Urgency Analysis of Learners' Comments: an Automated Intervention Priority Model for MOOC

Laila Alrajhi[1], Ahmed Alamri[1], Filipe Dwan Pereira [2] and Alexandra I. Cristea[1]

[1] Computer Science, Durham University, Durham, UK,
[2] Computer Science, Federal University of Roraima, Boa Vista, Brazil
[1]{laila.m.alrajhi, ahmed.s.alamri, alexandra.i.cristea}@durham.ac.uk
[2]filipe.dwan@ufrr.br

Abstract. Recently, the growing number of learners in Massive Open Online Course (MOOC) environments generate a vast amount of online comments via social interactions, general discussions, expressing feelings or asking for help. Concomitantly, *learner dropout*, at any time during MOOC courses, is very high, whilst the number of learners completing (*completers*) is low. Urgent intervention and attention may alleviate this problem. Analysing and mining learner comments is a fundamental step towards understanding their need for intervention from instructors. Here, we explore a dataset from a FutureLearn MOOC course. We find that (1) learners who write many comments that need urgent intervention tend to write many comments, in general. (2) The motivation to access more steps (i.e., learning resources) is higher in learners without many comments needing intervention, than that of learners needing intervention. (3) Learners who have many comments that need intervention are less likely to complete the course (13%). Therefore, we propose a *new priority model for the urgency of intervention* built on learner histories – past urgency, sentiment analysis and step access.

Keywords: MOOCs, FutureLearn, Comments, Priority in Intervention.

1 Introduction

Today, with the successful development of MOOC environments, they are playing a vital role in education. In an online world, learners can access knowledge and numerous high-quality resources [1]. This attracts a large learner cohort with different abilities. At the same time, the dropout rate is high enough to be a serious problem. There are many reasons for dropping out, including *learners' need for instructor intervention* [2].

MOOC platforms have an asynchronous discussion forum tool that provides a venue for learners to communicate with others [3]. It is a crucial component and can be utilised in different ways, involving social interaction, discussion, or as an essential part of a teaching strategy [4]. Also, it is the main communication tool between learner and instructor [5] for feedback, support, and encouragement [6].

Instructors' interventions are an essential teaching activity in MOOCs, to help learners [7]. However, due to the high ratio of learners-to-instructors, it is very hard to monitor all learners' comments. Thus, the problem of detecting urgent posts has stirred researchers to solutions primarily framed towards a text classification problem [8] [9] [10]. However, such approaches did not consider the study of the learner's behaviour.

We conjecture it as essential to understand learners' behaviours before proposing intervention. Hence, after analysing the distribution of comments that need intervention, we additionally explore the relation between high-frequency commenters and their behaviours, in terms of their access and completion rates. We define *high-frequency (HF) commenters* as learners who have many comments that need intervention, and formulate the following research questions:

RQ1: What is the behaviour of learners who need an urgent intervention?

- *RQ1.1: Is there a relationship between the number of comments written by the learners that need urgent intervention and the average number of comments?*
- *RQ1.2: Is there a relation between high-frequency (HF) commenters and their number of steps accessed?*
- *RQ1.3: Is there a relation between HF commenter number and completion-rates?*

RQ2: Can we design an effective intervention priority framework based on behaviour?

2 Related Work

Before the era of MOOCs, researchers were already analysing the need of *instructor intervention* in discussion forums in asynchronous virtual learning environments [11]. Recently, instructor intervention is one of the hot research directions for MOOCs [12]. The most common approaches focused on the use of text classification methods [8] [9] [10] [13]. Some were based only on Natural Language Processing (NLP), others involved other features. They deployed different types of machine learning algorithms (shallow and deep neural networks models). Other relevant attempts [14] [15] [16] predicted urgency as one of three different tasks (confusion, sentiment and urgency), but they also involved only text-based methods.

