Minute-Paper Dashboard: Identification of Learner's Misconceptions Using Topic Modeling on Formative Reflections

Aditya Vadapally
Department of Computer Science
Kennesaw State University
Marietta, GA, USA
avadapal@students.kennesaw.edu

Nasrin Dehbozorgi
Department of Software Engineering
Kennesaw State University
Marietta, GA, USA
dnasrin@kennesaw.edu

Dinesh Chowdary Attota
Department of Computer Science
Kennesaw State University
Marietta, GA, USA
dattota@students.kennesaw.edu

***Abstract*—This work in progress proposes a method for automated analysis of students' short reflections using NLP to get insights into their challenges and learning outcomes in the course. The importance of self-reflection in engineering education has been emphasized more recently as it improves the students' learning experience and helps instructors to remove students' learning gaps. Towards this goal, educational scholars have developed different types of reflection tools as well as analysis methods to get feedback on students' learning outcomes. Reviewing narratives of the reflections is time taking specially in large class settings. One of the known approaches to get feedback from students are mini-reflections called 'minute paper' that asks students to answer two questions briefly about what they learned and what they didn't learn in the class. Although these short surveys reduce the narrative load and help in the quicker review of the reflections still they require instructors to review them one by one. In this work, we apply clustering methods to the students' reflection responses in a software engineering course to extract their challenge areas as well as learning outcomes from each session of the class. The result of the analysis is visualized in a dashboard that dynamically shows an overview of the challenging topics and learning outcomes based on their weight and frequency of their occurrence in students' responses. This means the more students mentioned a certain topic as their challenge or learning outcomes, the topic acquires higher weight as a result. The application of this dashboard helps educators to get quick and real-time insight into students' misconceptions in a formative style. It enables them to narrow students' learning gap by discussing the challenging topics in the upcoming class sessions. We found this method to be very helpful in both improving students' learning experience as well as creating an open channel for students to communicate their misunderstandings with confidence and feel being heard and supported.**

***Index Terms*—Topic modeling, GSDMM, NLP, reflection analysis, software engineering**

I. Introduction

Reflective learning occurs when an individual's inquiry into a topic or event helps them develop their understanding and/or remove their misconceptions about that certain topic [1]. Reflective writing as a teaching technique is controversial due to its epistemological shortcomings [2], yet it is critical for students to better cope with ambiguity and complicated circumstances [3]. Reflection has been shown to help both students and teachers recognize and address areas of improvement [4]. Research suggests that students who use different methods to evaluate their learning and are critical of their achievements are better performers and get better academic outcomes [5]. Reflections can help students improve self-awareness, critical thinking, and problem-solving skills [6]. Students' feedback on their educational experiences influences the development of transferable skills. Students who self-assess and take responsibility for their education may flourish as self-learners [7].

Critiquing one's own beliefs and assumptions about their potential and progress is an important introspection that impacts cognition. Instructors and educators also benefit from reflection. Through critical inquiry and self-reflection, they can address ethical ideals and reevaluate content delivery approaches to improve students' learning. There are different methods for helping students have a reflection on their learning and get feedback from them. Polling students is a good approach to getting their opinions about a subject. Another option is to provide students with one-minute writing tasks that ask them to reflect on their experiences in class [8]. Because one-minute reflections are short students get more focused on synthesizing their ideas and concisely writing them. Although evaluating students' short reflections and extracting the challenging topics from them is more feasible for more instructors compared to reading long narratives of reflections, still this requires a substantial amount of time. Automating this process of text mining and extracting topics from them can greatly help educators to focus on the pace and modification of content delivery in their courses to better address students' needs.

To extract information from the voluminous text data on student evaluations, clustering short text documents is one of the most important text analysis strategies available to researchers [9] [10]. Compared to ordinary text clustering, short text clustering presents a greater challenge [11]. It is because of the briefness of the text and sparsity, noise, and high dimensionalities that are added into the text analytic process [12]. Short sentences include a lot of noise and only give a few contextual hints when applying typical data mining methods to them. As a result, researchers have experimented different techniques to short text clustering such deep representation learning, convolutional neural networks and statistical semantics [11] [13] [14]. To uncover latent semantic structure from text corpora, traditional topic modeling methods like probabilistic latent semantic analysis [15] and latent Dirichlet allocation (LDA) [16] are extensively used. These algorithms treat each document as a collection of topics, each with its word distribution. In this paper we utilize the Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM) model [17] to extract the themes from students' short reflections on their challenges and learning outcomes. The GSDMM model which performs well on short text data, assumes the input document contains only one topic as opposed to the aforementioned models that consider a document as a distribution of topics. In the following sections of this paper, we briefly overview the related work and present our model with a sample case study of implementing the model in a software engineering class.

