7-2019

An intelligent platform with automatic assessment and engagement features for active online discussions

Michelle L. F. CHEONG
Singapore Management University, michcheong@smu.edu.sg

Yun-Chen CHEN
Singapore Management University, jeanchen@smu.edu.sg

Bing Tian DAI
Singapore Management University, btdai@smu.edu.sg

Citation

CHEONG, Michelle L. F.; CHEN, Yun-Chen; and DAI, Bing Tian. An intelligent platform with automatic assessment and engagement features for active online discussions. (2019). *Advances and Trends in Artificial Intelligence: 32nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2019, Graz, Austria, July 9-11, Proceedings*. 11606, 730-743. Research Collection School Of Information Systems.

Available at: https://ink.library.smu.edu.sg/sis_research/5106

Published in Advances and Trends in Artificial Intelligence: 32nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2019, Graz, Austria, July 9-11, Proceedings, pp. 730-743.
https://doi.org/10.1007/978-3-030-22999-3_62

An Intelligent Platform with Automatic Assessment and Engagement Features for Active Online Discussions

Michelle L. F. Cheong(✉), Jean Y.-C. Chen, and Bing Tian Dai

School of Information Systems,
Singapore Management University, Singapore, Singapore
{michcheong, jeanchen, btdai}@smu.edu.sg

Abstract. In a university context, discussion forums are mostly available in Learning and Management Systems (LMS) but are often ineffective in encouraging participation due to poorly designed user interface and the lack of motivating factors to participate. Our integrated platform with the Telegram mobile app and a web-based forum, is capable of automatic thoughtfulness assessment of questions and answers posted, using text mining and Natural Language Processing (NLP) methodologies. We trained and applied the Random Forest algorithm to provide instant thoughtfulness score prediction for the new posts contributed by the students, and prompted the students to improve on their posts, thereby invoking deeper thinking resulting in better quality contributions. In addition, the platform is designed with six features to ensure that students remain actively engaged on the platform. We report the performance of our platform based on our implementations for a university course in two runs, and compare with existing systems to show that by using our platform, students' participation and engagement are highly improved, and the quality of posts will increase. Most importantly, our students' performance in the course was shown to be positively correlated with their participation in the system.

Keywords: Discussion forum · Natural Language Processing · Student engagement

1 Online Discussion Forums

Asynchronous online discussion (AOD) forums have been used in many learning contexts to enable students to learn from their peers as a community, and it was mentioned in [1] that "perhaps the most important form of active learning is discussion." Despite the popularity of using threaded forums in Learning Management Systems (LMS), it was argued that they "might not be the best technology to support the interactive and collaborative processes essential to a conversational model of learning" [2]. Some researchers noticed that in threaded forums, students tend to post condensed versions of the explanations of their own ideas rather than responding to the ideas of others [3, 4]. Also, the discussions are often not deep and remain at the surface level such as sharing and comparing information [5]. In [6], they summarized other constraints imposed by threaded discussion forums which include students' tendency to attend to unread posts and most recent posts rather than posts with important content, difficulty in promoting interactive dialogues, provides little support for convergent processes, and the lack of timely feedback. Thus, threaded forum discussions found in LMS do not foster productive online discussions naturally, and developing alternative discussion environments is needed to offer better support in asynchronous online discussions. Several suggestions that future discussion environments should improve upon was provided by [6]. First, it is to foster an online community, provide timely feedback, encourage information sharing and support collaborative problem solving. To achieve these objectives, some incentive mechanism may be designed into the environment. Second, the lack of convergence requires a system which can go beyond knowledge sharing to include active processing and synthesizing of the information provided by the community. Third, the use of multi-functional environments or systems that can integrate new media technology to facilitate learning at different phases, such as asynchronous and synchronous discussions. Fourth, to design appropriate instructional activities or strategies that can improve the quality of the discussion.

In [7], the focus is the Starburst system, which uses visualization techniques to present discussion posts as dynamic hyperbolic tree. Their results found that students were more purposeful in selecting which discussion threads to read and they did it in a more connected fashion. In [8], the authors developed and tested a mobile interactive system to compare social knowledge construction behavior of problem-based asynchronous discussion in e-learning and m-learning environments and found that using additional environments led to more options for students, and that using mobile devices positively influenced students' learning performance. In a study by [9], they compared how students provided online interactive feedback using two different systems, Blackboard being the normal threaded forum and the Annotation System developed by [10]. They found that students showed fewer evaluative feedback but more feedback with suggestions for revisions when using Annotation System than Blackboard. Thus, all three pieces of work [8–10] show that software design and interface can indeed influence learner engagement positively.

