2020 International Conference for Emerging Technology (INCET)
Belgaum, India. Jun 5-7, 2020

Short Answer Type Discussion Forum Analysis and Assessment

Pankaj Choudhary
School of Computer Engineering and Technology
MIT Academy of Engineering
Pune, India
pankajacks@gmail.com

Shriram R
School of Computer Science and Engineering
VIT Bhopal
Bhopal, India
shriram.r@vitbhopal.ac.in

***Abstract*— A recent research study on the online discussion forum shows significant improvement in the learning growth of students. The discussion forum is a very useful pedagogical tool and plays a very important role in providing interaction among the participant in online courses. The study shows the use of discussion forums in the online course helps the participant in the better learning experience and improved critical thinking through collaborative learning, the course performance, and the cognitive presence is also increased. The researcher suggests that the discussion forum should be the part assessment tool. This paper investigates the use of a question answer type of discussion forum for the assessment of student posts based on the content analysis and the relevance of content to the discussion topic. A similarity measure is calculated for assigning the grade to the student answer and the result is compared with the teacher grade shows the significant results.**

Keywords*— *Discussion forum, Forum Analysis, Content Analysis, answer assessment.

I. INTRODUCTION

Online learning has an exponential growth in recent years due to the advancement in technology. In classroom learning, face-to-face interaction has an advantage [1] and challenges [2]. Online discussion forum plays an important role to overcome the challenges of the traditional classroom such as large class size raises the problem of encouraging the participants and solve the query of individuals. The online interaction with student-to-student or student-to-teacher help allots in solving the student query. The interaction between the participants is either synchronous or asynchronous and there are multiple ways to carry out [3]. Online discussion forum helps the learner to understand the matter of concern more clearly and it is the most efficient way of doing the interaction among participants and that s why it is the most widely used tool in online courses [4]. The participation in the online discussion forum is based on self-interest and the participant of such courses can discussion their queries with other peer learners which enable the collaborative learning environment. An online discussion forum allows the participant to suggest the discussion topic and in turn, they will start to post, comment, vote and view the post of other learners on that topic [5]. Critical thinking [6] and knowledge construction [7] is also the expected outcome of online discussion. All the research reviewed in this paper shows the common interest of online discussion, to enable communication among the participant of the online course.

Use of online discussion forum in online courses are generating a huge amount of data and that enable us to get the insight into the discussion forum. There is a need to do the analysis on the data generated through online discussion and major ad-vantage to do the analysis is the data is available in large amounts and already in the digital format [4]. Analyzing the discussion forum content is done using quantitative or qualitative analysis or both of them can be used. In quantitative analysis, the number of posts created by the participants, commented, viewed and the number of relevant word count in the post is considered and in qualitative analysis, the relevance of post content, quality of content is considered. In most of the online courses, participation in the discussion forum is mandatory for the learner and the study also suggests that it could be used as part of assessment [8,9]. Some of the studies in this field are focused on to check engagement level of the participant in the discussion forum, learning the outcome of discussion forum [10], prediction student performance based on participation in discussion forum [11], learner cognitive presence in an asynchronous online discussion [7,12].

To do the assessment of the discussion forum post is a time-consuming job when the number of participation in discussion forum increases and so the number of the post also get an increase. In this paper, two methods are used for the automatic assessment of discussion forum posts based on content matching. The question-answer type of discussion forum is considered for the analysis where the teacher will post the discussion topic and the student has to post their answer before reading the post of other students. A set of the model answer is collected from the teacher and student post content is matched with the model answer. The best match score is considered as the final grade.

In the next sections, related work is discussed in section 2, the content analysis of student answer is discussed in section 3, Dataset used in section 4, result and analysis in section 5, conclusion and future work in section 6.

II. RELATED WORK

The related work done for the analysis of discussion forum content is discussed in this section.

