10-2020

Automated discussion analysis - Framework for knowledge analysis from class discussions

Swapna GOTTIPATI
Singapore Management University, SWAPNAG@smu.edu.sg

SHANKARARAMAN, Venky
Singapore Management University, venks@smu.edu.sg

GOKARN MALLIKA NITIN
Singapore Management University, mallikang.2015@sis.smu.edu.sg

Automated Discussion Analysis - Framework for Knowledge Analysis from Class Discussions

Swapna GOTTIPATI, Venky SHANKARARAMAN, Mallika NITIN Gokarn
School of Information Systems, Singapore Management University
swapnag,venks,mallikag.2015@smu.edu.sg

***Abstract*—This research full paper, describes knowledge management of class discussions using an analytics based framework. Discussions, either live classroom or through online forums, when used as a teaching method can help stimulate critical thinking. It allows the teacher to explore in-depth the key concepts covered in the course, motivates students to articulate their ideas clearly and challenge the students to think more deeply. Analysing the discussions helps instructors gain better insights on the personal and collaborative learning behaviour of students. However, knowledge from in-class discussions and online forums is not effectively captured and mined due to lack of appropriate automated tools. In this paper, the authors propose an automated discussion analysis (ADA) framework that provides a starting point providing guidance on the development of automated tools for performing different analysis on live classroom or online discussions. AI technology plays a key role in developing tools for knowledge representation and analysis. We propose software systems based on ADA framework and AI technology. The paper then describes a case study where one model of the ADA framework, individual behaviour analysis model, has been applied to automate the analysis of the online discussion forum used in a postgraduate course.**

Keywords—discussions, online discussion forum, live classroom discussion, discussion analysis, AI technology, framework

I. INTRODUCTION

Learning is an interactive process between the student, the teacher, and the subject matter. Learning is enabled by various teaching approaches that motivate students' desire to learn and empowers students to think about the subject matter on their own [1]. Classroom discussion is a sustained exchange between teachers and their students with the purpose of developing students' capabilities or skills on a specific topic. It provides a cooperative learning method that promotes students to interact with other students and the instructor. According to the model illustrated by the "learning pyramid" developed at the National Training Laboratory (NTL), when used as a teaching approach, classroom discussions is an active study method that can lead to greater retention of information and material studied, and higher academic achievement. Classroom discussions provide opportunity for effective personalized and collaborative learning across various education programs. In-class discussions, as well as online discussion forums should be carefully designed and executed by the instructors to stimulate student thinking, and increase participation and engagement [2, 3].

Analysing the discussions helps instructors gain better insights on the personal and collaborative learning behaviour of students. Thus providing directions for making appropriate changes to the content and delivery so as to enhance student learning behaviour. For example, by analysing the individual student participation in the discussion forum enables the instructor to provide participation grade and necessary feedback for effective personalised learning; by analysing the topics that the individual student focusses upon will enable the instructor to discover the strengths and weakness of each student with regard to specific topics in the course [1]. Additionally, discussions provide a rich source of content knowledge that can help students enhance their problem solving and collaborative learning skills. For example, the posts on specific topics can be summarised and shared with all the students in the cohort, thus supporting topical revisions or improving the problem solving skills for projects or assignments. Based on literature review, we understand that most of the online discussion forum data is not effectively used due to the fact that the manual process of generating insights and high quality information is a very tedious task [4, 5]. Additionally, knowledge from in-class discussions is not effectively captured and mined due to lack of appropriate automated tools.

Artificial Intelligence (AI) technology plays a key role for developing representations and reasoning about cognitive insights, for representing knowledge and enabling reasoning, and for measuring collaborative activity [5]. AI provides a technology platform to help deep understanding of the learner's background, strengths, and weaknesses. Traditional AI systems are restricted to tutoring and instructing. Yet there is great potential for AI to create a meaningful impact in education domain. AI methods are drawn from variety of disciplines, including data mining, machine learning, and analysis of psychometric tests using statistics, text analytics, NLP, information visualization, and computational modelling. In our research, we explore various AI methods to integrate with the task of automated analysis of discussion posts.

In this paper, we study the research question- "How classroom discussions can be converted into knowledge using automated tools?" The methodology adopted combines previous research work in the area of discussion analysis and latest developments in NLP and machine learning technologies to develop an Automated Discussion Analysis (ADA) framework. This framework provides a starting point providing guidance on the development of automated tools for performing different analysis on live classroom or online discussions. It elaborates on the main components of a classroom discussion analysis model and the interactions between these components. It also illustrates the main aspects to be taken into account when implementing discussion analysis tools based on artificial intelligence techniques such as data mining, machine learning, text analytics, and natural language processing techniques. The analysis of this framework conducted by applying on the discussion forum data and evaluate the findings.

Software applications accelerate the analysis of the outcomes from the framework. Therefore, based on ADA framework, we propose software systems with example features useful in achieving educational goals. Furthermore, in this paper, we present a case study where components of the framework are applied in the discussion forum within a course in our school

This paper is structured as follows. Section II provides the literature survey aligned to our work. Section III describes the ADA framework. In section IV, we present an overview of educations systems that can be developed based on the ADA framework. In Section V and VI through two case studies, we describe the application of ADA framework for individual behaviour analysis and content analysis respectively. The main conclusions from our research and future work are presented in Section VII.

