Topic Modeling for Evaluating Students' Reflective Writing: a Case Study of Pre-service Teachers' Journals

Ye Chen, Bei Yu, Xuewei Zhang, Yihan Yu
Syracuse University
New York, USA
{ychen129, byu, xzhang77, yyu41}@syr.edu

ABSTRACT

Journal writing is an important and common reflective practice in education. Students' reflection journals also offer a rich source of data for formative assessment. However, the analysis of the textual reflections in class of large size presents challenges. Automatic analysis of students' reflective writing holds great promise for providing adaptive real time support for students. This paper proposes a method based on topic modeling techniques for the task of themes exploration and reflection grade prediction. We evaluated this method on a sample of journal writings from pre-service teachers. The topic modeling method was able to discover the important themes and patterns emerged in students' reflection journals. Weekly topic relevance and word count were identified as important indicators of their journal grades. Based on the patterns discovered by topic modeling, prediction models were developed to automate the assessing and grading of reflection journals. The findings indicate the potential of topic modeling in serving as an analytic tool for teachers to explore and assess students' reflective thoughts in written journals.

Categories and Subject Descriptors

K.3.1 [**Computer Use in Education**]: *Computer-assisted instruction (CAI)*; I.2.7 [**Natural Language Processing**] *Text analysis*; I.5.4. [**Applications**] *Text processing*

General Terms

Algorithms, Measurement, Performance, Experimentation.

Keywords

Text mining, Topic modeling, LDA, Reflection, Journal writing, Automated grading, Learning analytics, Education.

1. INTRODUCTION

According to Boud et al. [1, p.3], reflection practices are those "intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciation." Journal writing is one of the commonly used reflective activities [2, 3]. In writing their journals after class, students step back and reflect on how they went through the learning activities, and what they have learned. They can reconstruct the class process, and see the separate aspects of the class together and derive important and meaningful knowledge from it [4]. Besides, in reflection, they make their own interpretation of the class activities and content to make sense of their learning experience. Mezirow believes that learning happens when they use this reflection-based interpretation to guide subsequent actions and understandings [5].

To enhance students' learning through reflection, teachers need to know whether students are engaged in reflective practice, and to probe what students are exactly caring and thinking about [3]. What course content was reflected in their journals? Did they talk about things not covered in class? How deep are they engaged in reflective thoughts? By answering these questions, teachers can assess student learning outcome, provide informative feedback, and adjust future teaching. Usually, teachers manually read and grade journals using a rubric or coding scheme [6, 7], which is labor-intensive and time-consuming, especially for large classes.

In this study we propose a new method that uses topic modeling to automatically explore and assess students' reflective writing. By extracting an optimal number of topics from the pre-service teachers' weekly journals for an education course, we were able to examine the relevance between common topics in student journals and the weekly content, additional thoughts that students mentioned in their journals, and the extent to which the topic relevance correlates with the received grade. Finally, we used the topic relevance factors, combined with other textual features like word count, to develop a writing quality prediction model.

2. LITERATURE REVIEW

Topic modeling is a family of computational methods that facilitate exploratory analysis of large text collections, extracting the common themes discussed in the corpora. LSA, an early topic modeling method, extracts salient topics by examining word co-occurrence. It has been widely used in analyzing and grading students' textual work. Sorour et al. [8] predicted students' final grade based on the topics models generated from using LSA on students' comments in course evaluation. Graesser et al. [9] applied LSA in an intelligent tutoring system that is capable of comprehending and grading students' written answers in tutorial dialogue. Wiemer-Hastings and Graesser [10] generated idea outlines of students' essay by clustering analysis on LSA semantic space. In addition, LSA was also used to track online learners' conceptual development [11], assess students' comprehension in reading tasks [12, 13], generate personalized feedback for students' summary writing [14], and evaluate students' contribution to group discussion [15]. However, Crain et al. [16] pointed out that LSA is problematic in overfitting training data as its parameters increase linearly with the number of documents.

Later, Blei et al. [17] developed Latent Dirichlet Allocation (LDA), a probability-based technique with more simplicity. Similar to LSA, LDA recognizes the coherent topics by finding the pattern of co-occurrence of words. But LDA treats each document as a random mixture of topics. Each topic can be understood as a probability distribution over a collection of words; while at the same time, a single document is represented as

LAK '16, April 25-29, 2016, Edinburgh, United Kingdom
© 2016 ACM. ISBN 978-1-4503-4190-5/16/04…$15.00
DOI: http://dx.doi.org/10.1145/2883851.2883951

a probability distribution over these topics. The Mallet topic modeling package [18] contains an extremely fast implementation of LDA, and has been used in many applications.

