A Time-Aware Approach to Detect Patterns and Predict Help-Seeking Behaviour in Adaptive Educational Systems

Raquel Horta-Bartomeu
Universidad Nacional de Educación a Distancia (UNED)
rhorta3@alumno.uned.es

Olga C. Santos
aDeNu Research group
Universidad Nacional de Educación a Distancia (UNED)
ocsantos@dia.uned.es

ABSTRACT

In distance education and some computer-assisted learning scenarios asking for help when needed is important. Some students do not ask for help even when they do not know how to proceed. In situations where a teacher is not present, this can be a serious setback. We aim to find an approach to learn about students' help-seeking behaviour by studying sequences of actions that end with the student asking for help. The goal is to be able to recognize those students who need help but fail to ask for it and offer them assistance. We propose to include the temporal context of user-platform interaction and suggest an ensemble model to learn from both general and personal tendencies.

Categories and Subject Descriptors

K.3.1 [**Computers and Education**]: Computer Uses in Education; I.2.6 [**Artificial Intelligence**]: Learning; I.5.4 [**Pattern recognition**]: Applications; G.3 [**Probability and Statistics**]: Markov processes, Time series analysis

Keywords

Adaptive systems, time series, educational data mining, personalized education

1. INTRODUCTION

Researchers have found help-seeking to be important in learning scenarios and observed that some students do not reach for help when they need it [4, 25]. When a teacher is not always present, and the student needs some level of self-discipline, not asking for help might be problematic as the student could end up wheel-spinning [18] or abandoning the task. The longitudinal nature of student-platform interactions leads us to think that taking into account the temporal context could be useful for analysing help-seeking behaviour. There is literature on help-seeking including temporal data, and some works have focused on performance prediction or have centred on specific knowledge topics. However, we have not found work that focuses on the behaviour around help-seeking actions, including temporal data and independent of student knowledge and task content. In this Master Thesis, we propose to represent student-platform interactions as sequences of actions, study whether sequential patterns exist in students' help-seeking behaviour and explore whether a prediction model could identify students that need help but do not ask for it.

2. RELATED RESEARCH

Time series studies are very common in natural sciences and some social sciences. Studies that make use of time series data can also be found in the field of educational sciences [5, 8, 9, 13, 14, 15, 17, 29]. We have reviewed existing works on both help-seeking behaviour and time series data analysis. In section 3 we highlight the specific differences between the works exposed here and what we propose to do.

2.1 Help-seeking behaviour

Knowing when to ask for help is important [4, 11]. [10, 11] agreed that it could improve resilience and efficacy. According to [10], help-seeking has been studied for years but the rise of new technologies opens new research opportunities on help-seeking in these new contexts.

Some works have focused on detecting specific situations that are known to be problematic. For instance, in classes where the teacher has more students than desired, it might be difficult for them to identify students who need help or are stuck. [18] developed a method using machine learning (ML) models to automatically predict wheel-spinning and decide how to intervene. [4] attacked both problems of asking for help too much and not enough by negotiating with the student. Rather than using ML models, they predefined a set of heuristics. A slightly different situation was studied by [32]. Their goal was to find a connection between student procrastination (i.e. intentionally delaying work) and their activities within different learning materials. They used data from a massive open online course (MOOC) platform and found two main study strategies: students who delayed work worked intensively for short periods followed by long pauses, while students who did not delay usually split the tasks into subtasks and worked more constantly but less intensively.

Other works have focused on knowledge tracing [6, 7, 24], however, we will not be considering student knowledge but their behaviour and interaction with the educational system.

2.2 Pattern recognition and sequence prediction

To cluster categorical sequences, one needs to define the function or method used to compare the sequences pairwise, that is, how to measure the distance between them. [5] analyzed activity frequency through the length of different online courses to study if different activity patterns were related to student performance. They used agglomerative hierarchical clustering (AHC) with the Levenshtein distance. [17] used a similar approach and found patterns in group problem-solving strategies analysing group behaviour of students working on interactive tabletops. [8] also used AHC with the Levenshtein distance and found 3 groups of similar study state sequences using data from a drill-and-practice learning environment in college mathematics.