Seethamraju R. analyze the effectiveness of discussion forum so that it can be used as one the pedagogical approach. A different group of students does the participation and the result show a significant improvement in learning [2]. Aragon P. et. al. reviews the statistical modeling of online discussion which helps in analyzing the growth of discussion forum. Also, a model is proposed which help to identify the new arrival of discussion topic temporally. The formation of the discussion topic is based on a model that uses a

978-1-7281-6221-8/20/$31.00 ©2020 IEEE

mechanism governed by a set of parameters θ that is obtained from the actual discussion topics [3]. Chiu TK et. al. has investigated the three common activities of discussion forum viewing, voting and commenting. A stepwise regression model is used to analyze the data collected from the two online course discussion forum and explored the factor which influences the peer learning and performance. Main influencer in peer learning and performance is by viewing the post and lesser by commenting. The finding of analysis shows the leaner s study purpose, weaker instructor-learner ties, and voluntary forum participation [5]. Wikle JS et. al. evaluate the discussion forum post based on the association be-tween them. The result shows the better result in some instances of discussion forum and in the other instance there is no association found. The quality of response and mastery of students in the discussion topic is also evaluated [6]. Marra R. et. al. has carried out the qualitative analysis of discussion forum post. Their research shows the comparison of two methods, the first Interaction Analysis Model (IAM) and the second is Newman, Webb and Cochrane method. Both of this content analysis protocol is used to analysis the one-week discussion forum content. For the better discussion content initially, the first phase student is not facilitated for a single student-led case study but the second phase student-facilitated case study for analysis. The second phase of discussion content is used for the analysis purpose because it contained a rich set of controversial issues. Issues raised by the student are the discussion topic and the student was asked to post their comment to indicate whether they agreed or disagreed with the issues [13]. Azevedo B. F. T. et. al. proposed a text mining technique for the qualitative analysis of discussion forum posts. Uses graph formalism techniques of text mining to represent the relevant term found in the discussion post. Analyzed the forum post based on the relevancy with the current discussion topic. Thematic Relevance Quotient (TRQ) is defined which is used for analyzing how relevant a text is within a certain discussion. The higher value of TRQ shows the important concept in the graph and have more associations between concepts [14]. Lyons M. et. al. uses linguistic inquiry and word count to identify the mental dis-tress of online communities by analyzing the content they have posted. The different format of mental distress is associated with the social, emotional and cognitive difficulties [15]. Wong JS et. al. proposed a mechanism that is used for the linking of the forum post to a cognitive level of bloom taxonomy. The analysis is performing over 50000 discussion forum interaction and the result shows an increase in the cognitive level of learners as the course progresses. The two important parts of the text that can have the most impact in the online discussion are pronoun and affective processes and LIWC can analyze 80 plus different parts of the text. The analysis of the forum message enables the instructor to set the course goal as per student expectation and identify the content which needs to expand [16]. Hu Q. et. al. developed a framework of a concept map for the forum (CMF) which is used to analyze the discussion forum. CMF is formed based on the two-point the most frequently used post and the content of the post. Forum post content is based on C programming course and the most frequent post belongs to two category concepts and debugging. CMF can be used to evaluate the answers automatically [17].

III. Proposed Work

Assessment of Question-Answer (QA) type discussion forum becomes tedious when there is a large amount of student participated in the course discussion forum. The assessment of QA type discussion forum post is considered for the analysis where a topic or question is posted by the teacher and all the student has to post their answer before seeing the post of other students. The analysis of this QA type forum is limited to assesses only a short answer type of question and the post should not exceed the length up to three to four lines or it should contain a maximum of 100 words.

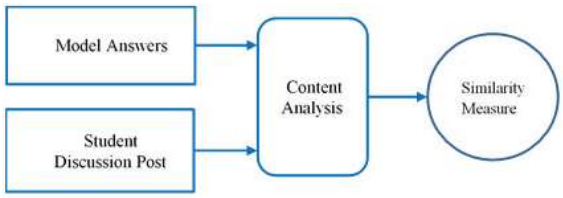

Fig. 1. Analysis Framework

The assessment of student posts is done using content analysis (see Fig. 1). The model answer is compared with the answer of students and based on the matching similarity the answer is graded.

