2019

Towards the Readiness of Learning Analytics Data for Micro Learning

Jiayin Lin
University of Wollongong, jl461@uowmail.edu.au

Geng Sun
University of Wollongong, gsun@uow.edu.au

Jun Shen
University of Wollongong, Massachusetts Institute of Technology, jshen@uow.edu.au

Tingru Cui
University of Wollongong, tingru@uow.edu.au

Ping Yu
University of Wollongong, ping@uow.edu.au

See next page for additional authors

Towards the Readiness of Learning Analytics Data for Micro Learning

Abstract

With the development of data mining and machine learning techniques, data-driven based technology-enhanced learning (TEL) has drawn wider attention. Researchers aim to use established or novel computational methods to solve educational problems in the 'big data' era. However, the readiness of data appears to be the bottleneck of the TEL development and very little research focuses on investigating the data scarcity and inappropriateness in the TEL research. This paper is investigating an emerging research topic in the TEL domain, namely micro learning. Micro learning consists of various technical themes that have been widely studied in the TEL research field. In this paper, we firstly propose a micro learning system, which includes recommendation, segmentation, annotation, and several learning-related prediction and analysis modules. For each module of the system, this paper reviews representative literature and discusses the data sources used in these studies to pinpoint their current problems and shortcomings, which might be debacles for more effective research outcomes. Accordingly, the data requirements and challenges for learning analytics in micro learning are also investigated. From a research contribution perspective, this paper serves as a basis to depict and understand the current status of the readiness of data sources for the research of micro learning.

Disciplines

Engineering | Science and Technology Studies

Publication Details

Lin, J., Sun, G., Shen, J., Cui, T., Yu, P., Xu, D., Li, L. & Beydoun, G. (2019). Towards the Readiness of Learning Analytics Data for Micro Learning. Lecture Notes in Computer Science, 11515 66-76. San Diego, United States International Conference on Services Computing

Authors

Jiayin Lin, Geng Sun, Jun Shen, Tingru Cui, Ping Yu, Dongming Xu, Li Li, and Ghassan Beydoun

Towards the Readiness of Learning Analytics Data for Micro Learning

Jiayin Lin[1], Geng Sun[1], Jun Shen[1,2], Tingru Cui[1], Ping Yu[1], Dongming Xu[3], Li Li[4], Ghassan Beydoun[5]

[1]School of Computing and Information Technology, University of Wollongong, Wollongong, Australia
[2]Research Lab of Electronics, Department of EE and CS, Massachusetts Institute of Technology, Cambridge, USA
[3]UQ Business School, The University of Queensland, Brisbane, Australia
[4]Faculty of Computer and Information Science, Southwest University, Chongqing, China
[5]School of Information System and Modelling, University of Technology Sydney, Sydney, Australia
jl461@uowmail.edu.au, {gsun, jshen, tingru, ping}@uow.edu.au, d.xu@business.uq.edu.au, lily@swu.edu.cn, ghassan.beydoun@uts.edu.au

Abstract. With the development of data mining and machine learning techniques, data-driven based technology-enhanced learning (TEL) has drawn wider attention. Researchers aim to use established or novel computational methods to solve educational problems in the 'big data' era. However, the readiness of data appears to be the bottleneck of the TEL development and very little research focuses on investigating the data scarcity and inappropriateness in the TEL research. This paper is investigating an emerging research topic in the TEL domain, namely micro learning. Micro learning consists of various technical themes that have been widely studied in the TEL research field. In this paper, we firstly propose a micro learning system, which includes recommendation, segmentation, annotation, and several learning-related prediction and analysis modules. For each module of the system, this paper reviews representative literature and discusses the data sources used in these studies to pinpoint their current problems and shortcomings, which might be debacles for more effective research outcomes. Accordingly, the data requirements and challenges for learning analytics in micro learning are also investigated. From a research contribution perspective, this paper serves as a basis to depict and understand the current status of the readiness of data sources for the research of micro learning.

Keywords: Micro Learning, Learning Analytics, Machine Learning, Data Mining, Data Insufficiency

1 Introduction

The rapid evolution of technologies and the changes in people's lifestyle make technology-enhanced learning (TEL) become a hot topic in recent years. In the context of big data, despite the technical difficulties, data problem is an obvious obstacle for the development of TEL research. As most models are driven by data, sufficient, consistent and complete data is the fundamental factor for system design, model construction, and evaluation for learning analytics. Learning analytics (LA) conventionally involves the evaluation, collection, analysis and reporting of data about learners and their contexts, for the purpose of understanding and optimising learning and environments which it occurs [1]; LA becomes increasingly important when the hype of big data and artificial intelligence are more and more distilled into the education sector.