LDA has seen increasing use in educational context. The applications have been of three broad types. First, LDA has been used as a learning analytic tool in MOOCs to analyze students' forum discussions. Reich et al. [19] adopted LDA to identify themes and patterns in students' MOOC forum discussion. Ramesh et al.[20] built topic models using LDA to explore forum discussion and predict students' survival. Ezen-Can et al.[21] applied LDA, along with clustering analysis of online discussion, to extract the topical themes discussed by students. Based on LDA, Hsiao and Awasthi [22] proposed a Topic Facet Model to uncover the content latent structure of students' discussion posts.

Second, researches have employed LDA to analyze students' essays. Gibson and Kitto [23] identified students' thoughts by discovering significant topics in their reflective texts. Southavilay et al. [24] used LDA and its extension DiffLDA to extract topics in students' collaborative writing with cloud writing tools (e.g. google doc). They tracked topic evolution during their writing process, and generated topic-based collaboration networks by linking topics with author contribution. Kakkonen et al. [25] applied LDA to automatic essay grading.

Third, LDA has been used for the purpose of document analysis or recommendation. Sekiya et al. [26] proposed a LDA-supported method to analyze course syllabus and compare curriculums. Kovanović, et al. [27] discovered the key themes in MOOC-related mainstream news reports. Chen et al. [28] investigated the common topics in learning analytics community by analyzing Twitter archives. Kandula et al. [29] developed a system with LDA to recommend relevant educational materials to diabetic patients.

Overall, topic modeling has not been fully exploited that looked into mining students' reflective writing in chronicle setting and automating the grading of their journals. In this study, we adopted the topic modeling technique LDA, as implemented in the Mallet package, to address these issues.

3. METHODS

3.1 Research context and data preparation

The context of this study was a 6-week undergraduate course in the School of Education at a northeastern university in the United States. The purpose of this course was to educate pre-service teachers on integrating technologies into teaching and learning. Students in this class were encouraged to learn through hands-on practices and writing reflection journals. In the reflection journal, students were encouraged to include a narrative description of their class experience and pick the things personally important to them.

Based on the journal content, the instructors graded students' journals into three levels. Level 3 is Excellent, if students demonstrate deep or integrated thinking and search for meaning inherent in class activities. For example, they may explain the pedagogical intention and principles behind activities, use analogy or examples to make a concept real and tangible, refer to external resources to justify their stands, or personalize the class content by connecting to their own experience. Level 2 is Great, if students comprehensively summarize the learning content or give detailed description of the class process. At this level, they configure their class experience by memorizing and recalling. Level 1 is Good, if students make simple description of class experience and reflect on the surface aspects of learning activities or knowledge content.

A total of 367 journals were collected from 80 students in three sections of the class. Table 1 showed the descriptive statistics of the data. The distribution of the grade levels was illustrated in Table 2. The main learning content and activities for each week are summarized in Table 3.

Table 1: Description of the data set

	2014 Spring	2014 Fall	2015 Spring	Sum
Week 1	31	16	25	72
Week 2	31	15	26	72
Week 3	31	15	26	72
Week 4	32	14	27	73
Week 5	33	15	30	78
Sum	158	75	134	367

Table 2: Grade level distribution

Grades	Excellent (3)	Great (2)	Good (1)
Journal entries #	121	130	116

Table 3: Weekly Topic

Week	Topic
Week 1	Course Introduction and Website Evaluation
Week 2	Word/Powerpoint as Teaching Tools
Week 3	Using Excel as a Tool
Week 4	Electronic Communication Tool
Week 5	Introduction to Assistive Technology

We retrieved the online journals from the Blackboard System, and then converted them from the Word format to text files using the tool *textutil*. The data were then ready for analysis.

3.2 Topic modeling

We first used the topic modeling algorithm provided in the MALLET toolkit to discover common topics discussed in students' reflective writings.

A major challenge for tuning topic models is the choice of *K*, the number of topics. Selecting the appropriate topic number is important for obtaining meaningful and useful topics [30]. If the number is too large, the topics will be redundant; if the number is too small, the different categories can't be separated from each other and the topics will be too broad.