II. Related Work

Weekly self-reflection writing allows students to reflect on the learning outcomes as well as their misconceptions. These reflection feedbacks help instructors identify gaps in student understanding and generate action items for the following sessions. However, analyzing student self-reflections is a tedious process that may be improved with automation.

Educators in academia have conducted surveys throughout the years to get feedback from students. Structured questions and Likert-scale [18] replies have been traditionally used to assess the students' opinions and feedback. Large amounts of data that are narrative make it difficult to objectively and properly extract topics from them. Researchers in [19] propose an assessment method that only asks 1 to 3 general questions and analyses the answers by applying Natural Language Processing (NLP) methods. Student-Directed Discussion Surveys (SDDS) technique gives a more important tool to augment or perhaps replace Likert scaled surveys. To evaluate their SDDS approach, students were given two open-ended questions three times during the semester. In order to do statistical comparisons, they quantified and rescaled the 8-9 Likert scaled responses of the larger calculus population into the normalized values between the range [-1, 1], with -1 being the most negative and 1 being the most positive. The SDDS findings indicate that this survey approach may be less biased than typical Likert scaled surveys, which often ask leading questions and are prone to acquiescence bias. The substantially greater variation associated with the Likert survey findings is indicative of this bias. SDDS eliminates acquiescence bias and false variation associated with negative or positive polar reactions.

The necessity to evaluate shorter messages has grown in importance as microblogging websites and social media platforms such as Twitter and Facebook have grown in popularity. Topic modeling is a kind of text mining technique that is used to uncover hidden subjects inside big collections of texts. The GSDMM is an excellent model for this purpose. The purpose of the work presented in [20] is to test the hypothesis that GSDMM outperforms LDA when applied to short text. In this model the assumption is that the input document has only one topic and short text inherently includes one main topic. Both models were tested on one large and two short text corpora. GSDMM outperformed LDA on coherence while LDA outperformed GSDMM on topic consistency.

In [21] the researchers provide a solution model that automates the analysis of students' self-reflections using NLP and text mining methods. The work in this study tests whether machine learning algorithms can identify difficult course topics from informal self-reflections. This research also analyzes how self-reflection can help instructors improve course content and results. To solve the issue, multiple solution models were tested in content mining for an undergraduate Information Systems course. LDA-bigrams, GSDMM-bigrams, and Word2Vec-based clustering models were employed. The result shows the best performing model was the Word2Vec+K-Means, while the GSDMM suffered data disparity issues. Another application of short text clustering is the analysis of the customers' review on products and services offered by providers on the web. This helps both providers and customers to have a summary of reviews which potentially will impact their decision of choosing certain products. LDA has been implemented in several text mining areas, and various studies have analysed text-based consumer complaints using this method. [22] [23]. While LDA is the most prominent technique to topic modeling, GSDMM enhances topic modeling for shorter texts. In one study [24] the authors compared the performance of LDA with GSDMM on the 6963 review data. For analysis purpose they applied two libraries of LDA which are GenSIm and Mallet. As a result the comparison of total three methods of GenSim LDA, Mallet LDA and GSDMM showed the GSDMM outperformed the other methods on analysing the short textual data.

III. Methodology

In this section, we present our approach to developing a model for automating the analysis of students' short reflections. In this study we used the minute paper approach [8] in which students provide short answers to two questions about what they learned and what they didn't learn in each class. Students' answers to these questions were utilized as inputs for the model, and the results of the model's analysis are presented in a dashboard to the instructor. To implement our approach we used GSDMM method [17] for extracting relevant topics from each session's dataset. The high-level architecture of the proposed approach is depicted in Fig 1.