In [17], the instructor intervention problem in MOOC forums was tackled by using the sequence of posts and combined features from these posts. They considered instructor posts as intervention. Chandrasekaran et. al. [18] proposed several studies on instructor intervention in Coursera forums. For instance, [19] proposed a taxonomy of pedagogical interventions for automated guidance to instructors. Moreover, [20] investigated discourse relations and used PDTB (Penn Discourse Treebank) based features to predict the need for instructor interventions. For position bias in intervention context [5] they showed that there is a bias in instructor intervention. They improved intervention classifier performance when they removed bias from the training data. In [3] they studied instructor intervention based on a deep learning model, and thread structure.

While these works provide solutions to the instructor's intervention problem, learner behaviours' relation to urgent intervention need remains unstudied. Specifically, we want to analyse how HF commenters behave on MOOC platforms. The main idea in this paper is enhancing intervention by prioritising it as an intelligent filtering system.

This priority is generated based on learner behaviour. To the best of our knowledge, the priority in intervention shown in this paper has not been seen before in the literature.

3 Methodology

This section presents our dataset and methodologies.

3.1 Dataset

The raw corpus dataset we utilised was provided by the FutureLearn platform [21], namely, the 'Big Data' course, Run 2. The course was conducted during 2016 on an over 9 weeks scale and, a.o., it contains English comments text. We then focus only on the first half of the course (5 weeks), with its subset of 5790 comments. This is done as early intervention on urgent comments is considered more appropriate than late intervention, as most learners tend to drop out in the first stages of the course [22] [23].

Gold Standard Corpus Creation. The collected 5790 text comments were manually labelled to assign urgency and they were annotated by domain experts. From these experts, two are instructors at the Department of Computer Science at the University; in addition, one is an author of this paper. We gave Agrawal et al. [24] instructions to annotators, to manually classify comments onto the urgency scale (1-7), (1: no reason to read the post – 7: extremely urgent: instructor definitely needs to reply). After completing the annotations, we excluded (four) comments containing anything other than (1-7). To validate these labels we used Krippendorff's α' [25]. However, we found that the agreement between these annotators was very low. To alleviate this, we converted the scale to binary (1 to 3→0 , 4 to 7→1). Then, we applied a voting process between the three annotators, resulting in a binary-class label as: 0 → Non-Urgent; 1 → Urgent. As this is real data, possibly unsurprisingly, the resulting data is biased towards the (Non-Urgent) class, with 883 comments as Urgent (15%) and the rest as Non-Urgent.

Dataset Statistics. The 5786 comments were created by 873 unique learners (commenters) in 5 weeks. Number of steps and comments per week appear in Table 1.

Table 1. Statistics of the gold standard corpus.

Week	# of steps (week)	# of comments (week)	# of active learners (week)	Average comments per learner (week)
1	11	2130	749	2.84
2	12	1600	419	3.81
3	15	1123	236	4.75
4	11	753	180	4.18
5	4	180	92	1.95

Fig. 1. (left) illustrates the number of comments written over 5 weeks. This number decreased gradually, dropping to 180 comments in the last week, from 2130 comments

in the first week (-99.9%). Every week has a different number of steps to complete. Thus, we also represented the number of comments per steps (Fig. 1, right) on the temporal axis. These numbers oscillate more – showing some topics to trigger more comments than others – although the overall numbers follow the downwards trend.

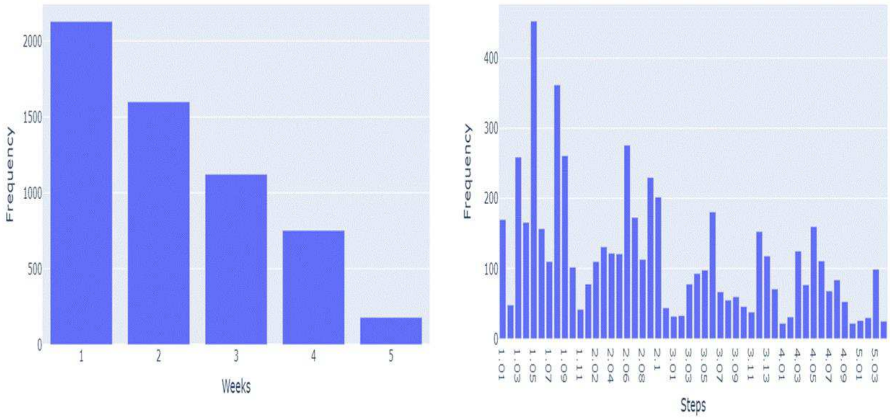

Fig. 1. The number of comments in every week (left) and in every step (right).