K is usually tuned manually by running topic modeling repeatedly for a set of potential *K* values, and then choosing the best *K* value with the most sense-making results based on prior knowledge in the task domain. In our task, we explored the *K* value in the range between 5 and 30 based on the following rationale. Because this is a six-week course with five weeks of instruction and one week of review, we assume the students would at least reflect on the weekly content from Week 1 to 5, and therefore expect $K>=5$. The students may discuss other topics that they consider relevant, but the additional topics are not expected to be large, at most several new and common topics each week Thus we assume $K<=30$. After fitting topic models with *K* from 5 to 30, the first author, who was also the instructor of this class, determined that the topic model with $K=10$ makes the most sense based on her prior knowledge. Each topic in the topic model is represented by a list of significant keywords, for example, 20 top keywords. We then manually labeled these topics based on the keywords in each topic.

The manually tuned results may be subject to human bias. In recent years, more research has been conducted to objectively evaluate the topic models [31] and estimate K, such as the "term-centric stability" measure developed by Greene et al. [32]. The basic idea behind this technique is that, a topic model with the optimal number of topics will be more robust and consistent in producing similar solutions on data from the same source [33].

Given a range of K values, e.g. (5, 30) in our task, the term-centric stability measure first fits 26 topic models, one for each K value. For each candidate K value, 10 sub-samples would be generated from the data set. Then 10 topic models would be generated, one for each sub-sample, resulting in 260 topic models in total. Since every topic is represented by a list of most significant keywords, an agreement score between two topics is defined as the level of agreement between the two word lists, inversely weighted by the word ranks. An agreement score between two topic models is further defined as the average agreement between every pair of term lists in these two topic models. The final stability score for each K is then calculated by averaging the agreement score among all the 10 topic models. The higher the stability score, the more robust the model is.

In our task, we found the highest stability score when K=10. See Figure 1 for the stability scores for K from 5 to 30. Thus, the manual tuning result and the automated tuning result consistently support that there are 10 common topics in our data. We then chose this model as our final model, and then obtained the topic distribution for each journal.

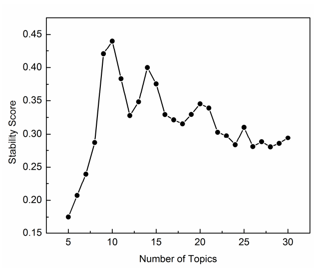


Figure 1: Stability analysis of different topic solutions

3.3 Correlation analysis and grade prediction

Several factors, such as journal's relevance to common topics, the sentiment that students showed in their writing, and the length of the writing may be related to the quality of the writing. Thus we did correlation analysis to test the significance of these factors, which would then be used to build prediction model. Students' sentiment level was analyzed using the program SentiStrength [34]. Polarity scores of both positive and negative sentiment were reported as the output, which ranged from 1 to 5 and -5 to -1 respectively.

Three types of classification algorithms were used to build prediction models: Naïve Bayes, Decision Tree J48, and Support vector machines (SVM). We selected the three algorithms because they have been commonly used in data mining of students-generated data in education context (e.g. [35, 36]). Three algorithms were implemented in Weka, and the accuracy of prediction models was evaluated through 10-fold cross validation.

4. RESULTS & DISCUSSIONS

4.1 Common topics in reflection journals

10 common topics were discovered from the students' reflection journals. Table 4 summarizes the samples of top keywords and the manual label that we assigned to each topic. We also compared the weekly teaching content against the discovered topic clusters, and found that our topic model has well recognized the weekly teaching content.

Besides the weekly content, students also mentioned some additional topics. Four such topics were discovered: age & computer (cluster #3), instructional strategies (cluster #9), positive sentiment toward class activities (cluster #7), and the topic about class activities in a general sense (cluster #5).