II. Literature Review

A. *Learning through Discussions*

Discussions used as a teaching method can help stimulate critical thinking. It allows the teacher to explore in-depth the key concepts covered in the course, motivates students to articulate their ideas clearly and challenge the students to think more deeply. Discussions help to promote learner-centred education, which "positions students as co-constructors of knowledge by situating them as active, disciplined participants in their education rather than passive receivers of pre-constructed 'truths" [6]. Discussions can be implemented in a course through live classroom and, or online discussion forum. Both these methods provide opportunities for students to actively engage in their learning process through active participation [7,8]. Using empirical evidence, Rudsberg et al, have been able to identify two specific learning processes that are enhanced; learning to specify the conditions for one's claim; and learning to find new solutions through arguments during classroom discussions [8]. Discussions are a primary mechanism that promote social interaction in the classroom [9]. These social interactions can help structure the knowledge acquisition process. As pointed out by Askell-Williams et al., "it encourages the students to put into the public domain their reasoning and understandings, which can then be augmented, examined, elaborated, critiqued and related to the understandings of other students" [10].

B. *Analysing Discussions*

As with any learning technique, it is very important that the instructor is able to analyse its effectiveness and also be able to assess the student performance. Therefore, one has to analyse the discussions in order to observe interaction patterns, student participation levels, content discussed, etc. Instructor can gather data pertinent to different aspects of discussions, that are relevant to all students and data specific to each student. For live classroom discussion, this data can be collected from video or audio recordings, and for online discussions the data is collected through the platform used for conducting the discussion forum. Using this data, the instructor can perform additional analysis to gain interesting insights. For example, these include: statements made by the students; statements made by the instructor; drawings; external references and links; total number of questions; intellectual level of questions; evolution of topics; frequency of participation; length of discussion responses; discussion between students vs discussion with an instructor; asking questions, of instructor, of other students; connection to previous learning; connection to personal experience; mapping intellectual level of the discussion to Bloom's cognitive levels [11].

Over the years, researchers have adopted a number of ways for analysing the discussions. On one dimension, the methods vary is in terms of whether it is classroom discussion or online discussion forum [12]. On another dimension, these methods include manual approaches to analysis and more recently, automated tools through the use of text analytics techniques. Mercer has classified the analysis methods as shown in Figure 1 [13].

Quantitative methods convert the transcribed discussion into counts of a specified set of relevant features using coding schemes. For example, a relative number of 'talk turns' taken by teachers and students. One approach to implement this method is through the use of "systematic observations", where an observer records the discussions and in some cases gestures. By observing the classrooms or video-recordings of the classroom, the observer manually maps what is seen or heard to the pre-defined set of categories and statistical methods are then applied to the collected data. This approach does have some limitations which besides others include; scalability when handling a lot of data across multiple classroom sessions; and, not all discussions can be mapped to pre-defined categories. More recently, researchers have been applying text analytics techniques. For example, these methods can automatically measure frequency of occurrence of certain words and further analyse which words occur together or see if a technical term introduced by the instructor is used by students in their group-based activity. This method allows analysis of discussions across many classroom sessions relatively quickly.

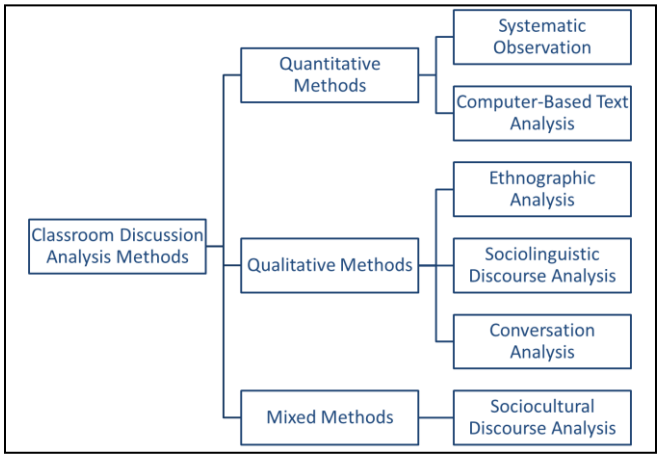

Fig. 1. Classification of classroom discussion analysis methods

Qualitative methods are focused on understanding the nature, patterns, and quality of spoken interactions through use of ethnographic, sociolinguistic and conversation analysis. Ethnographic analysis aims to capture detailed description of observed events, through continuous and close involvement in the social environment that is being studied.