In this paper, we will discuss about our intelligent Q&A platform, called CAT-IT, which is an integrated platform with the Telegram mobile app and a web-based forum which are synced in real time. The platform is capable of automatic thoughtfulness assessment of the questions and answers posted, using text mining and Natural Language Processing (NLP) methodologies, and provide instant thoughtfulness score prediction to encourage the students to improve their posts, thereby invoking deeper thinking resulting in better quality contributions. In [11], the author proposed four dimensions in assessing text quality including contextual, intrinsic, representation and accessibility, of which, we focused only on contextual in terms of relevancy, and representation in terms of ease of understanding and interpretability. Thus, we will assess posts using a thoughtfulness score, to measure if a statement contains insightful reasoning and relevance to the issues discussed [12]. We present the effectiveness of our newly developed Q&A platform in engaging students to conduct active online discussion, and how the thoughtfulness score prediction encouraged higher quality posts, leading to improved students' performance in the course.

2 System Design and Architecture

2.1 System Design and Features

CAT-IT is a student-centric platform where the interactions in the form of asking questions and providing answers, are all student-driven. We have chosen this design as we believe that while the traditional tutor-student delivery is important, student-student interaction is also important for academic engagement as all participants are equals and there will be no power relationship issues which may hinder active discussions [13]. We have designed six special features to improve and maintain student engagement and discuss the details of each feature below.

Avatar Identity for Anonymity. This is to overcome one main challenge in Q&A platforms, where poorly formed questions and inaccurate answers provided will be ridiculed [14], thus students tend to be wary about how their posts may be viewed by their peers. Our students participate in the platform using their individual avatar identity to provide them a “safe” environment to participate without the fear of being ridiculed. By understanding the “real deal” when students ask questions and provide answers according to their own understanding and perceptions, which could be repeated questions or inaccurate answers, it will allow course instructors to know exactly what the learning challenges are in a more timely fashion and address them during in-class session immediately. In addition, as our data collection involved human subjects, using the avatar identities will allow us to collect data according to IRB requirements.

Gamification and Leaderboard. Gamification is the process of making something more game-like, but it does not involve creating a complete game. It works with something pre-existing that is not a game, and should engage people with something for a purpose that is not solely to entertain or engage them [15]. Several recent works, both online [16] and classroom [17] have utilized gamification as an attractive strategy to enhance student engagement in learning the course content. Points are usually earned and accumulated when a certain desired learning behavior has been attained by the participant, and a leaderboard is used to instill a sense of competition among the participants by displaying the ranks of the participants’ performance. In our system, we display the real-time ranking of the avatars on a leaderboard based on their cumulative thoughtfulness score earned, to incite students’ desire to stay at the top of the leaderboard.

QA Coins and Bounty. Earning in-game coins is part of gamification. The amount of QA coins earned for a new post is based on a multiplication factor of the thoughtfulness score earned for that particular post. The multiplication factor f(x) follows $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{e}^{-\boldsymbol{x}}$, where x is the cumulative thoughtfulness score of that student. The purposeful design of using such a function is to reward students who have low cumulative thoughtfulness score (x) to earn more QA coins, as a form of encouragement to participate. In many gamified platforms, in-game coins are usually used for making in-game purchases such as buying capes and swords to don their avatar characters. However, in our system, each student only has an avatar identity but does not have an avatar character. Thus, the QA coins earned can only be used to “buy” quick response for the best answer provided within a time limit. This further gamifies the process of participation for both the asker and the answerer.

Auto-routing. One of the key success factors of QA platforms is timely feedback. Posted questions which are not answered will lead to loss in interest and engagement, which may result in the eventual failure of the system [18, 19]. Our system will automatically route an unanswered question to the top five and bottom five cumulative thoughtfulness score students, and also to students who have zero participation, once the time limit is reached. Time limit is default to 24 h if the question does not have a bounty and its associated time limit. By routing to the top five students, we hope that students who have more expertise will provide assistance, while by routing to the bottom five and zero participation students, we hope that the less active students can be nudged to participate.