Among research topics in the TEL area, a notable term, micro learning, represents a novel educational service, which provides users with small chunks of personalized learning materials [2]. Such service aims to enable learning activities being carried out by effectively making use of users' fragmented time, and this service can be applied to different online learning platforms such as traditional learning management system (LMS) and MOOC. We refer to the whole integrated processing flow of a micro learning service as a micro learning system. Micro learning is a representative interdisciplinary research topic, which is promoted by data mining and machine learning and consists of various technical topics such as recommendation, resource fragmentation, content analysis, and prediction.

In this paper, we focus on discussing the vital application stages that involve LA, and the data requirements and challenges for micro learning. It should be clarified that the strict definition of LA is beyond the scope of this paper. As this paper particularly reviews a portion of representative prior studies, we have left the systematic literature review of the data requirements and challenges as an opportunity for expanded future work. The remainder of this paper is organized as follow. In Section 2, before discussing the details of the data sources that used in each processing stage or various LA tasks, we first describe the structure and the key components of our proposed micro learning system. We discuss several representative segmentation and annotation studies and the data sources used in literature in Section 3. The data sources used in the recommendation, prediction, and analysis tasks are discussed in Sections 4 and 5. In Section 6, we summarize the data requirements and challenges for the micro learning research. This paper is concluded in Section 7.

2 Micro Learning System

As aforementioned, the provision of micro learning is an online educational service, which aims to utilize users' fragmented spare time and offer users small pieces of personalized learning materials. To realize this online service, the system requires several components for learning material preparation and personalized decision-making. Based on the service requirements, a micro learning system consists of three core parts (i.e., non-micro learning material segmentation, learning material annotation, and learning material recommendation) and several prediction and analysis modules, which work together to provide a complete personalized online learning service. In this section, before diving into the details of the data readiness of micro learning, we firstly outline the typical structure of a micro learning system and the data flow of this system, as shown in Figure 1.

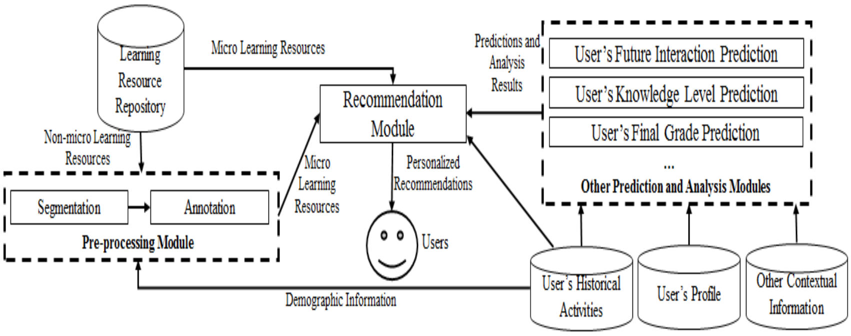

Fig. 1. The Structure and the Data Flow of the Proposed Micro Learning System.

2.1 Pre-processing Model

On many online learning platforms such as LMSs, the duration of consuming a learning material is longer than 15 minutes [2]. For a micro learning service, non-micro learning materials are logically segmented into coherent knowledge points prior to be delivered to learners. After the segmentation process, an annotation step is required to make learning resources both machine understandable and human understandable, to overcome the lack of interpretability of the micro learning resources. As discussed in many studies [3, 4], proper annotation, indexing, or tagging are essential for the retrieval, recommendation, and reuse of resources.

2.2 Recommendation and Prediction Models

The recommendation phase is central to a micro learning service, which greatly determines what information will be delivered to the users. In the educational domain, as prior studies [5, 6] point out, with the plethora of online learning resources and increasingly frequent formal and informal learning interactions, users can benefit a lot from services which help them quickly and precisely identify the suitable learning resources. Moreover, in the era of big data, more and more learners demand an intelligent system to help them pick suitable information or filter out the irrelevant one [7].

To realize a comprehensive micro learning service, a system should also involve some intelligent prediction and analysis modules to boost decision-making, such as behaviour prediction and performance prediction. With a large number of users, it is impractical to manually analyse each user's learning history or requirements. The prediction and analysis results, such as a user's knowledge level and the difficulty level of learning material, could enrich the information for the system's decision-making processes.