[13] found that a large subgroup of MOOC participants might have been engaging by watching video lectures without doing the assignments. Their methodology consisted in constructing, for each student, vectors of states representing their engagement trajectories through the course. They computed the distance between trajectories by assigning a numerical value to each label and calculating the L1 norm.

[14] proposed a method that would capture the clusters' number and size evolution over time. They transformed log data sequences into Markov chain models. Then, they computed the pairwise similarities by computing the expected transition probabilities using the stationary distribution over the actions. They used the Jensen-Shannon divergence and the Hellinger distance between the expected transition frequencies of the Markov chains (more details in [21], as cited in [14]). They used k-means with an evolutionary clustering method that tracks the evolution of the similarities over time by smoothing the similarity matrices ([31] as cited in [14]). [9] also modelled student behaviour using Markov chains. They randomly generated Markov chain priors and assigned each sequence to the prior most likely to generate it. Then, each prior would be updated to the Markov chain generated using its associated sequences. These last two steps were repeated until less than 5% of the sequences would change their prior. As they stated, this method is similar to k-means but with the clustering being dependent on the Markov chains instead of on a similarity measure performed directly on the sequences.

[29] was able to detect unprofitable learning experiences and predict student performance by using time series data. They used dynamic time warping (DTW) to measure the distance between sequences and performed hierarchical clustering to find clusters. DTW was proved to be useful; however, to the best of our knowledge, it is not suitable for categorical sequences but only for numerical ones and has therefore been ruled out as a possible approach to our specific problem.

Artificial neural networks (ANNs) are known to be useful for a vast variety of tasks. When it comes to time series, the most used ones seem to be recurrent neural networks (RNN) and long short-term memory networks (LSTM). RNNs' main limitation is their difficulty to work with long sequences due to a vanishing gradient problem [12]. While LSTMs solve this issue, they usually take quite a long to train and have difficulties in capturing long-term dependencies in long sequences [12, 30]. Finally, a novel approach called transformer networks was introduced by [30]. This approach introduces what the authors called an *attention mechanism*, which solves the long-term dependency problem in LSTMs. [2] used LSTMs in a multi-module system to analyse the relationship between intent and user actions in interactive systems. [16] used time-aware LSTMs (T-LSTM), a special type of LSTM that can handle time irregularities, to model student knowledge state in continuous time. They conducted an empirical experiment and discovered that they outperformed regular LSTMs, logistic regression and recent temporal pattern mining (RTPs). [15] used RTPs along with support vector machine (SVM) and logistic regression to predict student performance and detect the need for intervention using students' answers to programming exercises. They were able to classify students within only 1 minute into the exercise.

3. EXPECTED CONTRIBUTION

To our knowledge, this would be the first work to use student-platform interaction data in form of sequences of actions to predict help-seeking behaviour while being independent of the topic being taught. Our approach differs from existing work by joining three main aspects. First, **help-seeking behaviour**: we have found works that linked the need for intervention with performance or student knowledge [6, 15, 16] instead of analyzing the behaviour surrounding actual help-seeking actions. Second, **time-awareness**: we have found works that have used cumulative data (e.g. number of attempts) to predict the need for intervention [18] but did not take into account the temporal context. Third, **topic-independence**: we have found works that did take into account the temporal context but focused on the content of student answers for specific topics [15, 23].

The approach we propose would include the temporal context, would not be dependent on the nature of the content being taught and would focus on user-platform interaction (i.e. clicking, typing, deleting, consulting theory, etc).

While this research is still in the early stages, we believe in the importance of students' affective state [26, 27, 28] and might consider including the affective context if possible. Finally, if we were to find successful results, we believe that richer predictions could be obtained by joining student knowledge information [6, 15, 16] along with the information learnt from help-seeking behaviour. However, this is out of the scope of this research.