Need Improvement and Up-Vote Buttons. This is a new feature which was added only in our second run, in response to students' feedback that some students asked questions for the sake of asking to earn thoughtfulness score which will be part of their final grade for the course, as reported in our earlier work [20]. Thus the "Need Improvement" button will allow students to suggest that a question or answer needs further improvement. This is similar to a down-vote in other platforms. However, from the education point of view, we believe that there are no stupid questions or stupid answers, and we do not encourage irresponsible down-voting. Thus, for every click of the "Need Improvement" button, one QA coin will be deducted from the student who clicked the button, so that students can be more mindful when doing so. Conversely, we hope to encourage students to acknowledge good contributions from their peers, and use the "Up-Vote" button as a form of positive recognition. By using these two buttons, we hope to capture some emotional cues in student-student interactions.

Periodic Questions Posted by Automatic Chat Bot. This is also a new feature added in the second run. In our first run, we reported that there were about five answers for every question asked [20]. This showed that students found it easier to provide answers than to ask questions. Thus, we allowed our chat bot to automatically ask questions periodically from a question bank, to drive participation from students in answering questions.

2.2 System Architecture

Figure 1 shows the system architecture of our integrated Telegram and web-based forum. The back-end includes databases and machine learning algorithm, automated by Python scripts, Telegram API and Google API. The system has three main databases, where two of them are Excel files, namely the Q&A corpus for the training of the machine learning algorithm for the prediction of thoughtfulness score, and the question bank where the chat bot will draw its periodic question from. The main database is based on MySQL Workbench with tables to store the user database and the user's associated posts, thoughtfulness scores and QA coins earned. There are two front-ends available, the web-based forum (Fig. 2) and the Telegram mobile app (Fig. 3). Students and instructors must register separately on both applications, using their university email address and Telegram account respectively. The platform will link both registrations to be the same user. Using the Telegram app, students can post questions, provide answers, and view posts contributed by other students and himself/herself. The web-based forum has full functionality that lists all conversations by threads, and in addition to what the Telegram app can provide, students can search for posts and view user information and leaderboard results. The real-time synchronization between the two interfaces provides students the option of participating using their computers or their mobile phones on-the-go, improving engagement as noted in [8].

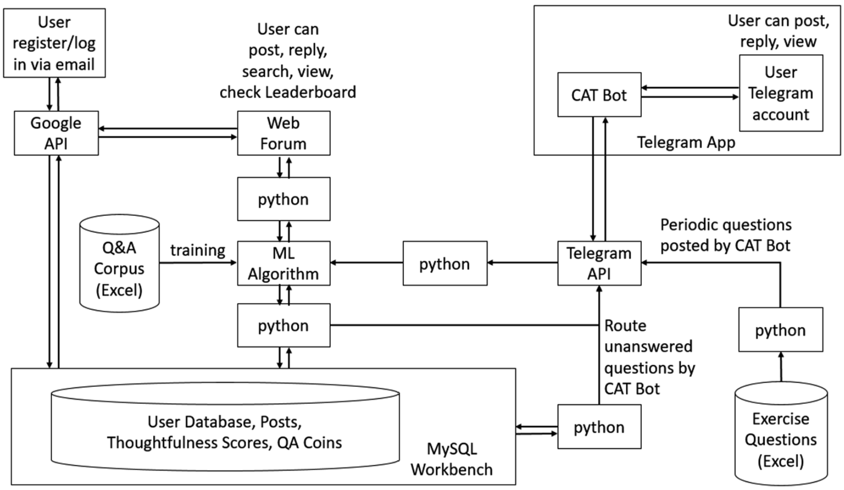

Fig. 1. Schematic diagram of system architecture

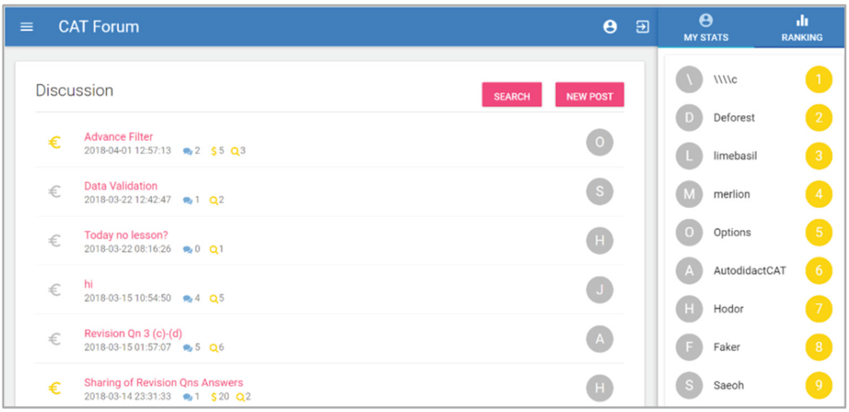

Fig. 2. Web-based forum to view, post and search for posts, and to view user info and ranking