3 Annotation and Segmentation

The task for pre-processing is to transform the non-micro learning materials to the micro learning ones, which consists of segmentation and annotation. Segmentation and annotation are mainly driven by the content information of the learning materials; hence, the readiness of the research data is vital for the research on segmentation and annotation.

As many learning materials are in video format, many studies of annotation and segmentation heavily rely on transforming the image and audio information to textual format [3, 8, 9]. Researchers extract textual metadata by applying Optical Character Recognition (OCR) and Automatic Speech Recognition (ASR) on the lecture videos [8]. The experiment in [8] only involves 20 randomly selected lecture videos from different speakers. The study [3] used natural language processing (NLP) techniques to further mine the extracted textual information in assisting the annotation process. The dataset used in this study was related to Objective Oriented Programming in French, but the details and source of this dataset were not given.

Some segmentation or annotation models are based on the users' demographic information. We refer this demography-based method as the 'crowd-wisdom' method. The log file of users' watching interactions of four edX[1] courses are analysed in the study [10]; the researchers argued that the re-watching peaks of the whole user population might be the crucial knowledge for LA because these peaks of the re-watching point can be used to further identify the boundaries of knowledge points. Another lecture annotation system is proposed in the study [11], but only 21 students were involved. The study [4] proposed a crowd-wisdom based model to integrate annotation results, but only an image dataset [12] was used. A semantic extraction model and a tagging model were proposed in [13] to annotate the online learning resources. However, the data source for model construction and validation was not elaborated in their paper.

4 Recommendation

Recommender system is a hot topic and has been studied for many years, but in the area of micro learning it is still in the embryonic stage. Insufficient, inappropriate and unknown data sources are the main challenges for the research of micro learning recommendation.

4.1 Insufficient Data Source

The term 'insufficient data' means the data used in a study can only partially reflect the underlying issues against the context of potentially bigger data, and the experiment result of recommendation might be biased.

The study [7] proposed a hybrid recommendation algorithm, which could reflect the timeliness of a learning procedure, but as few as 30 students were involved in this study. Metadata for Architectural Contents in Europe (MACE) and TravelWell da-

[1] https://www.edx.org/

tasets were used [14] for training the usage context-boosted recommender system. However, both datasets contain very few users and only a fraction of subjects. One prior study [15] used a convolutional neural network (CNN) to model the latent factors based on the BookCrossing dataset [16]. However, in the e-learning domain, the type of learning materials could be in the format of video, audio, and text, the conventionally trained model was not sophisticated enough for micro learning. The fault tolerance and the capability of self-optimization prompt swarm intelligence and evolutionary computing to be applied in many studies for learning path optimization [17, 18]. However, the sizes of experiments in these study were still defective, as only 80 students were observed [17] and only one chapter of mathematics teaching material was used in [18]. Due to the heterogeneity, insufficient users and/or subjects cannot generally reflect the latent patterns of e-learning scenarios or the whole user population in the real-life micro learning environments.

4.2 User Behaviour Prediction

User behaviour prediction and analysis are also important for the system's decision-making. As discussed in the study [31], the way how a learner may interact is worthwhile to understand in order to provide fine-grained insights into what particular content may be improved for further modification or adaptation of the learning activities. But only investigating one course [28] is inadequate for training and validating the model proposed in [31]. Another prior study [32] suggested that different watching patterns might represent different cognitive levels, where the users' next behaviours and future performance could be predicted by clicking interactions. However, only one course [33] comprising 48 lecture videos was examined by authors.

5 Data Requirements and Challenges

As discussed in the previous sections, most models involved in the LA systems are data-driven. For a micro learning system, the required data source can be roughly classified into five categories: user' historical learning and interaction records from log files, users' profile and items' content information stored in the relevant databases, and other contextual information captured by the platform and its various plug-ins. The summarization of the utility of the different data types is shown in Table 1.