4. RESEARCH QUESTIONS

We aim to study whether our proposal would be feasible and for that we present two research questions:

Q1 Are there temporal patterns in students help-seeking behaviour?

Q2 Can temporal student-platform interaction data be used to detect students who need help but do not ask for it?

5. PROPOSED METHODOLOGY

We will be dealing with both supervised and unsupervised problems: we will be using clustering algorithms towards answering Q1 and prediction algorithms towards answering

Q2. We propose to perform clustering (Q1) as a preliminary step to a more complex system (Q2). Clustering can lead to interpretable results and reveal information that could be useful to pedagogical experts while some prediction methods are more powerful but may act as a black box. As well as considering less interpretable methods, the system proposed in Q2 addresses personalization. We expose the methodology we intend to follow, and the methods we have considered so far.

5.1 Data

The dataset to be used is yet to be found or constructed. Efforts are being made to find a suitable dataset. Some promising options are being considered but are yet to be confirmed. Even though, the characteristics that we look for in a dataset have been defined. The dataset should contain action logs that originated from the interaction between a student and a learning platform that has some kind of help tool that the student can choose to use. Each log should include, at least: (1) action type, (2) action start time, (3) action end time, (4) student identification (anonymized) and (4) exercise identification.

Given that some actions are continuous rather than instantaneous, we will need to decide how to represent this characteristic. As an example, a student might consult the theory section of the system just for 10 seconds, or they could spend 5 minutes consulting the content. It would be desirable that those two cases were not represented in the same way and that duration was taken into account. When using Markov chains, if we consider action durations, the probabilities of staying in the same state will always be 0, and the duration would not be taken into account. To solve this, we could consider splitting the actions into time slots. We will need to take into account that some other actions might be instant actions, with practically no duration, e.g. submitting an exercise. We will need to make sure that the model we use does not undermine these actions. Finally, if possible, we might consider including *idle* actions, that is, time in which the student does nothing.

5.2 Clustering

To answer **Q1**, we encounter two main decisions: how to determine the distance between sequences and which clustering algorithm to use.

5.2.1 Distance between the sequences

The main challenge of dealing with sequential data is that they cannot be directly fed to traditional clustering algorithms. First, one needs to decide how to represent the sequences and define how to compare them. We have decided to try two methods for representing the distance between sequences: Markov chains and the Levenshtein distance. Markov chains represent a sequence by considering the probability of going from one state (i.e. action) to another. The basic form of a Markov chain only considers the current state to predict the next one. This could be a limitation and therefore n-order Markov chains could be considered. In a Markov chain of order n, n previous steps are taken into account. On the other hand, the Levenshtein distance is a type of edit distance, that is, the minimum changes required to transform one sequence into another. The Levenshtein distance considers insertions, deletions and substitutions.

5.2.2 Clustering algorithms

Taking into account existing work, we have narrowed the search for a clustering method down to two: hierarchical clustering and k-means. The main drawback of k-means is the requirement of a predefined number of clusters, which in our case is unknown. Hierarchical clustering has the advantage that the number of clusters can be chosen a posteriori, however, it can be expensive when dealing with large datasets. K-means is usually a fast algorithm, although it might depend on the chosen distance metric [19].

5.3 Prediction

Towards answering **Q2**, we propose a prediction system; its characteristics are presented in this section.

5.3.1 System structure

While we want to take advantage of how students in general behave, we want to provide a personalized learning experience. To do so, a student's personal traits and tendencies must be taken into account. Therefore, we aim to take advantage of the general traits of student behaviour while preserving the personal study tendencies of each student, thus combining an inter-subject with an intra-subject approach. To achieve this goal, we propose an ensemble system composed of three blocks. We name the system SHEmblE (Sequence analysis of Help-seeking behaviour with an ensEMBLE model for Educational systems)

Firstly, we will have a prediction model that will be trained with all the available data. We will refer to this model as the *common model* as it will be shared among all students. We expect it to be able to learn the general patterns of help-seeking behaviour if those exist.