Who is, however, writing these comments? To inspect the distribution of the number of, what we call, *active* learners (commenters) who wrote the comments every week and step, we visualised them as shown in Fig. 2.

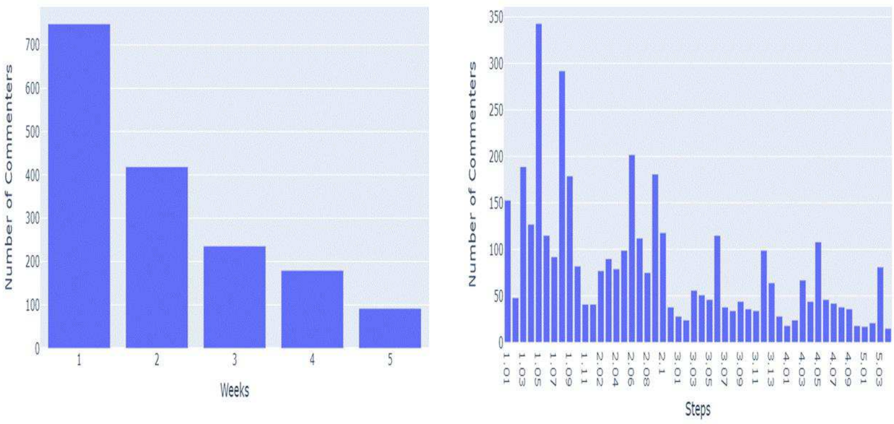

Fig. 2. Active learners (commenters) in every week (left) and in every step (right).

Next, we observed comments that need urgent intervention, to focus on their trend. Hence, we visualised a line graph over the 5 weeks, to explore how urgency changed over time (Fig. 3, left). Overall, the first weeks had a higher percentage of comments needing intervention (Fig. 3, left), drawn from a higher number of comments (Fig. 1, left). The fluctuation from week 4 to 5 is due to the drastic drop in overall comments. We also visualised percentages of urgent comments for every step, (Fig. 3, right), which showed high fluctuation. We further graphically compared results between Urgent and Non-Urgent comments number across (weeks, steps) in Fig. 4 (left, right).

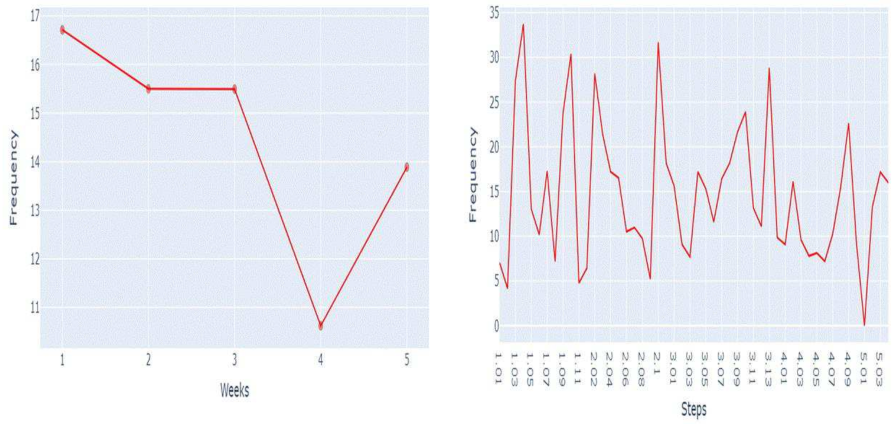

Fig. 3. The percentage of urgent comments for every week (left) and for every step (right).