5.1 Data Requirement for Recommendation

For a recommender system, users' historical learning activities are indispensable for training and validating the models. In most cases, the information about user's historical activities only exists in the log files and cannot be crawled from online learning platforms or Websites. As discussed in [15], data-driven recommendation methods require extensive historical data, which is difficult to obtain from the e-learning system. A recommending decision should also be made by referencing the contextual information of current learning activity, users' profile, and items' profile. For research purposes, some information components such as the resource descriptions are open to the public and can

Table 1. The Utility of Different Types of Data

<table>
<tr><th colspan="3">Data Type</th><th>Utility and Description</th></tr>
<tr><td colspan="3">User-Item Rating Matrix (Learning Records)</td><td>In most cases, users' historical rating is indispensable for constructing a recommender system.</td></tr>
<tr><td rowspan="3">Content Information</td><td colspan="2">Textual Information</td><td rowspan="3">Main information sources for segmentation and annotation.</td></tr>
<tr><td colspan="2">Audio/Video Information</td></tr>
<tr><td colspan="2">Other Metadata</td></tr>
<tr><td rowspan="5">User's Interaction</td><td rowspan="3">Learning Interaction</td><td>Clickstream</td><td rowspan="5">Main information sources for the tasks of performance prediction and analysis. Various prediction and analysis models, such as early prediction and learning path design, are based on the users' interaction records. Demographic information extracted from users' interaction can also be used to improve the recommendation and segmentation results.</td></tr>
<tr><td>Comments</td></tr>
<tr><td>User's Access log</td></tr>
<tr><td colspan="2">Quiz/Exam Performance</td></tr>
<tr><td colspan="2">User's Sequential Learning History</td></tr>
<tr><td colspan="3">User's Profile</td><td>The main source of the information about user characteristics. Such age and learning interests, which could be used in the user-based collaborative filtering recommender system.</td></tr>
<tr><td colspan="3">Contextual Information</td><td>For the decision making, contextual information is used as supplementary, which could be time, location, or anything included in the learning activity. It is the key for a context-aware system.</td></tr>
</table>

be crawled from the online learning website. However, some information only exists in the log file or can only be captured via extra plug-ins.

5.2 Data Requirements for Pre-processing

Before the commencement of recommendation, there is a pre-processing stage to get micro learning materials ready and mine the user's relevant information. The segmentation and annotation strategies are mainly based on the content of the learning materials. In addition, some crowd-wisdom based segmentation [10] and annotation [4, 11, 13] models rely on the demographic information about user's historical interactions. Most content information of the learning material is open to the public while the demographic information is not. On the contrary, the vast majority of prediction models and analysis processes are based on the users' historical interaction data such as clickstream, quiz performance and users' comments.

Table 2. The Problems of Used Datasets.

Problem of the Dataset	Descriptions
Insufficient Data [3, 7, 8, 11, 14, 17, 18, 25-27, 32]	Many research teams can only access and use a fraction of learning materials, or only a small number of users or learners are involved in the experiments. Many case studies in this field are based on relatively small samples. For the research in the big data context, the samples involved in the experiments should be able to cover most cases of the real-life application scenarios
Inappropriate Data [4, 15, 19, 20]	Due to lack of dataset, some research can only use simulated dataset or the dataset from other domains for experiments. The simulated dataset could be problematic, as in most cases we have very little prior knowledge about the data prior to the experiments or data analysis. Furthermore, the dataset from irrelevant domains could misrepresent the real situation.
Unknown Data Source [3, 7, 13, 22-24]	Many studies do not mention the source or the details of the datasets that they used. A study involving unknown data source make itself impossible to be followed and improved by other researchers, which seriously impedes the development of this research area.

5.3 Current Challenges

Considering the datasets used in the prior studies, except of non-publicity of the dataset, there are three other main obstacles for the research of micro learning, as summarized in Table 2.

Moreover, the datasets from different sources are isolated. This is not brought to sufficient attention from previous studies. Unlike research in some other domains, which often have standard datasets such as ImageNet[2] for object recognition; due to the vague information of learning resources and different curriculum structures, the datasets from different online learning platforms are isolated, making it challenging for next research to reuse the existing data source. For example, the study [8] captured the textual information from video content, and [14] used the co-occurrence information to boost the collaborative filtering result. The textual information is useful in mining semantic information among the learning resources, which may further boost the recommender system as proposed in the study [14]. However, because of the different sources, these two datasets cannot be fused directly.

Although there are initiatives to push a non-profit sharing of research-oriented MOOC data [34], unfortunately, most data from several learning platforms (e.g., edX and Coursera[3]) are still partially open to researchers, or merely open to their partners. Hence, most research teams can but get access to very limited datasets. Researchers demand more complete and diverse data to drive the decision-making system of the

[2] http://www.image-net.org/

[3] https://www.coursera.org/

online learning service. Hence, effective data fusion is another gap at present and worth for future research.