Table 4: Topic and the key terms

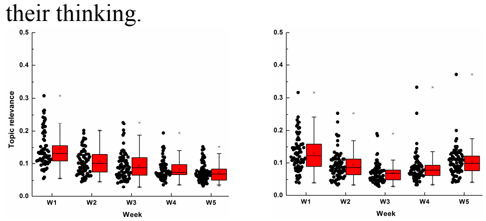
Cluster #	Topic label	Keywords
0	Word-PPT-Story	activity students word story create powerpoint make created read creating power point writing pumpkin order stories show book creative microsoft
1	Class introduction (activity)	lesson work teaching activities learn understand group groups ideas teach concept working topic video lessons instructional task strategy subjects walking
2	Assistive technology	picture children pictures words assistive disabilities ipad centers visual student technologies instructions software web sentences young read called reading station
3	Age & computer	teacher good today information questions computer easy ide made kids making time find elementary answer interactive younger age website specific
4	Excel use	excel grade create class make data grades reflection microsoft assignment survey classes book directions teachers graphs instructions question taught chart
5	General	students great program student learn tool learning classroom tools feel worked helps involved extremely step makes easily walk peers opportunity
6	Class introduction (content)	technology learning ways ide reflection websites classrooms skills types computers integration education experience discussed content helped stations multiple focused teach
7	Positive sentiment-assignment	class helpful project mini ide interesting fun lot thought reflection things found week enjoyed people gave idea beneficial thing today?€?s
8	Communication technology	teachers website communication teacher synchronous asynchronous edmodo school time assignments site social webinar discussed parents educational online diagram inspiration talked
9	Instructional strategies	classroom class learned future today important strategies programs incorporate make knowledge integrate management easier asked share observed setting give apply

4.2 Topic evolution over time

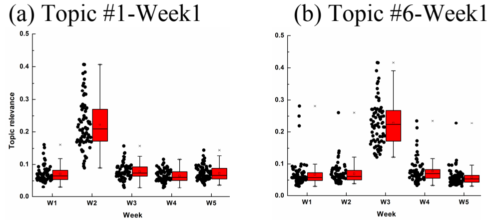
The topic model not only lists the most significant keywords for each topic cluster, but also estimates the topic distribution for each document. We hypothesized that if the topic model makes sense, then each weekly topic should yield highest relevance among journals in the corresponding week, and the four additional topics should occur more evenly in each week.

The results in Figure 2 confirmed our hypothesis. In Figure 2, the x axis represents the journal entries from Week 1 to Week 5, and the y axis represents the relevance proportions of a journal entry to the topic in a particular week. Figures 2 (c)-(f) shows that journal entries in Weeks 2-5 were significantly more relevant to the topics in the corresponding weeks than other weeks. The average relevance to Week 1 topics, which were of overview nature, was slightly higher than that to other weeks' topics.

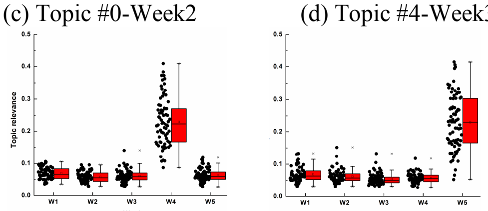
Figure 2 serves as a validity check of our finding that the topic model is able to accurately recognize the weekly teaching topic from students' reflection writing, which has laid a foundation for automatically assessing to what extent students reflect on their learning experience and how much knowledge they have taken away from class. If students reflect at a surface level, they might not clearly articulate what they have learned. In this case, there would be more statements simply expressing opinion as "*I like today's class*", "*This was fun*", or "*I learned a lot from class activities*" without pointing to a particular topic. Or, students might even talk about things that are irrelevant to the class content. In contrast, if students go beyond the surface level, they are more likely to connect to specific teaching topics or learning activities and to retrieve more details about their learning experience. The more such connections they have built, the more possible that they construct meaning and understanding from their experience. The elements they retrieved from the reality can serve as the mental objects that they can manipulate on in order to develop their thinking.



(a) Topic #1-Week1 (b) Topic #6-Week1



(c) Topic #0-Week2 (d) Topic #4-Week3

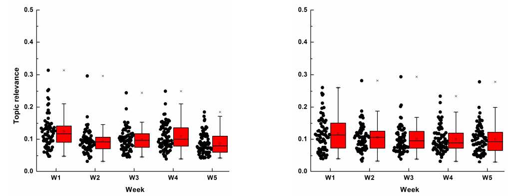


(e) Topic #8-Week4 (f) Topic #2-Week5

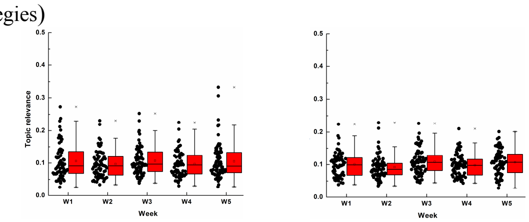
Figure 2: Distribution of weekly topics

In reflections, students were encouraged to identify the things of their personal importance. Thus, we can discover what students care about by examining the most common themes in addition to the weekly teaching topics. In each week's class, students explored different computer programs and study how to integrate them into teaching. It appeared that, in reflections, students were interested in picking practical matters. For example, they mentioned/discussed the appropriate age of using the programs, e.g. "*A question that I thought of while working on this was what age group is excel applicable towards?*", "*it may not be appropriate for students younger than fifth grade*", or proposed the instructional strategies that could help integrate the technology into their future teaching, e.g. "*Some classroom strategies I learned about in this class are to put the students in groups in different ways*". Their reflections also involved sentiment information expressed through terms like *"helpful", "interesting", "fun", "enjoyed",* and *"beneficial"*.