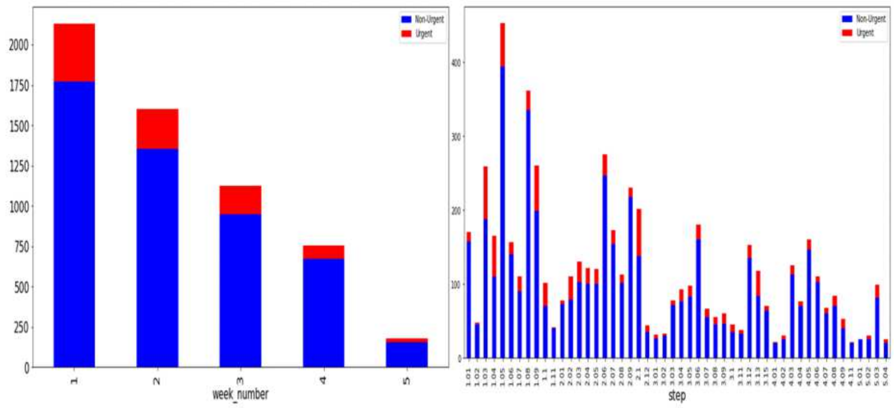

Fig. 4. Comparing Urgent and Non-urgent comment numbers for every week (left) and every step (right).

3.2 Exploring Urgency and Learner Behaviour

As an initial step, to understand learners' behaviour in writing comments, we explored the relationship between the number of comments written by the learners who need urgent intervention with the average number of comments. Then, to explore the effect of urgency on learner behaviour, we explored the relationship between HF commenters and their learning behaviour – here we simply compared it to the number of step accesses. We defined a learner who needs urgent intervention (*HF commenters*) as per equation 1; let *n*: number of comments, *u(c):* urgent comments and *c*: a comment.

$$HF\ Commenters = \frac{\sum_{n=1}^{\infty} u(c)}{\sum_{n=1}^{\infty}(c)} = 1 \qquad (1)$$

We calculated the average number of step access for each group (Non-Urgent) and HF commenters (Urgent) to track how every group behaves on the platform.

Finally, we addressed completers with respect to their need of intervention. We defined completers according to equation 2; where, *total access steps*: number of total access per learner, *total course steps*: total number of steps in a course.

$$Completer = total\ access\ steps \geq total\ course\ steps * 0.80 \qquad (2)$$

We define completers as in Eq. 2 because, in spite of the large number of previous studies, a formal definition of learners dropout is lacking [26]. Therefore, we went with the definition in [23], we defined completers are learners whose number of steps accessed is equal or higher than 80%.

3.3 Priority in Urgent Intervention

In this study we propose a new intervention framework designed to add prioritising to urgent comments based on learners' history, to assist instructors' decision, optimise their time and ability to adapt their intervention. We begin by supposing that, when the instructor intervened, some of these comments were potentially urgent. Then, for these potentially urgent instances we add *priority* (high-, mid- or low), depending on the *learner risk level*. The idea is to focus on learners, understand their behaviours and do a segmentation based on 3 variables (urgency, sentiment analysis and number of accesses).

Our model includes two phases (see Fig. 5), first phase (prediction phase): using a supervised classifier to predict if the comments need a response urgently or not. Second phase (intervention priority phase): takes the output of the previous phase (urgent comments) as input. Then, adds a priority to these comments based on the history of learners who wrote these comments using unsupervised machine learning (clustering). Therefore, based on these groups we assign different priorities to comments.