Secondly, we will have what we call a *personal model*. Each student will have each own personal model trained with their own data, if any. We expect this model to be able to learn the personal tendencies and preferences of a student.

Finally, a third model will combine the predictions of common and personal models. We call this model the *ensemble model* and we expect it to learn how to combine the predictions the best way possible.

We will focus on students who regularly ask for help for the general model. However, from those students, we will take into account interactions that exhibit help-seeking behaviour as well as those in which the student does not need help to successfully reach their goal. Sequences from students who never ask for help will not be included as we cannot know if the student did really not need assistance, or they simply never ask for it.

We are aware that data size will be a concern regarding the personal model. Its goal is to provide individualization, and thus, we believe it is an important part of the system [1, 7, 22]. Therefore, the ensemble model could take into account the amount of data with which the personal model was trained in order to weigh the predictions properly.

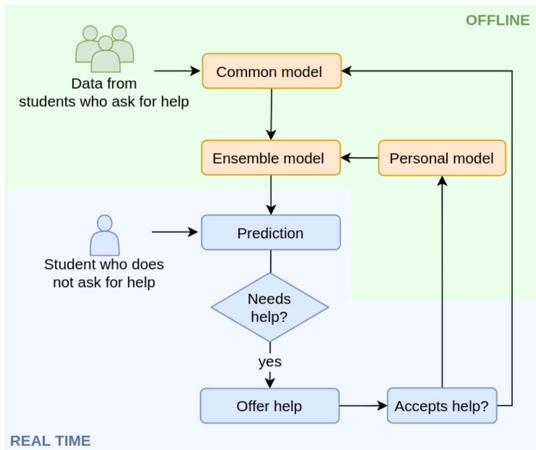

Figure 1: General schema of SHEmblE, the proposed system.

The ultimate goal, if this system was to achieve good results, would be to implement it in a real educational system. Figure 1 represents the overall structure of the proposed system. The idea would be to be able to detect, in real-time, students that need help and offer it to them. Apart from collecting logs from students that ask for help themselves, whenever we offer help we would save their response as well. The scope of this work comprises the common, personal and ensemble models, the rest could be the object of study of future research.

5.3.2 Prediction algorithms

Time series prediction has been the challenge of many works in literature. This work deals with categorical time series, in other words, categorical sequences. It has been narrowed down to three methods: artificial neural networks, hidden Markov models, and recent temporal patterns. As exposed in section 2, ANNs have been used in problems involving time series data and showed promising results, different types found in the literature will be considered (e.g. LSTM, T-LSTM, transformers). HMMs have also been useful for predicting and classifying action sequences. [20] found that HMMs needed fewer training samples and less CPU time while performing similar to LSTMs. Finally, RTPs [3] have been successful at similar tasks. [15] used them and managed to detect students that needed intervention only one minute after starting an exercise. While their data consisted of attributes of the students' answers' content and ours will consist of interaction data, we believe that a similar approach could be applied to our particular task.

5.3.3 System training and evaluation

The system we propose is going to be composed of three different models. These models will be evaluated independently and altogether. We intend to evaluate the common model by performing a variation of the leave-one-out cross-validation (LOOCV) in which in each iteration the whole data of one student is left out for validation. Regarding the personal model, the dataset will be split by the sequences' student id and for each student, a LOOCV will be performed. The performance of the personal model will be assessed by combining all the performances (eg. mean and standard deviation) and special attention will be paid to possible outliers. Finally, the ensemble model block will need to be fed the predictions of the other two models. Therefore, a whole new dataset E will need to be constructed such that:

- Consider the set of n students $S = \{s_i | i \in \{1..n\}\}$.
- Each student s_i has got m_i sequences $Q_i = \{q_{ij} | j \in \{1..m_i\}\}$
- The instance e_{ij} will correspond to the sequence j of the student i and will contain at least 2 features:
  - The output of the common model trained using the sequences from students other than s_i.
  - The output by the personal model trained using the sequences of student s_i other than q_{ij}.