Figure 3 illustrated the relevance between the weekly reflection and these four additional topics. These four topics were almost evenly distributed across five weeks, indicating that students maintain their interests on these topics throughout the course, regardless of the weekly teaching content. One of the benefits of mining the frequently mentioned themes in students' journals is to help teachers get to know about what knowledge is valuable from students' perspective and what knowledge have been retained in students' mind after class. The findings based on topic modeling could guide teachers to develop future teaching content and activities that are tailored toward students' needs, and thus boost their motivation in learning. For example, in the studied context, when teaching a certain type of technology, the teachers might consider to incorporate more discussions on the appropriate age for using this technology, and more inquires that investigate the effective strategies for integrating the technology to teaching practices.



(a) Topic #3 (Computer & Age) (b) Topic #9 (Instructional Strategies)



(c) Topic #7 (Positive Sentiment) (d) Topic #5 (General)

Figure 3: Distribution of additional topics that are personally important to students

4.3 Correlation analysis & Grade prediction

Weekly topic relevance (Pearson $r=.15$, $p<.01$) and word count (Pearson $r=.607$, $p<.001$) were found to be significantly correlated with journal grade. No significant correlation was found between journal grades and any sentiment levels (positive, negative, and overall sentiment) or any additional topics.

Using the word count and the weekly content relevance as attributes, we built three classifiers based on decision tree, naïve Bayes, and SVM algorithms respectively. All classifiers outperformed the random guess baseline (.333). Among them, the Naïve Bayes model obtained the highest accuracy .651. The J48 model performed at similar level of accuracy .649, while SVM's accuracy .594 is slightly lower. Decision tree model illustrated the rules for prediction: all writings with length <=183 words were classified as level 1; those with length in the (183, 262] range were classified as level 2; the ones longer than 262 words were further examined by their weekly content relevance – level 2 if relevance <= .15, level 3 otherwise.

5. CONCLUSIONS

In this study, by using the topic modeling approach, we conducted an exploratory analysis of pre-service teachers' reflection journals. The results suggest the potential of topic modeling in analyzing reflection journals, and indicate that topic modeling can contribute to the construction of analytic tool for formative assessment of students' learning. Future work may include experimenting on more factors that might relate with journal grades, and building a comparison-oriented prediction model by extracting the characteristics of Excellent level journals.

6. REFERENCES

[1] Boud, D., Keogh, M., and Walker, D. 1985. Reflection: Turning experience into learning. London: Kogan Page.

[2] Boud, D. 2001. Using journal writing to enhance reflective practice. New Directions for Adult and Continuing Education. 2001, 90 (2001), 9-18. DOI=10.1002/ace.16

[3] Kember, D. 1999. Determining the level of reflective thinking from students' written journals using a coding scheme based on the work

of Mezirow. International Journal of Lifelong Education. 18, 1(1999), 18-30. DOI=10.1080/026013799293928

[4] Collins, A., and Brown, J. S. 1988. The computer as a tool for learning through reflection. Springer US

[5] Mezirow, J. 1990. How critical reflection triggers transformative learning. Fostering critical reflection in adulthood, Jossey-Bass, San Francisco, CA, 1-20.

[6] Dyment, J. E., and O'Connell, T. S. 2011. Assessing the quality of reflection in student journals: a review of the research. *Teaching in* Higher Education. 16, 1(2011), 81-97.

[7] Lee, H. J. 2005. Understanding and assessing preservice teachers' reflective thinking. Teaching and Teacher Education. 21, 6(2005), 699-715.

[8] Sorour, S., Goda, K., and Mine, T. 2015. Correlation of topic model and student grades using comment data mining. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (Kansas City, MO, USA March 4-7, 2015). ACM, New York, NY, 441-446.

[9] Graesser, A. C., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Tutoring Research Group, T. R. G., and Person, N. 2000. Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments*. 8, 2(2000), 129-147.