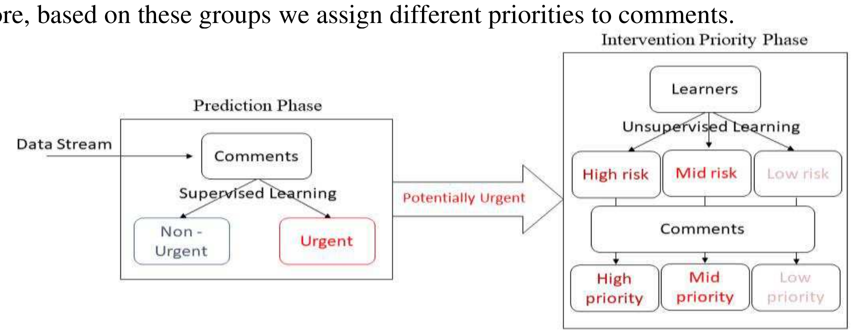

Fig. 5. Priority in urgent intervention framework.

Prediction Phase. Here we apply the state-of-the-art in text classification, Bidirectional Encoder Representations from Transformers (BERT) [27] to predict urgency.

Intervention Priority Phase. We study the behaviour of learners based on three variables (urgency, sentiment analysis and step access). We selected these three variables because they address RQ1.2 and RQ1.3. Moreover, a sentiment analysis study [13] found a negative correlation between urgency and sentiment analysis; meaning urgent comments correlate with negative sentiments. The processing was as follows:

Urgency. To find the learners for whom most of their comments need intervention, we calculate the number of urgent comments for each learners. After that, we clustered in an unsupervised manner all the learners into three groups, by assigning each learner based on the number of urgent comments, to a specific cluster.
Sentiment Analysis. We analysed every comment to extract sentiment polarity into three categories (positive, negative, and neutral) sentiment using the VADER tool. We selected this tool because it is a well-known tool and some researches proved that VADER outperforms Text Blob in social media [28] [29]. Then we found the overall average value of sentiments for each learner and created sentiment clusters, low sentiment number indicating high-risk learners.
Steps Access. For each learner, we calculated the number of step accesses. Then we clustered learners into three groups, based on these values. A high step access number is an important indicator of learning activity, possibly connected to high motivation.

For every variable (urgency, sentiment analysis and step access), we clustered all learners into three groups, by applying natural breaks optimisation with the Fisher Jenks algorithm [30] as it works on one dimensional data. Therefore, every learner has three scores that represent the three clusters' variables (urgency, sentiment analysis and steps access). We calculated an overall score for every learner as in Eq. 3.

$$Overall_{score} = urgency_{cluster\text{-}score} + sentimentAnalysis_{cluster\text{-}score} + stepAccess_{cluster\text{-}score} \quad (3)$$

Thus, the overall score will be between (0-6). Then, we mapped the overall score onto different levels of risks: Higher than 3→ High risk; Higher than 1→ Mid risk; Others → Low risk. Then, we segmented learners as below:

- *High risk*: learners who have high overall score from three variables (urgency, sentiment analysis and access steps).
- *Mid risk*: learners who have middle overall score from three variables.
- *Low risk*: learners for whom overall score from three variables is low.

Based on these levels of risks we computed the priority to the intervention for all potentially urgent comments – see Algorithm 1.

Algorithm 1. Priority of Intervention (C, U, S, M)

Input:
i) C: Stream of potentially urgent comment instances.
ii) U: Number of urgent comments for each learner.
iii) S: Average value of comments' sentiment for each learner.
iv) M: Number of steps access for each learner.
Output:
i) Urgent comments with the priority intervention results.
Method:

```
  Build 3 learner clusters for Urgency.
  Build 3 learner clusters for Sentiment Analysis.
  Build 3 learner clusters for Steps Access.
  Compute the Overall Score.
  if Overall Score is higher than 3 then High risk learner.
     Urgent comment = high priority intervention.
  else if Overall Score is higher than 1 then Mid risk learner.
     Urgent comment = mid priority intervention.
 else
     Low risk learner. Urgent comment = low priority intervention.
 end if
End Algorithm
```

4 Results and Discussions

RQ1.1: Is there a relationship between the number of comments written by the learners that need urgent intervention and the average number of comments?

To inspect learners' writing behaviour, we transformed an average number of comments into an urgency bar chart (1 urgent comment, 2 urgent comments, etc), as shown in Fig. 6. Interestingly, we observed that, usually (but not always), if a learner writes more comments that need intervention, they tend to write more comments in total. This is useful in that they do not 'give up' and present longer time to be 'dealt with'.