Cleaning and preparing data is the important initial step in text processing by NLP models. Textual data tend to have much noise and are unstructured. The phase of removing noise from text and making it ready to be processed in developed models is called preprocessing. In this study we applied the standard text-preprocessing methods of tokenization, stemming, lemmatization, and stop words removal to the datasets

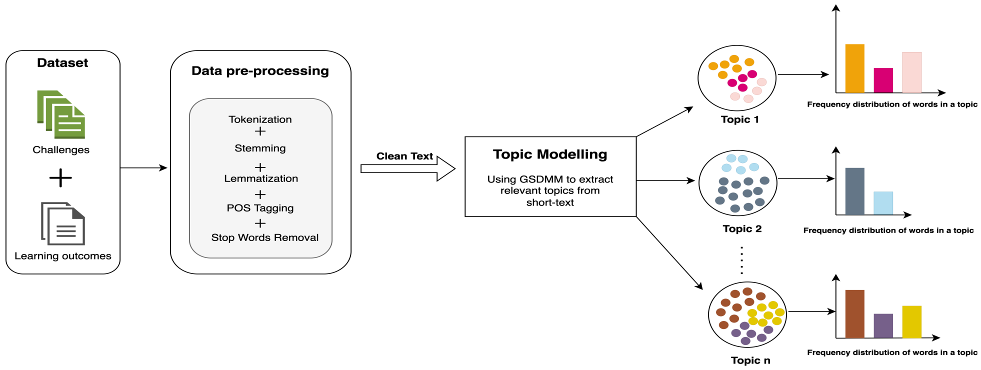

Fig. 1: High level architecture of proposed approach

of the challenges and learning outcomes that were collected from each class session.Tokenization is used to break down the dataset's raw text into short words and phrases. We have used a multi-word expression (MWE) tokenizer in this study. NLTK provides a rule-based ”add-on” tokenizer called the multi-word expression tokenizer. Tokens may be reorganized into multi-word phrases once the text has been parsed and tokenized using a tokenizer function. Several word truncation rules were applied to word tokens to eliminate morphological and inflectional ends from them. PorterStemmer, one of the most well-known English stemming algorithms, was used to do this.

As opposed to a stemmer, which uses language-based principles, lemmatizers need a full vocabulary and the knowledge of words' morphology to correctly lemmatize texts. WordNet [25] is a large, open, and free linguistic database that shows how words in the English language are linked together systematically. We have used the lemmatization process to turn the root stem into root words. Root stems must be changed to root words for similarity-checking process. After cleaning the data with the above pre-processing methods, we pass the data to a clustering model to extract relevant topics of a session from the corresponding data.

A. *Topic Modeling using GSDMM*

Topic modeling is the process of analysing unstructured and unlabeled text input and categorizing it into the topics it represents. LDA is a prominent topic modeling technique. However, LDA suffers from two fundamental drawbacks: 1) it presupposes that the text has a variety of topics, and 2) it operates poorly when dealing with a short text. We have used another recently introduced topic modeling algorithm known as GSDMM to detect significant topics from a short text. GSDMM extracts only one topic per document.

In the following, we present a case study of applying the proposed model to students' short reflections in a software engineering class.

B. *Case study*

During Spring 22 the students of a software engineering class were asked to fill out a form which was called an ”exit ticket” at the end of each lecture as a part of the course syllabus. The total number of students was 25 and they were asked to highlight what they learned and what they found challenging in each session of the class. To test the proposed model on the actual reflection data we used data from 9 class sessions during the semester. The learning outcomes and challenges were stored as separated datasets fed into the model. After prepossessing the collected textual data, the topics of each session of the class were extracted and exported in JSON format. The extracted topics and their frequency are plotted on a dashboard (user interface) for visualization. The bar graph depicts the frequency distribution of topics of both challenges and learning outcomes. We assigned a weight to the topics such that the topics that were repeated by more students in each session got higher weight. The dashboard presents the list of topics in order based on their assigned weights. A sample output of the result for one session is presented in figure 2. For example here in the learning outcome section, two students mentioned the concept of ”Node package” was interesting to them while other topics were mentioned only once by other students. As a result this keyword gets higher weight in learning outcomes. The topics of each session are presented on one page and the instructor can navigate through multiple sessions. Furthermore in each session by clicking on each topic the corresponding responses about that topic will be listed so that the instructor can get a better idea of the context.