To calculate the similarity measure two existing method is used. First is Cosine Based Similarity and Second is Levenshtein Distance both are used to match the model answer with the student answer in this paper. The first method counts the matching word between the two strings and the second method matches the pattern between the string. The result of both the method is compared and discussed in section 5 results and analysis.

IV. Dataset

The discussion forum is created for the engineering graduate student and the number of participants in the discussion is 45. Every student has to post the answer to the asked question in 70 to 100 words before viewing the post of other students. The sample is shown in Table 1 where the first column is student id, second is the post of student and third is the grade given by the teacher out of 10. The model answer dataset is also required to calculate the similarity of student answer with the given model answer. The teacher can add as many as possible correct answers for better performance and accuracy.

TABLE I. Sample Dataset of Discussion Forum Post

Id	Sample posted answer of the student	Grade
1	An API is a messenger that takes request and; tales the system what we want to do and; then returns the response back to us	6.7
2	API, or "Application Program Interface", is a set of routines and protocols that provide building blocks for computer programmers and web	6.3
3	APIs are interfaces that build connectivity between the client and a server to import services from the server (3rd party) which is provided	6.9

V. Result and Analysis

The model answers for the discussion topic are tokenized and keywords are used for matching with the model answer keyword using the cosine similarity. In the pattern matching

model answer string is compared with the student answer and similarity is calculated. The result shows that pattern matching gives better results as compared to cosine similarity (see Fig. 2 & Fig. 3) As shown in Fig 4 & Fig. 5 the error in cosine similarity is higher as compared to the error of pattern matching. The average error in cosine similarity is -1.88 and in pattern matching, it is -0.81.