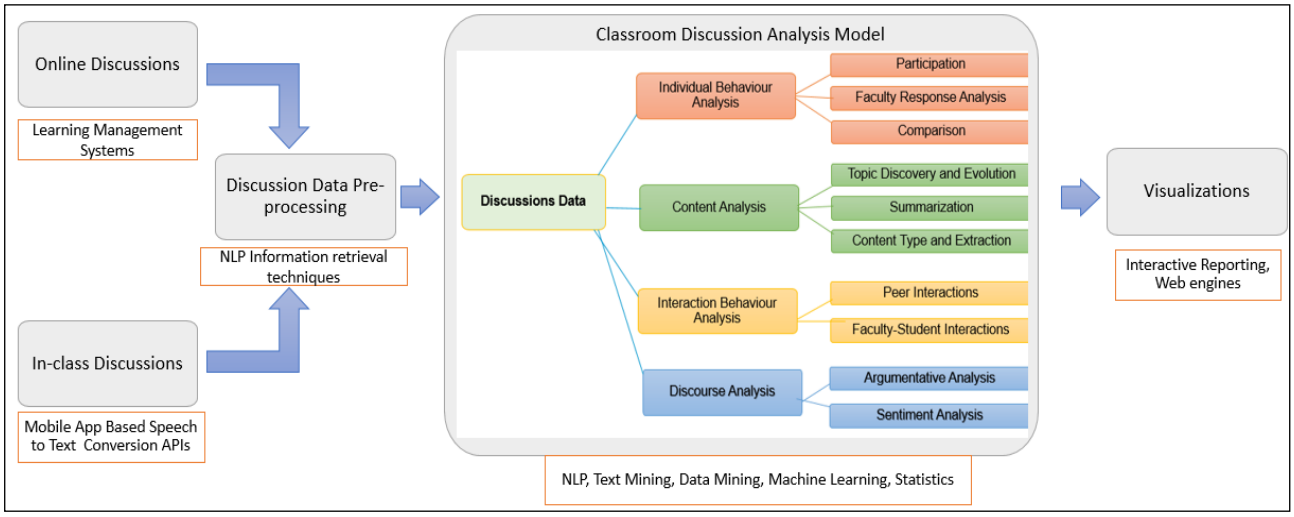

Fig. 2. Automated Discussion Analysis (ADA) framework

For example, (Maybin, 2006) studied how students took up and developed certain ideas and themes, by capturing all the "talk" of the students both in class and outside class during break periods over several months [14]. Sociolinguistic analysis aims to observe the relationship between the forms and structures of language and its uses in the discussion. This can include identification of distinctive sound patterns, vocabulary and grammatical constructions. Using sociolinguistic analysis, Swan studied the extent to which girls and boys dominated classroom interactions [15]. Conversational analysis is based on the tenet that humans use words to perform certain social actions such as describing, questioning, agreeing, offering, etc. Hence analysis of the word utterances can provide insights into the social interactions that happen during classroom discussions [16]. Stokoe studied the discussions of university students in a seminar class to determine the topics types of topics developed in the conversations, the opening topics that drive the discussion, alignment with topics covered during the discussion, and the topics of talk students treat as relevant, appropriate and legitimate to their discussions [16].

Sociocultural discourse analysis is a hybrid method that uses both quantitative and qualitative methods to focus on the content, function, and the ways shared understanding is developed in a discussion [12]. For example such analysis can identify if a specific technical term such as "microservices" which is introduced by the instructor during the lecture is taken up by students later in their group discussion.

Using the above research work as a foundation, we developed the "Automated Discussion Analysis Framework" that uses mobile app, natural language processing (NLP), text mining, data mining, machine learning and statistical techniques so that discussions either online or in classroom can be analysed to provide further insights to enhance learning.

III. Automated Discussion Analysis (ADA) Framework

In spite of the above research work from educational psychologist, from a tools perspective, in a university setting, research questions pertaining to how to develop automated tools to analyse discussions largely remain unclear and unanswered. Therefore, to help better understand how emerging techniques in text mining, and machine learning can be applied to discussion analysis, we present an automated discussion analysis framework (see Figure 2). As shown in Figure 2, the four components of the proposed framework are:

- Discussion Capture Component-Techniques and tools for capturing the discussions
- Data Pre-Processing Component-Techniques and tools for pre-processing the raw discussion data
- Discussion Analysis Models- Models used for analysing the processed data
- Analysis Reports-Visualization of the analysis.

A. *Discussion Capturing Component*

Discussions can be captured through multiple methods. Live audio discussions can be captured using mobile devices such as mobile phone or wearable microphones. Subsequently the audio can be converted into text through use of text conversion APIs such as Cloud Speech API, IBM Watson Speech to Text service and Amazon Transcribe. Online discussions are captured using discussion forums that are usually available in the Learning Management System. The key criteria to be considered when developing this component is the accuracy. If the text captured is inaccurate, subsequent analysis will not produce useful results.

B. *Data Pre-processing*

The objective of data processing is to collect and prepare the raw text data so that it can be used as input for the discussion analysis model. The data preparation stage relies on the learning management systems employed in the institutions. A well-developed system, aids in collecting online discussions effectively with eliminating noise such as "html tags". On the other hand, the in-class live audio discussions pose challenges related to spelling errors, missing words, syntax issues and impure language. Qualitative data needs to be cleaned and represented in structured format for extracting useful information. Some common challenges include; noise words such as "a", "an", "for" etc., which are of little value in helping subsequent analysis, same form of words such as "project", "projects" that represent the same topic and the sentiment words embedded within the textual feedback such as "too fast", "not easy" etc. Data processing stage handles such data challenges and prepares the data for the next stage. Some example methods used for data

processing include tokenization, stopword removals, word normalization, parts of speech tagging etc. [17, 18].

C. Discussion Analysis Models

Combining natural language processing, machine learning and text mining techniques [17, 19], universities can collect and analyse discussion data along the four areas as shown in the Figure 2; individual behaviour analysis, content analysis, interaction behaviour analysis and discourse analysis. In this sub-section, we describe each analysis model and suggest example inputs, outputs and techniques that can be used for each model.

1) Individual Behaviour Analysis

Individual behaviour analysis mainly focusses on the quantitative analysis of discussion data such as individual student participation statistics and comparison statistics [20]. The goal is to generate descriptive statistics on the numbers of forum postings, posting rates, class participation rates and relative comparison with other session participation rates. Such numerical scores are useful for auto-grading the students.