6 Conclusion and Future Research

In this paper, we discuss and review different datasets used in the representative prior studies on e-learning. For the different processing stage of a micro learning system, the requirements of the data types vary a lot. User's historical rating of learning materials is an essential factor for constructing a recommender system. However, the segmentation and annotation of learning materials rely on the learning content and the demographic information. Other performance prediction and analysis models are primarily based on the user's historical learning records. As discussed above, insufficiency, inappropriateness, and non-publicity of the datasets, as well as the difficulty of data fusion are the main challenges that we are facing and need to deal with.

As most models involved in the proposed system are data-driven, the idea behinds these optimization and analysis strategies have a significant overlap with the other data-driven research topics in the TEL domain. Even though this paper is under the topic of micro learning, many views derived from discussion and analysis of this paper can be extended to other e-learning related research topics. We expect this paper can also support the future research of other TEL related studies. Furthermore, this paper calls for efforts on the construction of effective public datasets for the research of TEL.

Acknowledgement

This research has been carried out with the support of the Australian Research Council Discovery Project, DP180101051, and Natural Science Foundation of China, no. 61877051, and UGPN RCF 2018-2019 project between University of Wollongong and University of Surrey.

References

[1] R. Ferguson, "The state of learning analytics in 2012: A review and future challenges". Knowledge Media Institute, Technical Report KMI-2012-01, 2012.

[2] G. Sun, T. Cui, J. Yong, J. Shen and S. Chen, "Mlaas: A cloud-based system for delivering adaptive micro learning in mobile mooc learning". IEEE Transactions on Services Computing, 2018. **11**(2), p. 292-305.

[3] M. Hendez and H. Achour. "Keywords extraction for automatic indexing of e-learning resources". in Computer Applications & Research (WSCAR), 2014 World Symposium on. 2014. IEEE, p. 1-5.

[4] X. Du, F. Zhang, M. Zhang, S. Xu and M. Liu. "Research on result integration mechanism based on crowd wisdom to achieve the correlation of resources and knowledge points". in International Conference on Innovative Technologies and Learning. 2018. Springer, p. 568-577.

[5] K. Verbert, N. Manouselis, X. Ochoa, M. Wolpers, H. Drachsler, I. Bosnic and E. Duval, "Context-aware recommender systems for learning: A survey and future challenges". IEEE Transactions on Learning Technologies, 2012. **5**(4), p. 318-335.
[6] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel and R. Koper, Recommender systems in technology enhanced learning, in Recommender systems handbook. 2011, Springer. p. 387-415.
[7] W. Chen, Z. Niu, X. Zhao and Y. Li, "A hybrid recommendation algorithm adapted in e-learning environments". World Wide Web, 2014. **17**(2), p. 271-284.
[8] H. Yang and C. Meinel, "Content based lecture video retrieval using speech and video text information". IEEE Transactions on Learning Technologies, 2014(2), p. 142-154.
[9] D. Dessì, G. Fenu, M. Marras and D. R. Recupero, "Bridging learning analytics and cognitive computing for big data classification in micro-learning video collections". Computers in Human Behavior, 2018. p. 1-10.
[10] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos and R. C. Miller. "Understanding in-video dropouts and interaction peaks inonline lecture videos". in Proceedings of the first ACM conference on Learning@ scale conference. 2014. ACM, p. 31-40.
[11] E. F. Risko, T. Foulsham, S. Dawson and A. Kingstone, "The collaborative lecture annotation system (clas): A new tool for distributed learning". IEEE Transactions on Learning Technologies, 2013. **6**(1), p. 4-13.
[12] P. Welinder, S. Branson, P. Perona and S. J. Belongie. "The multidimensional wisdom of crowds". in Advances in neural information processing systems. 2010. p. 2424-2432.
[13] D. Cernea, E. Del Moral and J. Gayo, "Soaf: Semantic indexing system based on collaborative tagging". Interdisciplinary Journal of E-Learning and Learning Objects, 2008. **4**(1), p. 137-149.
[14] K. Niemann and M. Wolpers. "Usage context-boosted filtering for recommender systems in tel". in European Conference on Technology Enhanced Learning. 2013. Springer, p. 246-259.
[15] J. Shu, X. Shen, H. Liu, B. Yi and Z. Zhang, "A content-based recommendation algorithm for learning resources". Multimedia Systems, 2018. **24**(2), p. 163-173.
[16] C.-N. Ziegler, S. M. McNee, J. A. Konstan and G. Lausen. "Improving recommendation lists through topic diversification". in Proceedings of the 14th international conference on World Wide Web. 2005. ACM, p. 22-32.
[17] Q. Zhao, Y. Zhang and J. Chen. "An improved ant colony optimization algorithm for recommendation of micro-learning path". in Computer and Information Technology (CIT), 2016 IEEE International Conference on. 2016. IEEE, p. 190-196.
[18] M. Chen, M. Tong, C. Liu, M. Han and Y. Xia. "Recommendation of learning path using an improved aco based on novel coordinate system". in Advanced Applied Informatics (IIAI-AAI), 2017 6th IIAI International Congress on. 2017. IEEE, p. 747-753.
[19] Y. Zhou, C. Huang, Q. Hu, J. Zhu and Y. Tang, "Personalized learning full-path recommendation model based on lstm neural networks". Information Sciences, 2018. **444**, p. 135-152.
[20] D. Wu, J. Lu and G. Zhang, "A fuzzy tree matching-based personalized e-learning recommender system". IEEE Transactions on Fuzzy Systems, 2015. **23**(6), p. 2412-2426.
[21] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context". Acm transactions on interactive intelligent systems (tiis), 2016. **5**(4), p. 19.
[22] G. Fenza, F. Orciuoli and D. G. Sampson. "Building adaptive tutoring model using artificial neural networks and reinforcement learning". in Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on. 2017. IEEE, p. 460-462.
[23] A. Al-Hmouz, J. Shen, R. Al-Hmouz and J. Yan, "Modeling and simulation of an adaptive neuro-fuzzy inference system (anfis) for mobile learning". IEEE Transactions on Learning Technologies, 2012. **5**(3), p. 226-237.