  Moreover, additional features could be added, such as the size of the dataset used to train the personal model, given that some students might have few or no data.
- The dataset E will then have $\sum_{i=1}^{n} m_i$ rows.

Once the dataset has been constructed, k-fold cross-validation can be performed. Figure 2 shows a graphical representation of the proposed evaluation method.

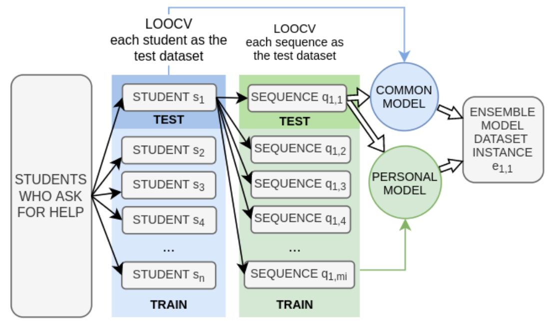

Figure 2: Graphical representation of the evaluation scheme and the generation of the dataset for the ensemble model.

6. CONCLUSIONS

We have not found works that aim to detect students who need help by analysing behaviour around help-seeking actions using time-aware user-platform interaction data. In this Master Thesis, we aim to study whether such data can be useful to predict help-request actions and propose an ensemble system that combines a shared model and a personal model so as to achieve individualization.

This work is still at a very early stage. Any feedback and ideas on this proposal are very much welcomed. Specifically, comments on the sequence representation, and the clustering and predictive model choices will be appreciated.

7. ACKNOWLEDGEMENTS

The work is partially supported by UNED's master's degree in Research in AI and the project INT2AFF funded under Grant PGC2018-102279-B-I00 (MCIU/AEI/FEDER, UE) by the Spanish Ministry of Science, Innovation and Universities, the Spanish Agency of Research and the European Regional Development Fund (ERDF).

8. REFERENCES

[1] M. Abell. Individualizing learning using intelligent technology and universally designed curriculum. *The journal of technology, learning and assessment*, 5(3), 2006.

[2] R. Agrawal, A. Habeeb, and C. Hsueh. Learning user intent from action sequences on interactive systems. In *Thirty-Second AAAI Conference on Artificial Intelligence*, page 59–64, New Orleans, Louisiana, USA, February 2-7 2018.

[3] I. Batal, D. Fradkin, J. Harrison, F. Moerchen, and M. Hauskrecht. Mining recent temporal patterns for event detection in multivariate time series data. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 280–288, 2012.

[4] C.-Y. Chou, K. R. Lai, P.-Y. Chao, S.-F. Tseng, and T.-Y. Liao. A negotiation-based adaptive learning system for regulating help-seeking behaviors. *Computers & Education*, 126:115–128, 2018. ID: 271849.

[5] R. Conijn and M. V. Zaanen. Trends in student behavior in online courses. In *3rd International Conference on Higher Education Advances*, pages 649–657, 2017.

[6] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

[7] A. T. Corbett, J. R. Anderson, V. H. Carver, and S. A. Brancolini. Individual differences and predictive validity in student modeling. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 1994.

[8] M. Desmarais and F. Lemieux. Clustering and visualizing study state sequences. In *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*, pages 224–227, 2013.

[9] C. Hansen, C. Hansen, N. O. D. Hjuler, S. Alstrup, and C. Lioma. Sequence modeling for analysing student interaction with educational systems. In *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017*, pages 232–237. International Educational Data Mining Society (IEDMS), 2017.

[10] S. Järvelä. How does help seeking help?–new prospects in a variety of contexts. *Learning and Instruction*, 21(2):297–299, 2011.

[11] S. A. Karabenick and R. S. Newman. *Help Seeking in Academic Settings: Goals, Groups, and Contexts.* Lawrence Erlbaum Associates, Inc, 2006.