In terms of qualitative analysis, to know if the participation or the post was thoughtful or useful, the instructors' response to the specific post or comment is combined with the student's post or comment using a scoring equation. For example, if the student made Comment A: "using the following example I would like to explain the problems with LDA algorithm" following which the instructor made the Comment B: "excellent contribution to the current discussion topic", then Comment A is give a high score, since the comment is deemed relevant to the discussion. In order to do this, the instructor's response is analysed using text mining techniques to discover the instructor's sentiment; positive, neutral or negative. Table I depicts the details of individual behaviour analysis briefly.

TABLE I. INPUTS, OUTPUTS AND TECHNIQUES FOR INDIVIDUAL BEHAVIOUR ANALYSIS

Inputs	Discussion posts and sentiment lexicons
Outputs	1. Summarised reports of participations rates, airtime rates, and relative comparisons. 2. Positive and negative sentiment summation along with the actual responses from the instructor for each participant summarised by time and topic
Techniques	Descriptive statistics measures, sentiment mining models

2) Content Analysis

Qualitative analysis of discussions can be conducted under three dimensions; Content analysis, Interaction analysis and Discourse analysis. Content analysis in this context refer to the study of discussions in order to discover knowledge from student posts or comments [21, 22, 23, 24]. Forums and classroom discussions are knowledge repositories, which contain useful knowledge relevant to the subject matter covered in the course. However, manually assimilating such large volumes of knowledge is very tedious. Therefore, content analysis tools enable the students to grasp the knowledge through the use of indexed navigation and summarization. Navigation through the knowledge is possible by defining the topics of the posts, types of the posts and topical evolution of the posts. The summarization of the knowledge is achieved via the concise topical summaries of the posts. Table II depicts the details of content analysis for discussions.

TABLE II. INPUTS, OUTPUTS AND TECHNIQUES FOR CONTENT ANALYSIS

Inputs	Discussion posts
Outputs	1. Discovery of topic and sub-topics from the posts or comments. Sometimes, the topics discovered can go beyond what is covered within the class to include student's own learning and experience. Topic evolution and connections can show how the topics can be interrelated. 2. Concise summaries of the related posts and comments which are clustered under the same topic.
Techniques	Clustering models, topics models such as LDA and text summarization techniques.

3) Interaction Behaviour Analysis

Discussion forums and classroom discussion provide a platform for interactive learning. Applications of interaction analysis include improvement of teaching style and learning achievement through reflection on classification of interaction type (Amidon, 1968). The analysis is based on learner's contribution behaviour and this can include overlap of concepts, pauses to assimilate and form ideas, echoes to clarify and support information, and repairs to correct, refute and disprove concepts. As a result, the discussions become more than just an assignment; students learn from each other and become more engaged in the learning process. Cognitive and social presence are the two key interactions that can be observed and analysed [25]. Furthermore, the interactions can also be analysed under peer-peer and instructor-student dimensions. Table III depicts the details of interaction behaviour analysis briefly.

TABLE III. INPUTS, OUTPUTS AND TECHNIQUES FOR INTERACTION BEHAVIOUR ANALYSIS

Inputs	Discussion posts, sentiment lexicons and domain specific lexicons
Outputs	1. Cognitive interaction statistics for higher order thinking based on explanations, suggestions, creating solutions etc. 2. Social interaction statistics such as emotions, open communications and group cohesions.
Techniques	Classification algorithms, sentiment mining techniques and social network models

4) Discourse Analysis

Soller's Collaborative Learning Conversation Skill Taxonomy describes a method of classifying conversational behaviours which can be used to distinguish effective and ineffective contributions to interactions in classroom and online discussions [26].

Discourse analysis is an approach which emphasizes the situated, constructed, and generative nature of language [27]. Combined with discourse analysis, Soller's learning framework provides the basis to profile the students in terms of interpersonal aspects and behavioural aspects. Discourse analysis of the discussions aids to discover how the conversations contribute to the development of meaningful learning for participants. Table IV depicts the details of discourse analysis on discussions.

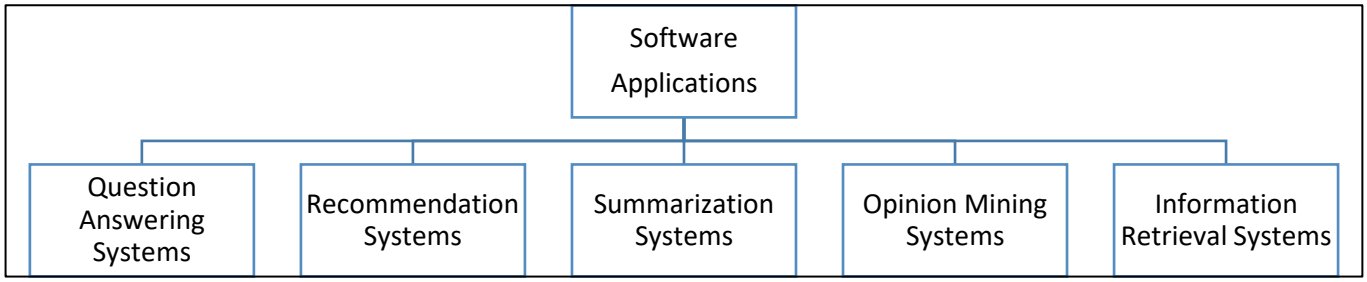

Fig. 3. Education systems based on ADA framework

TABLE IV. INPUTS, OUTPUTS AND TECHNIQUES FOR DISCOURSE ANALYSIS

Inputs	Discussion posts, discourse analysis treebanks, learning frameworks
Outputs	1. Aligning student's profile based on interpersonal and behavioural aspects to Soller's learning framework. 2. Relationships between discourse and cognitive knowledge.
Techniques	Syntax analysis models, semantic analysis models, classification algorithms, and scoring models

D. *Visualizations*

Visualizations are a key component where the goal is to provide user friendly summaries of the qualitative and quantitative analysis results obtained from the data analysis models. Effective visualization helps instructors to analyse and reason about data and evidence. Numerical data may be summarized using dots, lines, or bars, to visually communicate a quantitative message. Text visualization techniques can be challenging and dependent on the outputs from the text mining techniques [20, 28, 29, 30]. These techniques make complex data more accessible, understandable and usable. Simple visualizations like basic charts to complex visualizations such as word clouds, trees, node-graphs and maps are some of the techniques proposed by various researchers to generate effective visuals of the qualitative analysis [25, 31, 32].