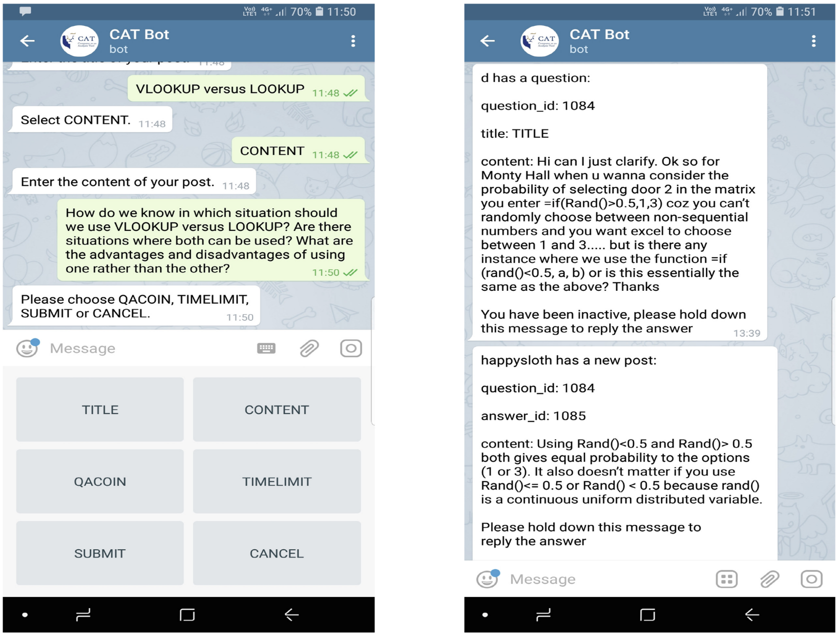

Fig. 3. Telegram app interface for posting question and answering question

3 Data Analytics and Applied Intelligence

3.1 Data Preparation

In order to train a machine learning algorithm to perform the prediction, the team required an initial data set which contains questions and answers representing similar content. An initial data set was crawled from a public forum on Excel (www.excelforum.com) over a four-month period, from June to September 2017. After careful selection based on relevance to the course content, a total of 2377 posts, containing 340 questions, were used to form the training data set. The two course instructors then manually labelled the posts independently on a scale of 0 to 5, based on agreed labeling rules where 0 means that the post is not thoughtful at all; 1 means that the post is a short sentence without details; 2 means that the post contains two to three sentences with some explanations using one or two Excel formulas; 3 means good explanation with example and formula; 4 means good explanation with example, formula and comparison; and 5 means clear explanation, examples, references, suggested formulas and interpretations. We tested inter-coder reliability using consistency index [21] and Cohen's Kappa statistics [22] and achieved 0.906 and 0885 respectively, representing high reliability and deemed that the labeling were consistent and reliable, and thus the labelled data can be used for training the machine learning algorithms.

To implement our system for other courses, the Q&A corpus and chat bot question bank related to the course will need to be prepared. A best performing machine learning model will need to be trained to perform the thoughtfulness prediction. Such work will take about three months at most and the system will be ready for deployment. In the case of a programming course where posts are computer programs, an online judge system will be more suitable. However, such online judge systems are usually used in competitive programming or used by course instructors to assess if the programs submitted by students are correct, and are not meant for discussion purposes. Thus, an integration with open source online judge API for code compilation and execution can be done in future to assess the computer program separately from the text discussion posts.

3.2 Data Pre-processing and Feature Selection

We pre-processed the text using tokenization; stop words removal; word stemming; trim and count the number of URL references and number of Excel formulas; part of speech tagging; and natural language parsing using the Stanford Parser (https://nlp.stanford.edu/software/lex-parser.shtml) to analyze the grammatical structure of the sentences. We generated structural features including average number of characters per word, average number of words per sentence, number of words, average parse tree height, and average number of subordinate clauses per sentence. For syntactic features, we generated average number of noun phrases, verb phrases, and pronounce phrases per sentence. In [23], it was found that discourse relations were correlated with the text quality. Thus, we included "expansion" relation where the second argument expands the first argument or moves its narrative forward; "comparison" relation which highlights the difference between two arguments; and "contingency" relation where an argument causally influences another argument. As discussed in [24], questions can be classified into factoid, list, definition, complex and target. We have adopted the similar concept and proposed an ordinal scale to take care of increasing complexity and discourse features in the different question types according to "where" = 1, "what" = 2, "how" = 3, "when" = 4, and "why' = 5. In addition, we also included whether the post is a question or an answer, and the number of reference links and Excel formulas.