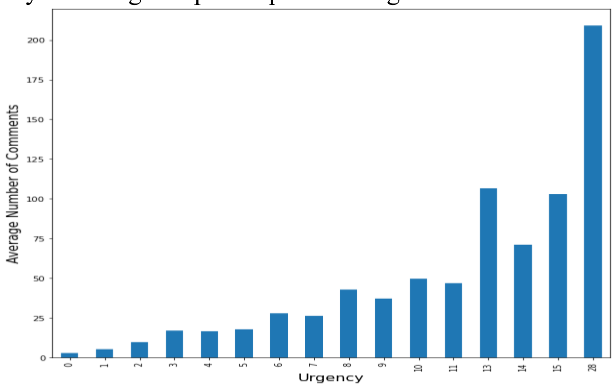

Fig. 6. Relation between urgent comments (urgency) and average number of comments.

RQ1.2: Is there a relation between high-frequency (HF) commenters and their number of steps accessed?

As (Fig. 7, left) shows, we calculated the average number of steps accessed for the HF learners or Urgent and Non-Urgent group. We found that, in general, both groups access learning materials, but the average number of steps access in the Urgent group was lower (33 steps). This difference is statistically significant (Mann-Whitney U test: $p < 0.05$). Consequently, the key observation indicates more learning activity and thus potentially increased motivation for learners with comments not needing intervention.

RQ1.3: Is there a relation between HF commenter number and completion-rates?

The result of the relation between urgency and completion is shown in (Fig. 7, right). As we can see, HF learners who require urgent intervention are less likely to complete the course only (13 %). This difference is statistically significant (Mann-Whitney U test: $p < 0.05$). From this result, we conclude that learners who need intervention tend not to complete the course. We think this is one of the reasons for the high dropout rate. This confirms the need for intervention for urgent comments.

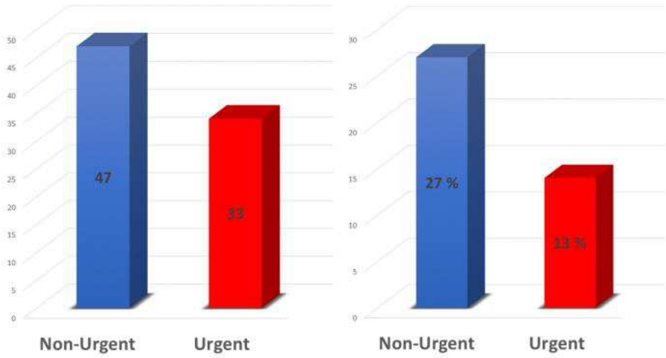

Fig. 7. For each group: average number of steps accessed (left), completion rate (right).

RQ2: Can we design an effective intervention priority framework based on behaviour? As per section 3.3, we proposed a framework containing two phases. We suppose that the instructor can decide to intervene after 5 weeks (our data). In the prediction phase, we used a stratified 5-fold cross validation to estimate the performance of classification model. To evaluate BERT, we measured accuracy averaged over two classes (Urgent, Non-Urgent), Recall, Precision and F1-score for the (important, minority) Urgent class (Table 2). We prioritise the Recall metric that gave us the rare Urgent cases rather than Precision – preferring to ensure we are capturing all urgent cases.

Table 2. The results of BERT model (Precision, Recall, F1-score for the Urgent class).

Accuracy	Precision	Recall	F1-Score
0.90	0.65	0.72	0.68

In the intervention priority phase, there are 387 commenters who have at least one comment that needs Urgent intervention. Table 3 shows the minimum (min) and maximum (max) for each variable in every cluster. For Urgency labelling, we used the label resulting from our manual annotators with voting mechanism, not the one predicted by a classifier, to increase accuracy.

Table 3. The minimum (min) and maximum (max) for each variable in every cluster.