C. *Results and Discussion:*

As this work in progress is at its early stages of development we have not collected substantial data to evaluate its accuracy and performance in multiple classes. We used qualitative evaluation by a human expert to check the extracted topics and if they match students' responses. Our qualitative evaluation shows that we have about 75 percent accuracy in

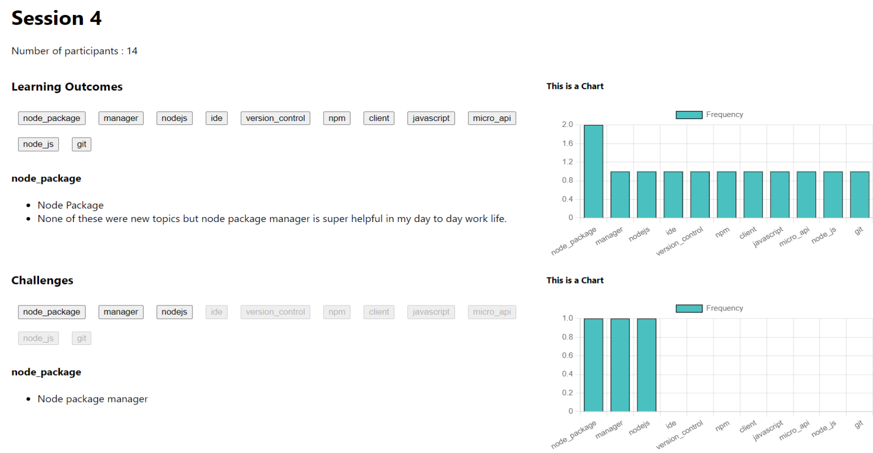

Fig. 2: Topic vs frequency graph of a session

the extracted topics with a limited sample size. More data is required to conclude the performance of this model. In the evaluation process, we found out that the algorithm needs some optimization to better extract the topics in the context. For example, if a student adds a negation at the beginning of a phrase in the challenge question, i.e. "I didn't find SOAP to be a challenge for me", the algorithm should not consider SOAP as a challenge. One may ask how the proposed model differs from world-cloud representations of the topics which are built-in in many online platforms. One main advantage of this model is that it extracts the topics in the given context not just based on the frequency. For this purpose, we have developed a custom dictionary that includes a list of context-related keywords. The unigrams are first extracted from the text input and are listed based on the frequency. The context-related keywords are marked to be added to the custom dictionary. When the algorithm extracts the topics using GSDMM it passes the topics through this filter and if the topic is outside the course context (custom dictionary) it will be dropped. We are working on the algorithm to make this process automated so that the model can generate the dictionary by learning from the context without human intervention.

IV. Conclusion and Future Work

In this work, we implemented a topic modeling method using GSDMM to extract relevant topics from students' reflections on their challenges and learning outcomes in a software engineering class. The results of this study seem to be promising in extracting relevant topics for short texts. This method used topic distribution frequency in visualizing the data in the developed dashboard. In future works, we will improve the model by adding the following functionalities: 1) tracking the learning outcome and challenges single student throughout multiple sessions (the model currently just navigates through sessions) 2) optimizing the algorithm to extract topics more semantically and to automatically develop a custom dictionary of context-related words for each course, and 3) finally we will use the developed model in different classes to collect additional data to have a more concrete data analysis and evaluation on the performance and accuracy of the model. This system has great potential to help both educators and students to provide a more productive learning experience. Instructors can make interventions in each class and course content based on the feedback they get from students. furthermore, by tracking the changes in students' challenges and learning outcomes from one session to another they can monitor their progress. This tool can also be used to form teams of students that can learn from each other based on what they learned and what they didn't understand well. Thus it can promote peer learning in classes.

References

[1] E. M. Boyd and A. W. Fales, "Reflective learning: Key to learning from experience," *Journal of Humanistic Psychology*, vol. 23, no. 2, pp. 99–117, 1983. [Online]. Available: https://doi.org/10.1177/0022167883232011

[2] R. Jordi, "Reframing the concept of reflection: Consciousness, experiential learning, and reflective learning practices," *Adult Education Quarterly*, vol. 61, no. 2, pp. 181–197, 2011. [Online]. Available: https://doi.org/10.1177/0741713610380439

[3] J. Fullana, M. Pallisera, J. Colomer, R. F. Peña, and M. Pérez-Burriel, "Reflective learning in higher education: a qualitative study on students' perceptions," *Studies in Higher Education*, vol. 41, no. 6, pp. 1008–1022, 2016. [Online]. Available: https://doi.org/10.1080/03075079.2014.950563

[4] A. Bandura *et al.*, "Prentice-hall series in social learning theory. social foundations of thought and action: A social cognitive theory," *Englewood Cliffs, NJ*, 1986.