3.3 Machine Learning Models and Selection

With the features, we trained four families of machine learning models to predict the thoughtfulness score. We tested Linear Regression (LR), Neural Network Regression (NN), Support Vector Machine (SVM) and Random Forest (RF). For NN, we used multi-layer perceptron with hyperbolic tangent (tanh) activation function. There are two hidden layers, where the first layer has 1000 neurons and the second layer has 100 neurons. We used Adam, a stochastic gradient descent optimization, with learning rate set at 0.001. For SVM, we have tested several kernels including linear, polynomial, sigmoid and radial basis function (RBF) and found that RBF performed the best. In RF, 10 trees with a maximum of six-feature split used, has the best performance. We used stratified 10-fold cross-validation to split the dataset and the results are given in Table 1. RF outperformed the other three models with the lowest mean squared error (MSE) and the highest R Square value, so we used the RF model to perform automatic thoughtfulness score prediction for the future posts contributed by the students.

Table 1. Machine learning models comparison

	LR	NN	SVM	RF
Mean squared error	0.590	0.276	0.319	0.080
R square	0.188	0.622	0.563	0.892

4 System Process Flow, Implementation and Results

4.1 System Process Flow

The process flow is depicted in Fig. 4. Student starts by asking the first question (Q). A question contains a title, content, optional QA coins as bounty and its associated time limit. Once a question is posted, students can provide answer (A) to the question or to an earlier answer with no depth limit. For every question or answer post, the machine learning (ML) model will predict the thoughtfulness score and prompt the student to improve the post. If the student chooses to improve the post, both posts and their respective thoughtfulness scores will be recorded. In our first pilot run, students were not shown the thoughtfulness scores, so the final post, which could be the first or second post (if student chose to improve the first post) will be posted. While in the second run, the thoughtfulness scores were shown and student can choose either the first or second post to be the final post. For questions with bounty and time limit, the best answer will earn the QA coins posted within the time limit. Once time limit is reached, any unanswered questions will be routed based on our auto-routing rule. For questions with no bounty, the time limit will be automatically set to 24 h.

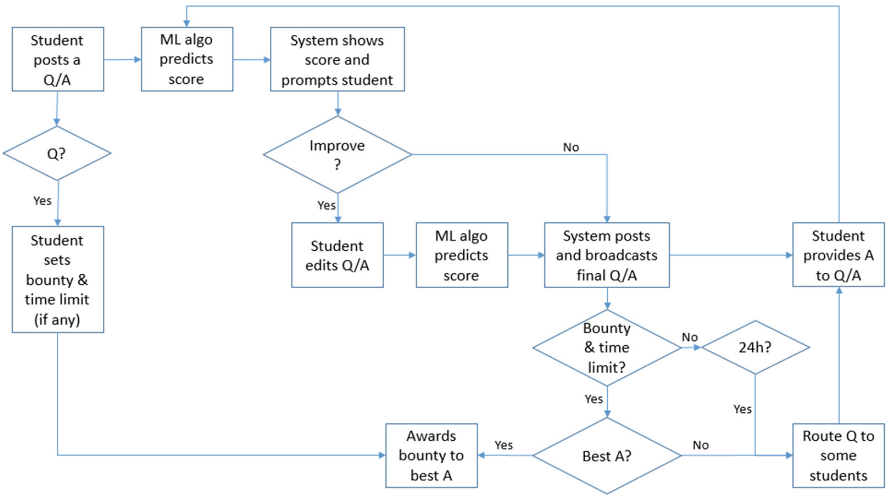

Fig. 4. Process flow

4.2 Implementation

The platform was implemented for a university course on spreadsheets modeling over two runs, with different number of sections, different number of students, different allocation of marks (0%, 5% or 10%) as a percentage of the overall assessment, and with the two additional features (Need-Improvement and Up-Vote buttons, and Automatic Chat bot) added in the second run. For Run 1, participation was mandatory for all sections (5% or 10%). For Run 2, only section G13 was allocation 0% which means voluntary participation (Table 2).