IV. EDUCATIONAL SYSTEMS BASED ON ADA FRAMEWORK

According to Woolf's AI grand challenges in education, "Interaction Data to Support Learning' is the third AI challenge that helps to address the goals of [5]. It is about exploring and leveraging the unique types of data available from educational settings and the use of this data to better understand students, groups, and the settings in which they learn. Our framework supports the vision of the challenge – "adaptable to acquiring and analysing educational data and discovery of novel and potentially useful information". The use of software assists and accelerates the qualitative analysis of posts from the discussion forum. Based on our framework, Figure 3 depicts the software applications that can be developed using the data from discussion forums to aid the education stakeholders in the teaching and learning process.

- Question Answering Systems: The goal of these systems is to find the best answer to a certain question. Based on content analysis of discussion forums, questions answering systems can be developed using NLP techniques for understanding the questions and generating the answers.
- Recommendation Systems: The goal of these systems is to provide recommendation in order to aid the user with decision making. Based on individual behaviour analysis, the recommendation of additional course resources can be generated based on a student's weaknesses with regard to the course topics. Recommendations can also be generated for the instructor on the topics where students have low participation and thus alerting the need for further interventions.
- Summarization Systems: The goal of these systems is to generate summaries by reducing the size and level of document detail, but retaining the main ideas and the general meaning of the content. Based on content analysis of the discussion forums, recommendation systems can be developed using machine learning algorithms for generating topic based summaries.
- Opinion Mining Systems: The goal of this system is to identify the sentiments of the participant with regard to certain topics. Based on the students' responses to the topic that is being discussed, it is possible to analyse their sentiment with regard to the topic and their ability to perform argumentative analysis. Discourse analysis and content analysis can be combined to build opinion mining systems.
- Information retrieval systems: The goal of these systems is to retrieve the relevant documents for a given query. Based on content analysis, the query can be processed using the topics and retrieve the relevant posts or generate summaries of the posts. Based on the individual behaviour analysis, search features can be added to discover insights using instructor responses to the student posts.

Due to space constraints, we described only a few example features of the education software systems that are useful for the instructors in the learning process. Each software system can be developed with additional features to address the vision proposed by Woolf [5]. In the following sections, through two case studies, we describe the application of ADA framework for individual behaviour analysis and content analysis respectively, which have developed in our university as a first step towards implementing the third AI challenge namely, "Interaction Data to Support Learning".

V. APPLICATION OF THE FRAMEWORK: CASE STUDY ON INDIVIDUAL BEHAVIOUR ANALYSIS

The study was conducted on a graduate course, "Text Analytics and Applications", offered by School of Information Systems, Our University [anonymised]. The course extended for 14 weeks with a one week term break,

study week and exam week. In this case study we used a discussion forum tool which was part of the Learning Management System. The discussion threads were posted by the instructor for certain selected weeks, especially those weeks that covered complex topics, and to encourage out of class research. Out of the 55 students enrolled in the course, 37 students participated in the discussion forum. More than 50% of the students had industry experience or were currently working in the industry. To motivate the students, every week, the instructor collected the analysis data from the discussion forum and used it to recap the content covered in the classroom session. A total of around 200 student responses were received over all the different discussion threads.

A. *Individual Behaviour Analysis*

For this case study, we describe how the individual behaviour analysis model was used by the instructor, specifically to understand the behaviour of each student and then compare with other student's in the class. Only Participation Analysis and Comparison analysis was done. Instructor Response Analysis was not considered in this case study. According to our framework, the techniques used are statistical models and exploratory visualization charts such as bar graphs and heat maps. Figure 5 shows the visualization generated for this analysis.

Using Figure 5, Part A, the instructor is able analyse how often each student contributes (i.e. the quantity) to the discussion across all sessions and compare it with other students. As seen from Figure 4, Student X has made very little contribution both in terms of the quantity as well as the airtime, whereas Student Y has made a lot more contribution both in terms of the quantity as well as the airtime.

B. *Comparison Analysis*

Based on the average airtime per contribution, a comparison analysis to compare an individual student's contribution to the rest of the class is derived, see Figure 4. We use z-score for this task, which calculates a students' average airtime with respect to the population mean of average airtime of all students normalized by the population standard deviation given by the formula,

Where x is the value of airtime for each student, μ is the population mean and σ is the standard deviation.