Cluster	Urgency 'min : max'	Sentiment Analysis 'min : max'	Steps Access 'min : max'
0	'1 : 3'	'27 : 75'	'35 : 52'
1	'4 : 9'	'7 : 24'	'15 : 34'
2	'10 : 28'	'-3 : 6'	'0 : 14'

Finally, to further validate the effectiveness of this proposed model, we computed the relation between different risk groups of learners identified (high, mid, low) and their completion-rates. The distributions are visualised in Fig. 8. From this box plot we note that most of completion-rates of high-risk learners are very low, whilst mid risk learners have average ones and for low risk learners, the completion ration is very high. This is further confirmation that our risk model, based on data from the first half of the course, and refining our potential urgency model, can correctly find learners at risk for not completing their course, and separate them from the other two milder risk groups.

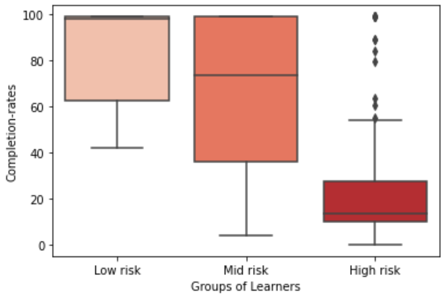

Fig. 8. Boxplot for groups of learners' risk and their completion-rates.

We need here to discuss what constitutes a good classifier of urgent intervention, in terms of best trade-off between false positives (incorrectly identifying learners requiring urgent intervention) and false negatives (failing to identify learners who require urgent intervention). We arguably interpreted it here by giving priority to intervention on urgent cases; hence false negatives were more problematic than false positives. Please also note that learners who need intervention but do not use the comments as communication means are not a target of this research; they would need other means of identification. We also do not compare with work associating comments to participation in MOOCs [31] [32]– as the focus here is on intervention. Further work can link with the work on pedagogical interventions for automated guidance to instructors [19], as well as evaluating how interventions guided by our procedure presented impact on learner progression.

5 Conclusion

In this paper we addressed the *automatic, intelligent intervention problem* in MOOCs. We offer an analysis of learner comments for *urgency*. We demonstrate that learners with high step access rate require less intervention to their comments, whilst step access of *high-frequency commenters* are less than that of other commenters. This might be due to a decrease in learners' motivation to continue accessing the course material, when they have many comments that need intervention. In addition, we confirmed that most course completers did not need much intervention to their comments. Based on these findings, we have constructed a *framework and algorithm for priority of intervention*, to encourage instructors to help their learners and support them by focusing on learners with high risk first, to improve the potential outcomes of the intervention. This framework can be used in intelligent system in MOOC environments. Future work can look into interventions guided by our procedure and its effect on learner progression, as well as using coefficients to allocate different importance to the three criteria (urgency, sentiment analysis and number of accesses) and other optimisation means for the performance of the intervention procedure.