Table 2. Two implementation runs

	# of sections	# students	Settings for different sections
Run 1	3	128	G1 (10%), G15 & G16 (5%)
Run 2	4	147	G1 (10%), G10 & G12 (5%), G13 (0%)

4.3 Comparing Participation Performance with Other Platforms

Table 3 below compares the participation rate and average number of post per active student of our platform with three other similar discussion platforms which were implemented in the past, and their performance results were reported in the literature.

BlikBooks [25] was founded in 2010 and used in more than one-third of UK HE institutions. Its impact on student-tutor and student-student interactions in an International Strategy Development module with 440 students were evaluated [26]. In terms of student-student interaction, only 71 comments were made by 53 students, giving it a low participation rate of 12.1% and 1.34 posts per active student. CaMILE is an anchored forum which was first developed in 1994 and was implemented for students in 17 classes for courses in computer science; chemical engineering; English; history; and literature, culture and communication [27]. Being an anchored forum, discussion topics were provided by the course instructors and thus a higher participation rate would be expected. However, the platform only achieved an average participation rate of 60% and each student only contributed an average of 5.2 posts. On the other hand, SpeakEasy [28] which was also implemented as an anchored forum with discussion

Table 3. Participation performance comparison

Platforms	# of participants	# of active participants	Participation rate	# of meaningful posts	Average # of post per active participant
1. CAT-IT – Run 1	128	101	78.9%	1025	10.15
2. CAT-IT – Run 2	147	123	83.7%	1128*	9.17
3. BlikBooks	440	53	12.1%	71	1.34
4. CaMILE	-	-	60.0%	-	5.2
5. SpeakEasy	180	173	96.1%	-	5.3

* excludes questions posted by chat bot

topics related to Science provided, involved 180 eighth grader, was able to achieve a high participation rate of 96.1% and each student contributed an average of 5.3 posts. Such a high participation rate was achieved due to two main reasons. Firstly, eighth graders would be expected to have more time available for discussions as compared to university students, and secondly, SpeakEasy was implemented as an anchored forum which was part of an assignment.

Comparing to these platforms, our CAT-IT has achieved higher participation rate (at least 78.9%) and higher average number of post per active student in both runs (almost double that of other systems), except when compared to SpeakEasy in terms of participation rate. It is important to note that CAT-IT is not an anchored forum, but a free-form Q&A platform, where all questions and answers are student-driven. In the case of university students who have countless number of time-competing tasks in hand, our platform was able to achieve such high participation rate and high average number of posts shows that it is an effective platform to encourage active student-driven discussion. Both performances will be discussed in the following sections.

4.4 Participation Improvement Due to Specific Features

For both runs, our platform has shown to achieve high participation rate. Of the six special features, we were able to track the improvements in participation due to three of them, specifically QA coins as bounty, auto-routing and automatic chat bot. In applying the QA coins as bounty, the intention is to improve the response time, as unanswered questions will lead to loss in interest and engagement as reported in [18, 19]. Results from Run 1 showed that the average response time for questions with QA coins was 55.6 s, while the average response time for questions without QA coins was 122.5 s. We tested the null hypothesis (H_0) that the average time to response to questions with QA coins is not significantly different than those without QA coins, and obtained a p-value of 0.0283 (<0.05, reject H_0), and t-statistic of −1.996. Thus, we can conclude that by using QA coins as bounty, questions will get faster response.

Our auto-routing feature is also used to reduce unanswered questions. When the time limit is reached, the system will automatically route the unanswered questions to the top five and bottom five cumulative thoughtfulness score students, and also to students who have zero participation. In Run 1, 13 unanswered questions beyond the time limit were auto-routed and 7 were answered, while in Run 2,110 unanswered questions were auto-routed and 20 were answered. This shows that with automatic routing feature, some unanswered questions will receive responses which would otherwise not be forthcoming. Finally, in Run 2, our CAT bot was able to ask questions from a question bank over a 6-week period from Week 2 to Week 7. A total of 140 questions (35 questions each for four sections) was asked by the CAT bot, and 127 of them received answers from the students. The average number of answers received per question for a CAT bot question is 1.96, which is slightly higher than 1.88 for a student question. While the difference is insignificant, it shows that students continue to participate and contribute answers regardless of whether it is a CAT bot question or student question, keeping the discussion active.