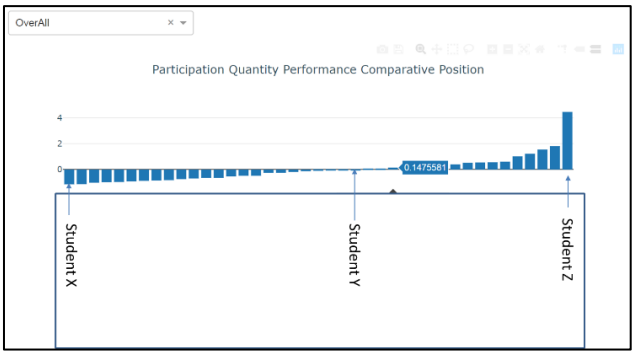

Fig. 4. Comparative analysis based on airtime of the contribution

The statistics from Figure 4 can be used to automatically calculate a score for the class participation for each student. If required, this analysis can also be conducted for each week, mapped to specific topics rather than the entire course.

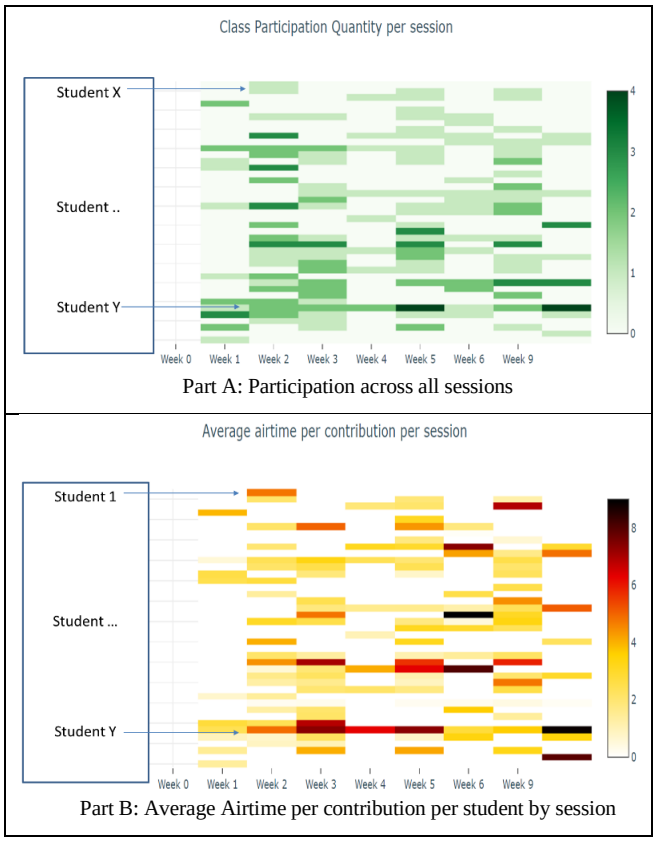

Fig. 5. Individual Behaviour Analysis- Participation analysis statistics

VI. Application Of The Framework: Case Study On Content Analysis

The study was conducted on the same dataset as mentioned in Section V. The key goal of this case study is to discover the topical insights from the discussion forums based on a text analytics approach. In particular, understanding the topics and sub-topics that are emerging in the discussions which can provide useful insights into the student learning process [33]. For example, if the students have discussed only the main topics that were covered in class, it indicates that the students are bounded to in-class learning and have not taken efforts to do further research on their own. If more sub-topics, which were not covered in the class, emerge from the main topic, it indicates the effort towards out of class learning process by the students. In this digital era, it is important for students to learn beyond the classroom, and further scaffold this learning, by instructors intervening to identify the links between the various sub-topics and providing a summary of the topical evolutions [34, 35].

A. *Topic and Sub-topic Analysis*

We applied clustering algorithms on the discussions to extract the sub-topics. Figure 6 shows the network graph for topics and sub-topics visualization. The topics are represented by yellow circles and the sub-topics are represented by red circle. The legend for the sub-topics is shown to the right. For the purpose of this paper, we are showing only 12 sub-topics instead of 20 for simplicity.

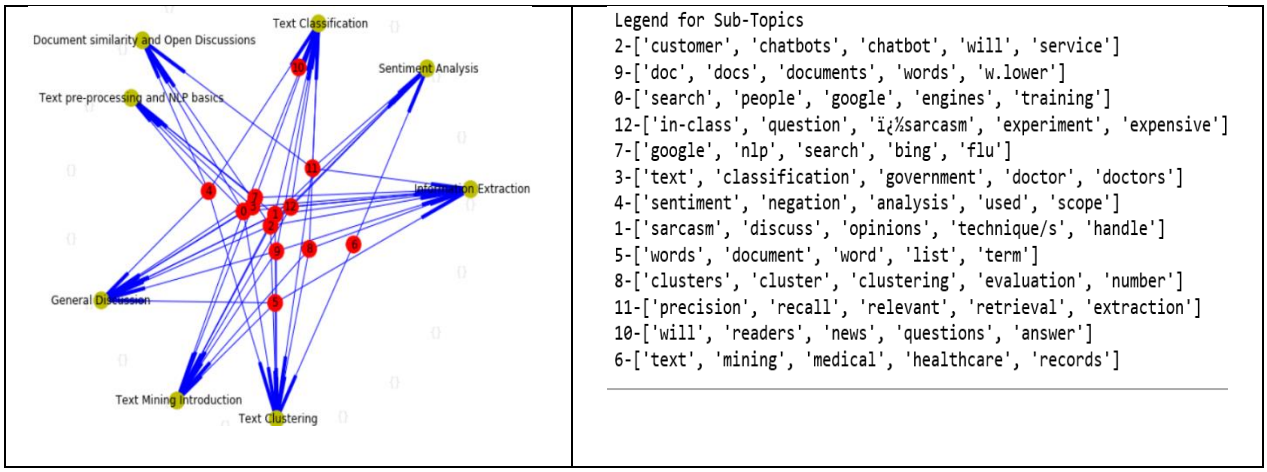

Fig. 6. Topics and Sub-topics network graph. The legend is displayed to show the sub-topics, red circles