[12] F. Karim, S. Majumdar, H. Darabi, and S. Chen. Lstm fully convolutional networks for time series classification. *IEEE access*, 6:1662–1669, 2017.

[13] R. Kizilcec, C. Piech, and E. Schneider. Deconstructing disengagement. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 170–179. ACM, 2013.

[14] S. Klingler, T. Käser, B. Solenthaler, and M. Gross. Temporally coherent clustering of student data. In *Proceedings of the 9th International Conference on Educational Data Mining (EDM 2016)*, pages 202–209, 2016.

[15] Y. Mao. One minute is enough: Early prediction of student success and event-level difficulty during novice programming tasks. In *Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019)*, pages 119–128, 2019.

[16] Y. Mao, S. Marwan, T. W. Price, T. Barnes, and M. Chi. What time is it? student modeling needs to know. In *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, pages 171–182, 2020.

[17] R. Martinez, K. Yacef, J. Kay, A. Al-Qaraghuli, and A. Kharrufa. Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In *4th International Conference on Educational Data Mining, EDM 2011*, pages 111–120. CEUR-WS, 2011.

[18] T. Mu, A. Jetten, and E. Brunskill. Towards suggesting actionable interventions for wheel-spinning students. In *13th International Conference on Educational Data Mining (EDM 2020)*, pages 183–193, 2020.

[19] E. Ofitserov, V. Tsvetkov, and V. Nazarov. Soft edit distance for differentiable comparison of symbolic sequences. 2019. Preprint available at https://arxiv.org/abs/1904.12562.

[20] M. Panzner and P. Cimiano. Comparing hidden markov models and long short term memory neural networks for learning action representations. In *International Workshop on Machine Learning, Optimization, and Big Data*, pages 94–105. Springer, 2016.

[21] L. Pardo. *Statistical inference based on divergence measures.* Chapman and Hall / CRC Press, 2018.

[22] Z. A. Pardos and N. T. Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 255–266. Springer, 2010.

[23] C. Piech, J. Huang, A. Nguyen, M. Phulsuksombati, M. Sahami, and L. Guibas. Learning program embeddings to propagate feedback on student code. In *International conference on machine Learning*, pages 1093–1102. PMLR, 2015.

[24] C. Piech, J. Spencer, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein. Deep knowledge tracing. In *Proceedings of Advances in Neural Information Processing Systems 28 (NIPS 2015)*, 2015.

[25] I. Roll, R. S. d. Baker, V. Aleven, and K. R. Koedinger. On the benefits of seeking (and avoiding) help in online problem-solving environments. *Journal of the Learning Sciences*, 23(4):537–560, 2014.

[26] S. Salmeron-Majadas, R. S. Baker, O. C. Santos, and J. G. Boticario. A machine learning approach to leverage individual keyboard and mouse interaction behavior from multiple users in real-world learning scenarios. *IEEE Access*, 6:39154–39179, 2018.

[27] S. Salmeron-Majadas, O. C. Santos, and J. G. Boticario. An evaluation of mouse and keyboard interaction indicators towards non-intrusive and low cost affective modeling in an educational context.

Procedia Computer Science, 35:691–700, 2014.

[28] O. C. Santos. *Emotions and personality in adaptive e-learning systems: an affective computing perspective*, pages 263–285. Emotions and personality in personalized services. Springer, 2016.

[29] S. Shen and M. Chi. Clustering student sequential trajectories using dynamic time warping. In *Proceedings of the 10th International Conference on Educational Data Mining (EDM 2017)*, pages 266–271, 2017.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.

[31] K. S. Xu, M. Kliger, and A. O. H. Iii. Adaptive evolutionary clustering. *Data Mining and Knowledge Discovery*, 28(2):304–336, 2014.

[32] M. Yao, S. Sahebi, and R. F. Behnagh. Analyzing student procrastination in moocs: A multivariate hawkes approach. In *13th International Conference on Educational Data Mining (EDM 2020)*, pages 280–291, 2020.