4.5 Higher Quality Posts

Once a post is created, our system will prompt the students to improve their posts. Table 4 below shows that in Run 1, where students were not shown the thoughtfulness scores for both attempts, 76.7% of the second attempts were indeed improved. For Run 2, students were shown the thoughtfulness scores for both attempts, and students have the choice to choose which attempt to be the final post, 69.8% of the second attempts were improved. We were surprised to see that a lower percentage of second attempts were improved in Run 2 as compared to Run 1. We offer a couple of plausible explanations why some students still chose the lower thoughtfulness score post as their final post. One is that students did not fully understand how to use the system to choose the higher thoughtfulness score post, and two is that students may have made a mistake in their selections. Nevertheless, our results show that when students chose to improve their posts when prompted, a high percentage of them managed to do so, thus leading to higher quality posts.

Table 4. Improvement in thoughtfulness scores on second attempts

	Number of posts with second attempt	% improved	% did not improve
Run 1	56	76.7%	23.3%
Run 2	53	69.8%	30.2%

4.6 Improved Student Performance in Assessments

We have analyzed and reported in our earlier paper [20] on the positive correlation between students' performance in the course with the thoughtfulness scores they earned, based on the students in Run 1. For Run 2, the same analysis was done and the results are shown in Table 5. It can be seen that for all four sections, the correlations between the average thoughtfulness score and final course performance all displayed positive correlations. The p-values for all are significantly lower than 0.05, except for Section G12 where the p-value is 0.0519, slightly more than 0.05. It is interesting to note that the average thoughtfulness score was higher for sections with allocation of 5% (G10 and G12), as compared to section G1 with allocation of 10%. This is again consistent with what we have reported earlier in our paper [20] which suggested that with a higher stake (10% versus 5%), the quality of posts may not be higher. In totality, we can conclude that students with higher thoughtfulness scores tend to perform better in assessments with appropriate extrinsic motivation, and 5% is better than 10% when higher quality posts are desired.

Table 5. Correlation between average thoughtfulness score and assessment performance

Section	Average thoughtfulness score	Final course performance	
		Correlation	p-value
G1 (10%)	2.03	0.5100	0.0008
G10 (5%)	2.42	0.5029	0.0013
G12 (5%)	2.26	0.3362	0.0519
G13 (0%)	2.16	0.3667	0.0200

5 Conclusions and Future Enhancements

Our intelligent platform with automatic assessment of post contributions and six engagement features has shown to be effective in encouraging active student-driven online discussion, and resulted in better student performance in the course. As compared to other platforms, our CAT-IT can achieve higher participation rate, higher average number of post, and higher quality posts. Most importantly, there is positive correlation between students' performance in the course and their participation in the system. We have successfully applied NLP to automatically assess posts' quality, and show that the thoughtfulness scores of posts increased when students were given the chance to improve their posts, encouraging mindful and purposeful attempts to ask better questions and provide better answers. Our system can be extended to other courses by replacing the Q&A corpus to train the ML model for prediction of thoughtfulness, and the Chat bot question bank. By integrating with open source online judge API for code compilation and execution in future, it will be possible to assess the correctness of computer programs which are contained in the posts, for programming related courses. In our future work, we will improve the NLP analysis to provide guidance to students on how and in what areas to make improvements to their posts, resulting in even higher quality posts.

References

1. Kember, D., McNaught, C.: Enhancing University Teaching: Lessons from Research into Award Winning Teachers. Routledge, Abingdon (2007)
2. Thomas, M.J.W.: Learning with incoherent structures: the space of online discussion forums. J. Comput. Assist. Learn. **18**, 351–366 (2002)
3. Hara, N.M., Bonk, C.J.M., Angeli, C.M.: Content analysis of online discussion in an applied educational psychology course. Instr. Sci. **28**, 115–152 (2000)
4. Larson, B.E., Keiper, T.A.: Classroom discussion and threaded electronic discussion: learning in two arenas. Contemp. Issues Technol. Teach. Educ. **2**(1), 45–62 (2002)
5. Gunawardena, C.N., Lowe, C.A., Anderson, T.: Analysis of a global on-line debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. J. Educ. Comput. Res. **17**(4), 397–431 (1997)