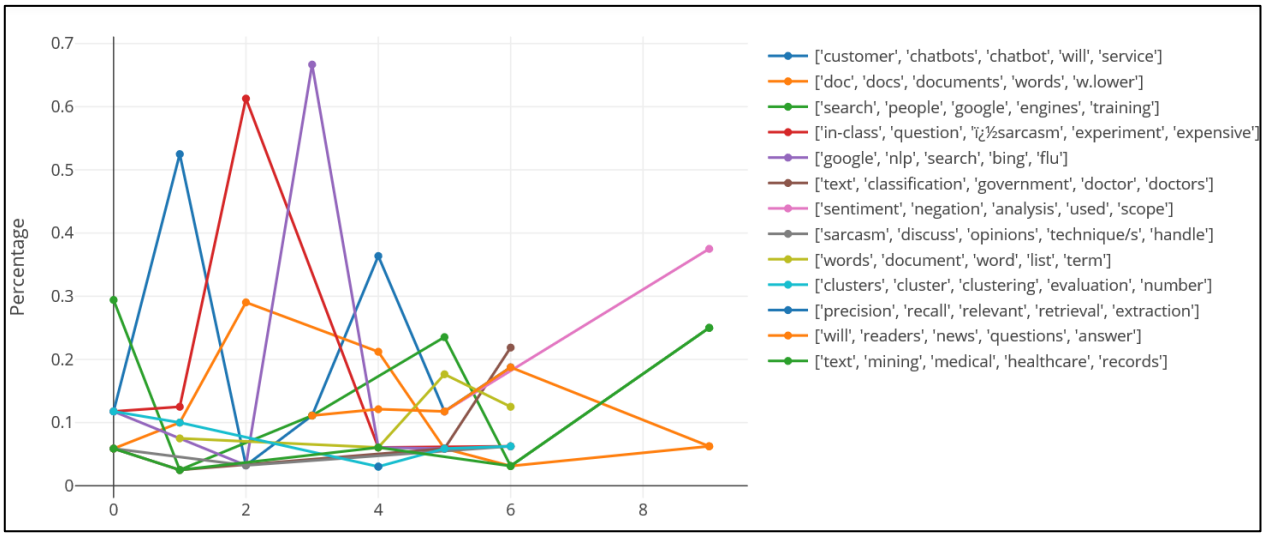

Fig. 7. Sub-Topic Evolution over the weeks. Interactive graph with hovering features

From Figure 6, we observe that sub-topics can be part of more than one main topic. For example topic, 6 - "medical and healthcare", appears under "text classification", "text mining introduction", and "clustering". This shows how students are connecting the sub-topics over various topics via the discussion posts. From such graphs, the instructor can identify the missing sub-topics and submit relevant posts under the main topic to lead the students in the learning process.

B. *Topic Evolution*

Figure 7 shows the sub-topic evolution over time. It depicts topic evolution over the weeks, given the percentage makeup of each week. A threshold value of minimum percentage makeup can be set using the slider below the graph. The interactive nature of the graph also enables the users to study each topic in detail and aid the instructor to decide on the need for intervention if the student misses the sub-topics. This chart is based on the results from the 12 clusters shown in Figure 6.

We observe some interesting patterns in this graph in terms of the short-lived vs repeated topics. Topics on "chatbot" and "healthcare" have occurred several times over the weeks. This is due to the examples that the students choose to apply the concepts within the given domains. The chart is also interactive so that the selection of topics can be done and allows further study of the sub-topics comparison in details.

VII. Conclusion

Discussions either in the classroom or through online discussion forums provide a number of benefits that make them a valued approach to enhance learning. Firstly, it allows students to explore different perspectives and consider different viewpoints on a topic through sharing of opinions, thoughts, and questions about course content. Secondly, through discussions students take ownership for their learning thus shifting the emphasis from teaching to learning. Thirdly, it provides opportunity to explore a topic more deeply than listening to a lecture on the topic. Finally, it provides a powerful mechanism to get students actively involved in the learning process. This in turn help a student learn better.

As any other pedagogy approach it is important for an instructor to be able to analyse the effectiveness of the discussion method by observing the student learning behaviour. Thus providing directions for making appropriate changes to the content and delivery so as to enhance student learning behaviour. In this regard, traditionally manual approaches have been adopted by educational psychologist. These approaches are limited to observing and analysing discussions in one classroom for one or more sessions but scaling them to more classrooms and sessions is constrained

due to enormous resource requirements. With recent developments in text mining, data mining and machine learning it is now possible to develop tools to automate this process.

In this paper, we have presented an automated discussion analysis (ADA) framework that describes various discussion analysis models along with the inputs, outputs and techniques for each. We show how this framework, has been applied to the discussion forum from a graduate course. The ADA framework can be use as basis to help develop different types of applications in the domain of discussion forums, such as question-answering systems, recommendation systems, summarization systems, opining mining systems and information retrieval systems. The main limitation of our work is that in this paper we only demonstrate the implementation of the ADA framework for "Individual Behaviour Analysis" and "Content Analysis". In order to fully validate the framework, our future work will be aimed at developing applications that address other dimensions of the ADA framework such as "Interaction Behaviour Analysis" and "Discourse Analysis", by leveraging the discussion data captured in a live classroom or from online discussion forums.

