for better learning.

Keywords: MOOCs, dropout, collective attention, learning analytics, pattern mining

1 INTRODUCTION

Alongside the increasing development of MOOCs accommodating open and flexible learning experiences at scale, the unusual high dropout rates beyond 90% of participations are alarming (Jordan, 2014). The dropout phenomenon was first treated as a sign of deficient quality, but later as an explicit expression for an individual's choice. This later counter-argument highlights the importance of focusing on how dropouts learn in MOOCs, and in what ways the learning design of MOOCs facilitates their learning. Nevertheless, few studies have taken the learning analytics approach to understanding how dropouts learn, as no or little behavior data of dropouts can be meaningfully addressed using the data mining approach. Furthermore, the behavior data of dropouts are often removed as "outliers" in traditional statistical analysis. To address such a problem, we build upon the earlier research by using the network model of collective attention to investigate how dropouts learn at the collective level in a MOOC. A key innovation is the focus on how to make sense of learning patterns by using a new method to model short, limited, and heterogeneous behavior trajectories left by dropouts.

2 METHODS

2.1 Context

'Introduction to Psychology (2018 autumn)' offered on XuetangX was selected as the case to study. This course offered 70 learning resources, including videos, quizzes, and an exam, within 13 units. 9508 learners participated in the course, and their behavior data were automatically stored in the database. Due to the incomplete records of behavior data (no. 5237) via mobile devices, deficient information of registration and exam (no. 2110), only 1892 dropouts out of 2161 were selected for this study. The dropout rate of 88% of this course is a typical rate frequently reported in the literature. About 200 dropouts registered before the course started, and about 400 dropouts registered after the course ended. The accumulated number of participations increased over time, and half of the learners have registered before the mid-term. This pattern of registration reflects on learners' choice to learn at their pace. The pattern of learners' visits is also a typical long-tail distribution. While about 700 participates accessed courseware in unit 1, the learners drop out over time, resulting in only 16 learners accessed the last unit – exam.

2.2 An Open-Flow Network of Collective Attention and its Minimum Spanning Tree

The classical social network is a closed model that fails to account for the high rates of attrition and steeply unequal participation patterns of learners. This study built upon our earlier research, and adopted the open flow network of collective attention (Zhang, Lou, Zhang, & Zhang, 2019) to model learners behaviors, and see this article for a comprehensive review of collective attention. This network model, using node to represent the learning resources and the link to represent the learners' sequential visits across the learning resources. At the collective level, the large body of learner's sequential visits resembles the flux of attention flowing in and out of the learning resources. The flux of such attention flow forms a network, in which two artificial nodes - 'source' and 'sink'- were added to represent the offline space. Thus, this network becomes an open and balanced model, which allows collective attention to flow in and out across online and offline spaces. As for individual learning resources, the inflow of attention equals the outflow.

In such a collective attention network, flow distance (Guo et al., 2015) measures the average first-arrival distance between nodes, by using the N-order Markov transition to calculate the probability that attention would flow in or out of a learning resource. A skeleton of the network, including all learning resources, was generated by using Minimum Spanning Tree (MST) (Kruskal, 1956). In such a tree structure, for any node, another node with the shortest flow distance to it was added until all the nodes were added to the tree containing a sum of flow distances, which is minimum. Zhang and her colleagues (2019) found that the amount of attention flow and flow distance were negatively correlated. Thus, the weight of link is calculated using the opposite of flow distance between two nodes, which implies that the likelihood that amount the attention flow (including direct and indirect) in and flow out between two nodes in such a network. The learning resources in the same color belong to the same unit, and the size of the node is proportional to the amount of attention flow in/out to this learning resource. Python and Gephi were used for data analysis and visualization.

3 RESULTS AND CONCLUSION

The minimum spanning tree represents a new structure formed by using real behavior data of dropouts (see Figure 1). Such a tree, representing the topological properties of the collective attention network, yields a lower bound on the cost of collective attention. As argued by Zhang et al. (2019), MOOC learning is pricey at the cost of the learners' attention, and thus the topological structure of such a tree contributes to the design of cost-effective learning resources to prevent learners from becoming overloaded.