6. Gao, F., Zhang, T., Franklin, T.: Designing asynchronous online discussion environments: recent progress and possible future directions. Br. J. Educ. Technol. **44**(3), 469–483 (2013)
7. Marbouti, F., Wise, A.F.: Starburst: a new graphical interface to support purposeful attention to others' posts in online discussions. Educ. Technol. Res. Dev. **64**, 87–113 (2016)
8. Lan, F.L., Tsai, P.W., Yang, S.H., Hung, C.L.: Comparing the social knowledge construction behavioural patterns of problem-based online asynchronous discussion in e/m-learning environments. Comput. Educ. **59**(4), 1122–1135 (2012)
9. van der Pol, J., van den Berg, B.A.M., Admiraal, W.F., Simons, P.R.J.: Exploration of an e-learning model to foster critical thinking on basic science concepts during work placements. Comput. Educ. **53**, 1–13 (2008)
10. vander Pol, J., Admiraal, W.F., Simons, P.R.J.: The affordance of anchored discussion for the collaborative processing of academic texts. Int. J. Comput. Support. Collab. Learn. **1**(3), 339–357 (2006)
11. Sonntag, D.: Assessing the quality of natural language text data. https://pdfs.semanticscholar.org/8715/1450d6fd7b3f77e83e0b5e66b9fb0b2d40e6.pdf. Accessed 27 Mar 2019
12. Gottipati, S., Jiang, J.: Finding thoughtful comments from social media. In: Proceedings of COLING 2012: Technical Papers, pp. 995–1010 (2012)
13. Kassens-Noor, R.: Twitter as a teaching practice to enhance active and informal learning in higher education: the case of sustainable tweets. Act. Learn. High Educ. **13**, 9–21 (2012)
14. Raphael, R.J.: The 20 dumbest questions on Yahoo Answers. PC World (2009). http://www.pcworld.com/article/184999/strange_questions_yahoo_answers.html. Accessed 27 Mar 2019
15. Andrews, D.: Gamification Systems Development: A practical Guide. The Advanced Services Group, Aston Business School, Birmingham, UK (2017)
16. Barata, G., Gama, S., Jorge, J., Gonçalves, D.: Improving participation and learning with gamification. In: Proceedings of the First International Conference on Gameful Design, Research, and Applications, pp. 10–17. ACM, New York (2013)
17. de Freitas, A.A., de Freitas, M.M.: Classroom Live: a software-assisted gamification tool. Comput. Sci. Educ. **23**(2), 186–206 (2013)
18. Chua, A.Y.K., Banerjee, S.: So fast so good: an analysis of answer quality and answer speed in community question-answering sites. J. Am. Soc. Inf. Sci. Technol. **64**(10), 2058–2068 (2013)
19. Saha, R.K., Saha, A.K., Perry, D.E.: Toward understanding the causes of unanswered questions in software information sites: a case study of stack overflow. In: ESEC/FSE (2013)
20. Cheong, M.L.F., Chen, J.Y.-C., Dai, B.T.: Integrated telegram and web-based forum with automatic assessment of questions and answers for collaborative learning. In: Proceedings of IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2018, Wollongong, Australia (2018)
21. Bennett, E.M., Alpert, R., Goldstein, A.C.: Communications through limited-response questions. Public Opin. Q. **18**(3), 303–308 (1954)
22. Hallgreen, K.A.: Computing inter-rater reliability for observational data: an overview and tutorial. Tutor. Quant. Methods Psychol. **8**(1), 23–34 (2012)
23. Shah, C., Pomerantz, J.: Evaluating and predicting answer quality in community QA. In: SIGIR (2010)
24. Chua, A.Y.K., Banerjee, S.: Answers or no answers: studying question answerability in stack overflow. J. Inf. Sci. **41**(5), 720–731 (2015)
25. Outsell: Blikbooks: Addressing the UK market for higher Ed E-textbooks. https://web.archive.org/web/20131213053912/, http://outsellinc.net/Headline.aspx?ID=455070. Accessed 27 Mar 2019

26. Crowder, M., Antoniadou, M., Stewart, J.: To BlikBook or not to BlikBook: exploring student engagement of an online discussion platform. Innov. Educ. Teach. Int. **56**, 295–306 (2018)
27. Guzdial, M., Turns, J.: Effective discussion through a computer-mediated anchored forum. J. Learn. Sci. **9**(4), 437–469 (2000)
28. Hoadley, C.M., Linn, M.C.: Teaching science through online, peer discussions: SpeakEasy in the knowledge integration environment. Int. J. Sci. Educ. **22**(8), 839–857 (2000)