References

1. Kay, J., et al., *MOOCs: So many learners, so much potential.* IEEE Intelligent systems, 2013. **28**(3): p. 70-77.
2. Hone, K.S. and G.R. El Said, *Exploring the factors affecting MOOC retention: A survey study.* Computers & Education, 2016. **98**: p. 157-168.
3. Chandrasekaran, M.K. and M.-Y. Kan, *When to reply? context sensitive models to predict instructor interventions in mooc forums.* arXiv preprint arXiv:1905.10851, 2019.
4. Mazzolini, M. and S. Maddison, *Sage, guide or ghost? The effect of instructor intervention on student participation in online discussion forums.* Computers & Education, 2003. **40**(3): p. 237-253.
5. Chandrasekaran, M.K. and M.-Y. Kan. *Countering Position Bias in Instructor Interventions in MOOC Discussion Forums*. in *Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications*. 2018.
6. Sokolovskaya, A., *Connectivist Knowledge Building, Collaborative Learning, and Social Presence in a Connectivist Massive Open Online Course: A Study of PLENK2010*. 2015, Concordia University.
7. Chandrasekaran, M.K., et al. *Using discourse signals for robust instructor intervention prediction*. in *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
8. Guo, S.X., et al., *Attention-Based Character-Word Hybrid Neural Networks with semantic and structural information for identifying of urgent posts in MOOC discussion forums.* IEEE Access, 2019. **7**: p. 120522-120532.
9. Sun, X., et al. *Identification of urgent posts in MOOC discussion forums using an improved RCNN*. in *2019 IEEE World Conference on Engineering Education (EDUNINE)*. 2019. IEEE.
10. Almatrafi, O., A. Johri, and H. Rangwala, *Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums.* Computers & Education, 2018. **118**: p. 1-9.
11. Lin, F.-R., L.-S. Hsieh, and F.-T. Chuang, *Discovering genres of online discussion threads via text mining.* Computers & Education, 2009. **52**(2): p. 481-495.
12. Adnan, M., et al., *Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models.* IEEE Access, 2021. **9**: p. 7519-7539.
13. Alrajhi, L., K. Alharbi, and A.I. Cristea. *A Multidimensional Deep Learner Model of Urgent Instructor Intervention Need in MOOC Forum Posts*. in *International Conference on Intelligent Tutoring Systems*. 2020. Springer.
14. Clavié, B. and K. Gal, *EduBERT: Pretrained Deep Language Models for Learning Analytics.* arXiv preprint arXiv:1912.00690, 2019.
15. Wei, X., et al., *A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification.* Information, 2017. **8**(3): p. 92.
16. Bakharia, A. *Towards cross-domain mooc forum post classification*. in *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. 2016. ACM.

17. Chaturvedi, S., D. Goldwasser, and H. Daumé III. *Predicting instructor's intervention in MOOC forums*. in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2014.
18. Chandrasekaran, M.K., et al., *Learning instructor intervention from mooc forums: Early results and issues.* arXiv preprint arXiv:1504.07206, 2015.
19. Chandrasekaran, M., et al., *Towards feasible instructor intervention in MOOC discussion forums.* 2015.
20. Chandrasekaran, M.K., et al., *Using discourse signals for robust instructor intervention prediction.* arXiv preprint arXiv:1612.00944, 2016.
21. FutureLearn.
22. Cristea, A.I., et al. *Earliest predictor of dropout in MOOCs: a longitudinal study of FutureLearn courses*. 2018. Association for Information Systems.
23. Alamri, A., et al. *Predicting MOOCs dropout using only two easily obtainable features from the first week's activities*. in *International Conference on Intelligent Tutoring Systems*. 2019. Springer.
24. Agrawal, A. and A. Paepcke. *The Stanford MOOCPosts Data Set*. Available from: https://datastage.stanford.edu/StanfordMoocPosts/.
25. Antoine, J.-Y., J. Villaneau, and A. Lefeuvre. *Weighted Krippendorff's alpha is a more reliable metrics for multi-coders ordinal annotations: experimental studies on emotion, opinion and coreference annotation*. in *EACL 2014*. 2014.
26. Sunar, A.S., et al., *How learners' interactions sustain engagement: a MOOC case study.* IEEE Transactions on Learning Technologies, 2016. **10**(4): p. 475-487.
27. Devlin, J., et al., *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.
28. Bonta, V. and N.K.a.N. Janardhan, *A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis.* Asian Journal of Computer Science and Technology, 2019. **8**(S2): p. 1-6.
29. Min, W.N.S.W. and N.Z. Zulkarnain, *Comparative Evaluation of Lexicons in Performing Sentiment Analysis.* JACTA, 2020. **2**(1): p. 14-20.
30. North, M.A. *A method for implementing a statistically significant number of data classes in the Jenks algorithm*. in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. 2009. IEEE.
31. Saadatdoost, R., et al., *Understanding MOOC learners: Insights from participation in Coursera MOOC.* International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 2019. **14**(1): p. 93-112.
32. Wong, J.-S., et al. *An analysis of MOOC discussion forum interactions from the most active users*. in *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. 2015. Springer.