REFERENCES

[1] Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.

[2] Howard, J. R. & Henney, A. (1998). Student participation and instructor gender in the mixed-age college classroom. The Journal of Higher Education, 69(4), 384-405

[3] Burge, E. (1994). Learning in computer conferenced contexts: The learners' perspective. Journal of Distance Education, 9(1), 19-43.

[4] Chaturvedi, Snigdha and Goldwasser, Dan and Daumé III, Hal. 2014. Predicting Instructor's Intervention in MOOC forums Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014

[5] Woolf, Beverly Park, H. Chad Lane, Vinay K. Chaudhri, and Janet L. Kolodner. 2013. "AI Grand Challenges for Education." AI Magazine 34:66-84.

[6] Boyd, P. (2008). Analyzing students' perceptions of their learning in online and hybrid first-year composition courses. Computers and Composition, 25, 224–243.

[7] Thomas, M. J. W. (2002). Learning within incoherent structures: The space of online discussion forums. Journal of Computer Assisted Learning, 18, 351–366.

[8] Rudsberg, K., Öhman, J & Östman, L. (2013). Analyzing Students' Learning in Classroom Discussions about Socioscientific Issues. Science Education, July 2013, Vol.97 (4), pp.594-62

[9] De Bruyn, Lisa Lobry(2004) 'Monitoring online communication: can the development of convergence and social presence indicate an interactive learning environment?', Distance Education, 25: 1, 67 — 81

[10] Askell-Williams, H & Lawson, M. J. (2005). Students' knowledge about the value of discussions for teaching and learning. J. Social Psychology of Education: An International Journal; Dordrecht Vol. 8, Issue. 1, (Mar 2005): 83-115.

[11] Good, T. I., and J. Brophy. 2000. Looking in classrooms. 8th ed. New York: Longman.

[12] Mercer, N., & Hodgkinson, S. (Eds.). (2008). Exploring classroom talk. Sage. London

[13] Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. British Journal of Educational Psychology, 80, 1-14.

[14] Maybin, J. (2006). Children's voices: Taik, knowledge and identity. Macmillan.

[15] Swann, J. (2007). Designing 'educationally effective' discussion. Language and Education, 21, 342-359.

[16] Stokoe, E. (2000). Constructing topicality in university students' small-group discussion; a conversation analytic approach. Language and Education, 14, 184-203.

[17] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. Introduction to Information Retrieval. Cambridge University Press, New York, NY, USA.

[18] M.F. Porter. An algorithm for suffix stripping, Program, 14 (3):130-137, 1980

[19] Charu C. Aggarwal and Cheng Xiang Zhai. 2012. Mining Text Data. Springer Publishing Company, Incorporated. 2012

[20] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy. (2015). Visualizing patterns of student engagement and performance in MOOCs. Pages 83–92, New York, New York, USA, 2014. ACM

[21] Marra, R., Moore, J., & Klimczak, A. (2004). Content Analysis of Online Discussion Forums: A Comparative Analysis of Protocols. Educational Technology Research and Development, 52(2), 23-40.

[22] Atapattu, T., Falkner, K. and Tarmazdi, H. (2016). Topic-wise classification of MOOC discussions: A visual analytics approach. Proceedings of the 9th International Conference on Educational Data Mining (EDM), Raleigh, NC, USA.

[23] Duval, E. (2011, February). Attention please!: learning analytics for visualization and recommendation. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 9-17). ACM.

[24] Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modelling for understanding MOOC discussion forums: a learning analytics approach. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (pp. 146-150).

[25] Garrison, D.R. (2007). Online community of inquiry review: social, cognitive, and teaching presence issues. Journal of Asynchronous Learning Networks, 11(1): 61—72.

[26] Soller, A.L. (2001). Supporting social interaction in an intelligent collaborative learning system. International Journal of Artificial Intelligence in Education, 12(1), 40-62.

[27] Taylor, S. 2001. "Locating and Conducting Discourse Analytic Research," in Discourse as Data: A Guide for Analysis, M. Wetherell, S. Taylor, S. Yates (Eds.), London: Sage, pp. 5-48.

[28] Alharbi, M.; Laramee, R.S. 2019. SoS TextVis: An Extended Survey of Surveys on Text Visualization. Computers 2019, 8, 17.

[29] Leony, D., Pardo, A., de la Fuente Valentín, L., de Castro, D. S., & Kloos, C. D. (2012, April). GLASS: a learning analytics visualization tool. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 162-163). ACM.

[30] Leony, D., Pardo, A., de la Fuente Valentín, L., de Castro, D. S., & Kloos, C. D. (2012, April). GLASS: a learning analytics visualization tool. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 162-163). ACM.

[31] Andy, C., 2013, Reports Forum Graph, Moodle. https://moodle.org/plugins/report_forumgraph

[32] Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008

[33] Simon, Y. K. Li, and Gary K. W. Wong. 2016. Visualizing the asynchronous discussion forum data with topic detection. In SIGGRAPH ASIA 2016 Symposium on Education: Talks (SA '16). ACM, New York, NY, USA, Article 17, 3 pages. DOI: https://doi.org/10.1145/2993363.2993367

[34] Ma, T., Qu, D., & Ma, R. (2016). Online topic evolution modelling based on hierarchical Dirichlet Process. IEEE International Conference on Data Science in Cyberspace, pp. 400–405.

[35] Jian, F., Yajiao, W., & Yuanyuan, D. (2018). Microblog topic evolution computing based on LDA algorithm. Open Physics, 16(1), 509-516.