[5] A. Ghanizadeh, "The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education," *Higher Education*, vol. 74, no. 1, pp. 101–114, 2017.

[6] F. Perels, T. Gürtler, and B. Schmitz, "Training of self-regulatory and problem-solving competence," *Learning and Instruction*, vol. 15, no. 2, pp. 123–139, 2005, development and Modification of Motivation and Self-regulation in School Contexts. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095947520500023X

[7] B. M., "The real powers of meditation," *New Scientist*, vol. 209, no. 2794, p. 32 – 35, 2011, cited by: 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-78650960934&doi=10.1016%2fS0262-4079%2811%2960048-1&partnerID=40&md5=8124a7f24f421f3cce6f94e9d02af9d7

[8] N. Dehbozorgi and S. MacNeil, "Semi-automated analysis of reflections as a continuous course," in *2019 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2019, pp. 1–5.

[9] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "A brief survey of text mining: Classification, clustering and extraction techniques," 2017.
[10] P.-y. Zhang and C.-h. Li, "Automatic text summarization based on sentences clustering and extraction," in *2009 2nd IEEE International Conference on Computer Science and Information Technology*, 2009, pp. 167–170.
[11] J. Xu, P. Wang, G. Tian, B. Xu, J. Zhao, F. Wang, and H. Hao, "Short text clustering via convolutional neural networks," in *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 2015, pp. 62–69.
[12] G. S. Reddy, "Dimensionality reduction approach for high dimensional text documents," in *2016 International Conference on Engineering MIS (ICEMIS)*, 2016, pp. 1–6.
[13] Z. Wang, H. Mi, and A. Ittycheriah, "Semi-supervised clustering for short text via deep representation learning," *arXiv preprint arXiv:1602.06797*, 2016.
[14] S. Seifzadeh, A. K. Farahat, M. S. Kamel, and F. Karray, "Short-text clustering using statistical semantics," in *Proceedings of the 24th international conference on world wide web*, 2015, pp. 805–810.
[15] T. Hofmann, "Probabilistic latent semantic analysis," 2013.
[16] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," in *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds., vol. 14. MIT Press, 2001. [Online]. Available: https://proceedings.neurips.cc/paper/2001/file/296472c9542ad4d4788d543508116cbc-Paper.pdf
[17] J. Yin and J. Wang, "A dirichlet multinomial mixture model-based approach for short text clustering," 2014.
[18] A. Joshi, S. Kale, S. Chandel, and D. K. Pal, "Likert scale: Explored and explained," *British journal of applied science & technology*, vol. 7, no. 4, p. 396, 2015.
[19] K. Hood and P. K. Kuiper, "Improving student surveys with natural language processing," in *2018 Second IEEE International Conference on Robotic Computing (IRC)*, 2018, pp. 383–386.
[20] J. Mazarura and A. de Waal, "tsc," in *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, 2016, pp. 1–6.
[21] O. De Lin, S. Gottipati, L. S. Ling, and V. Shankararaman, "Mining informal amp; short student self-reflections for detecting challenging topics – a learning outcomes insight dashboard," in *2021 IEEE Frontiers in Education Conference (FIE)*, 2021, pp. 1–9.
[22] K. Bastani, H. Namavari, and J. Shaffer, "Latent dirichlet allocation (lda) for topic modeling of the cfpb consumer complaints," *Expert Systems with Applications*, vol. 127, pp. 256–271, 2019.
[23] A. Dewinta and M. I. Irawan, "Customer complaints clusterization of government drinking water company on social media twitter using text mining," in *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 2021, pp. 338–342.
[24] S. İ. Omurca, E. Ekinci, E. Yakupoğlu, E. Arslan, and B. Çapar, "Automatic detection of the topics in customer complaints with artificial intelligence," *Balkan Journal of Electrical and Computer Engineering*, vol. 9, no. 3, pp. 268–277, 2021.
[25] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.