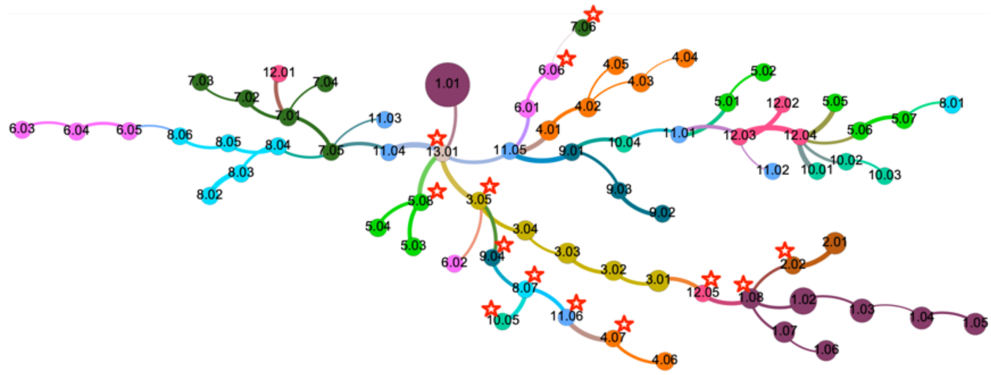

Figure 1: The minimum spanning tree of learners' collective attention flow

The visiting pattern of dropouts presents a long-tail distribution, which also reflects on the MST. As shown in Figure 1, the size of nodes belonging to Unit 1 is the largest, while the size of node 13.01, the exam, is the smallest. This illustrates that the amount of attention flow decreases from earlier units to later units. However, it is interesting to discover that the largest node 1.01, which the most of collective attention flows in and out, is not the center of this MST. It is instead an isolated leave connecting to the central node of the exam, and the rest learning resources belonging to Unit 1 are clustered together lie leaves on the other side. Instead, the exam Unit (13.01) is the distinct center of the MST, which implies that the cost of collective attention is minimum by giving the assessment a central role to connect with other learning resources.

We can also see in this tree a separation of the learning resources into three large branches of resources and several twigs of varied lengths. It is interesting to note that quizzes are likely to be at the crossing of the main branches, such as quiz 1.08, 3.05, 5.08, and 8.07 (marked in red star in Figure 1). Notably, several quizzes across different units (e.g., 3.05, 4.07, 8.07, 9.04, 10.05, 11.06) form a cluster, and quizzes are also likely to act as the bridges between video resources. For example, Quizzes in unit 1, 2 and 12 serve as the bridges to link unit 1, 2, and 3. One possible explanation for this result is that quizzes are not used by dropouts to evaluate their studies, instead dropouts use quizzes as a learning strategy to guide their study.

This preliminary exploration of how dropouts learn in a selected MOOC only sheds light on behavior patterns using the model of collective attention and its MTS. Likely learning intentions and learners' profiles, as argued in the literature, play a significant role in constructing similarities in patterns of learning behaviors. In our future work, how to incorporate nonstructural properties, such as intention, capacity, time, etc. in the model of collective attention is to be seriously considered.

REFERENCES

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses Massive Open Online Courses. *International Review of Research in Open and Distance Learning*, *15*(1).

Kruskal, J. B. (1956). On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Proceedings of the American Mathematical Society, 7(1), 48.

Guo, L., Lou, X., Shi, P., Wang, J., Huang, X., & Zhang, J. (2015). Flow distances on open flow networks. *Physica A: Statistical Mechanics and Its Applications*, *437*, 235–248.

Zhang, J., Lou, X., Zhang, H., & Zhang, J. (2019). Modeling collective attention in online and flexible learning environments. *Distance Education*, *00*(00), 1–24.