[24] F. A. Dorça, V. C. Carvalho, M. M. Mendes, R. D. Araújo, H. N. Ferreira and R. G. Cattelan. "An approach for automatic and dynamic analysis of learning objects repositories through ontologies and data mining techniques for supporting personalized recommendation of content in adaptive and intelligent educational systems". in 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT). 2017. IEEE, p. 514-516.

[25] T.-Y. Yang, C. G. Brinton, C. Joe-Wong and M. Chiang, "Behavior-based grade prediction for moocs via time series neural networks". IEEE Journal of Selected Topics in Signal Processing, 2017. **11**(5), p. 716-728.

[26] C. G. Brinton, S. Buccapatnam, M. Chiang and H. V. Poor, "Mining mooc clickstreams: On the relationship between learner behavior and performance". arXiv preprint arXiv:1503.06489, 2015.

[27] C. G. Brinton and M. Chiang. "Mooc performance prediction via clickstream data and social learning networks". in Computer Communications (INFOCOM), 2015 IEEE Conference on. 2015. IEEE, p. 2299-2307.

[28] M. Chiang. Networks: Friends, money, and bytes. Available from: https://www.coursera.org/course/friendsmoneybytes. Princeton University 2012;

[29] C. Brinton and M. Chiang. Networks illustrated: Principles without calculus. Princeton University 2013; Available from: https://www.coursera.org/learn/networks-illustrated.

[30] G. Kórösi, P. Esztelecki, R. Farkas and K. Tóth. "Clickstream-based outcome prediction in short video moocs". in 2018 International Conference on Computer, Information and Telecommunication Systems (CITS). 2018. IEEE, p. 1-5.

[31] M. Shridharan, A. Willingham, J. Spencer, T.-Y. Yang and C. Brinton. "Predictive learning analytics for video-watching behavior in moocs". in Information Sciences and Systems (CISS), 2018 52nd Annual Conference on. 2018. IEEE, p. 1-6.

[32] T. Sinha, P. Jermann, N. Li and P. Dillenbourg, "Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions". arXiv preprint arXiv:1407.7131, 2014.

[33] M. Odersky. Functional programming principles in scala. 2012; Available from: https://www.coursera.org/learn/progfun1.

[34] G. Lopez, D. T. Seaton, A. Ang, D. Tingley and I. Chuang. "Google bigquery for education: Framework for parsing and analyzing edx mooc data". in Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale. 2017. ACM, p. 181-184.