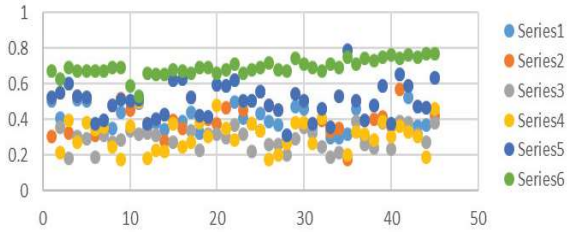

Fig. 2. Sample score v/s actual score using cosine similarity

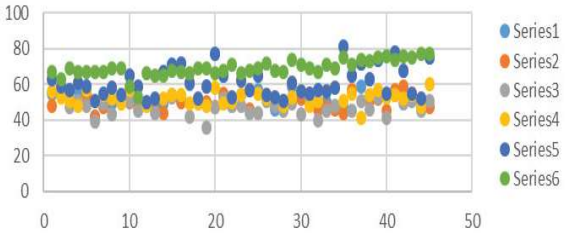

Fig. 3. Sample score v/s actual score using pattern matching

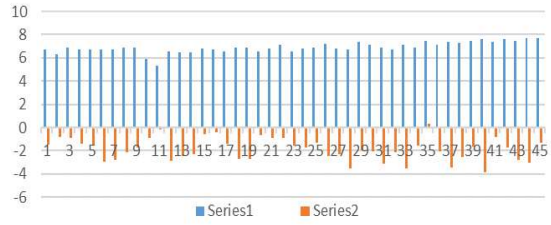

Fig. 4. Error in cosine similarity

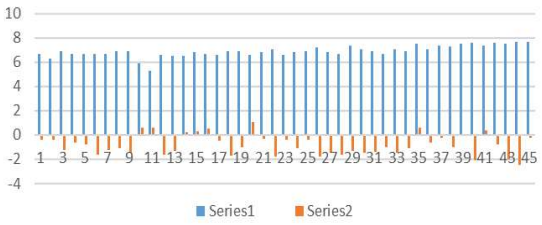

Fig. 5. Error in pattern matching

VI. Conclusion & Future Work

Discussion forum analysis is the most recent research topic in the area of education because of the large amount of digital data available for the analysis. Discussion forum analysis is carried out to check whether it is a good pedagogical tool for learning, better learning outcome, cognitive level of the student in learning and student performance prediction. The content for the analysis of discussion forum taken from the online course and the type of discussion forum considered for the analysis in this study is the question-answer type of forum where the teacher has to post the discussion topic and students have to post their answer. The content analysis method is used to measure the similarity of the model answer with the posted answer of the student to the discussion topic and their relevance. The result shows better accuracy when compared with the grade given by the teacher. In the future, the different advance and intelligent model for the measure of similarity and relevance of the post has to be used. Also, the dataset from different do-mains and groups of the student is to be considered to analyze the discussion forum content and do the comparative analysis between the methods. Finally, the discussion forum server as the better learning tool in an online course or in the traditional classroom to carried out the collaborative learning among the student and it can be considered as one of the summative assessment tools.

References

[1] Miliszewska, I., Is it fully on or partly off ? The case of fully-online provision of transnational education. Journal of Information Technology Education: Research, 6(1), 2007, pp.499-514.

[2] Seethamraju, R., Effectiveness of using online discussion forum for case study analysis. Education Research International, 2014.

[3] Aragón, P., Gómez, V., García, D. and Kaltenbrunner, A., Generative models of online discussion threads: state of the art and research challenges. Journal of Internet Services and Applications, 2017, 8(1), p.15.

[4] Holtz, P., Kronberger, N. and Wagner, W. Analyzing internet forums. Journal of Media Psychology, 2012.

[5] Chiu, T.K. and Hew, T.K., Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. Australasian Journal of Educational Technology, 2018, 34(4).

[6] Zhang, J., Scardamalia, M., Lamon, M., Messina, R. and Reeve, R., Socio-cognitive dynamics of knowledge building in the work of 9-and 10-year-olds. Educational Technology Research and Development, 55(2), 2007, pp.117-145.

[7] Noroozi, O., Weinberger, A., Biemans, H.J., Mulder, M. and Chizari, M., Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. Computers & Education, 2013, 61, pp.59-76.

[8] Dennen*, V.P., From message posting to learning dialogues: Factors affecting learner participation in asynchronous discussion. Distance Education, 2005, 26(1), pp.127-148.

[9] Mazzolini, M. and Maddison, S., When to jump in: The role of the instructor in online discussion forums. Computers & Education, 2007, 49(2), pp.193-213.

[10] Wikle, J.S. and West, R.E., An Analysis of Discussion Forum Participation and Student Learning Outcomes. International Journal on E-Learning, 2019, 18(2), pp.205-228.

[11] Lopez, M.I., Luna, J.M., Romero, C. and Ventura, S.. Classification via clustering for predicting final marks based on student participation in forums. International Educational Data Mining Society 2012.

[12] Shields, K.D., Crowdsourcing Cognitive Presence: A Quantitative Content Analysis of a K12 Educator MOOC Discussion Forum 2017.

[13] Marra, R.M., Moore, J.L. and Klimczak, A.K. Content analysis of online discussion forums: A comparative analysis of protocols. Educational Technology Research and Development, 2004, 52(2), p.23.

[14] Azevedo, B.F.T., Behar, P.A., and Reategui, E.B. Qualitative analysis of discussion forums. In IADIS MULTI CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS 2010, pp. 251-258.

[15] Lyons, M., Aksayli, N.D. and Brewer, G., Mental distress and language use: Linguistic analysis of discussion forum posts. Computers in Human Behavior, 2018, 87, pp.207-211.

[16] Wong, J.S., Pursel, B., Divinsky, A. and Jansen, B.J., November. Analyzing MOOC discussion forum messages to identify cognitive learning information exchanges. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community 2015, (p. 23). American Society for Information Science.

[17] Hu, Q., Huang, Y. and Deng, L., August. A Method for Analysis of Online Discussion Forum in MoodLe. In 2018 13th International Conference on Computer Science & Education (ICCSE), IEEE, 2018, pp.1-4.