[10] Wiemer-Hastings, P., and Graesser, A. C. 2000. Select-a-Kibitzer: A computer tool that gives meaningful feedback on student compositions. *Interactive Learning Environments*. 8, 2(2000), 149-169.

[11] Wild, F., Haley, D., and Bülow, K. 2011. Using latent-semantic analysis and network analysis for monitoring conceptual development. *Journal for Language Technology and Computational Linguistics*. 26, 1(2011), 9-21.

[12] McNamara, D. S., Boonthum, C., Levinstein, I. B., and Millis, K. 2007. Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In *Handbook of latent semantic analysis*, T.Landauer, D. S.McNamara, S.Dennis, and W.Kintsch Eds. Mahwah, NJ: Erlbaum, 227-241.

[13] Millis, K., Magliano, J., Wiemer-Hastings, K., Todaro, S., and McNamara, D. S.2007. Assessing and improving comprehension with latent semantic analysis. In *Handbook of latent semantic analysis*, T.Landauer, D. S.McNamara, S.Dennis, and W.Kintsch Eds. Mahwah, NJ: Erlbaum, 207-225.

[14] Wade-Stein, D., and Kintsch, E. 2004. Summary Street: Interactive computer support for writing. *Cognition and Instruction*. 22, 3(2004), 333-362.

[15] Streeter, L. A., Lochbaum, K. E., LaVoie, N., and Psotka, J. E. 2007. Automated tools for collaborative learning environments. In *Handbook of latent semantic analysis*, T.Landauer, D. S.McNamara, S.Dennis, and W.Kintsch Eds. Mahwah, NJ: Erlbaum, 279-290.

[16] Crain, S. P., Zhou, K., Yang, S. H., and Zha, H. 2012. Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. In *Mining Text Data,* C. Aggarwal, C., C. Zhai Eds. NY: Springer, 129-161.

[17] Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3 (2003), 993-1022.

[18] McCallum, A. 2002. MALLET: A Machine Learning for Language Toolkit. Retrieve from: *http://mallet.cs.umass.edu*.

[19] Reich, J., Tingley, D. H., Leder-Luis, J., Roberts, M. E., and Stewart, B. 2014. Computer-assisted reading and discovery for student generated text in massive open online courses. *Journal of learning analytics*. 2, 1(2014), 156-184.

[20] Ramesh, A., Goldwasser, D., Huang, B., Daume III, H., and Getoor, L. 2014. Understanding MOOC discussion forums using seeded LDA. In *Proceeding of 9th Workshop on Innovative Use of NLP for Building Educational Applications* (Baltimore, MD, USA, June 26), 28-33.

[21] Ezen-Can, A., and Boyer, K. E. 2013. Unsupervised classification of student dialogue acts with query-likelihood clustering. In Proceedings of *Educational Data Mining* (Memphis, Tennessee, USA, July 6-9, 2013). 20-27

[22] Hsiao, I. H., and Awasthi, P. 2015. Topic facet modeling: semantic visual analytics for online discussion forums. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (Poughkeepsie, New York, US, March 16-20, 2015). ACM, New York, NY, 231-235.

[23] Gibson, A., and Kitto, K. 2015. Analysing reflective text for learning analytics: an approach using anomaly recontextualisation. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (Poughkeepsie, New York, US, March 16-20, 2015). ACM, New York, NY, 275-279.

[24] Southavilay, V., Yacef, K., Reimann, P., adn Calvo, R. A. 2013. Analysis of collaborative writing processes using revision maps and probabilistic topic models. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (Leuven, Belgium, April 8-12, 2013). ACM, New York, NY, 38-47.

[25] Kakkonen, T., Myller, N., and Sutinen, E. 2006. Applying latent dirichlet allocation to automatic essay grading. In *Proceedings of the 5th International Conference on Natural Language Processing* (Turku, Finland, August 23-25, 2006), Berlin/Heidelberg: Springer, 110–120.

[26] Sekiya, T., Matsuda, Y., and Yamaguchi, K. 2015. Curriculum analysis of CS departments based on CS2013 by simplified, supervised LDA. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (Poughkeepsie, New York, US, March 16-20, 2015). ACM, New York, NY, 330-339.

[27] Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., and Hatala, M. 2015. What public media reveals about MOOCs: A systematic analysis of news reports. *British Journal of Educational Technology*. 46, 3(2015), 510-527.

[28] Chen, B., Chen, X., & Xing, W. 2015. Twitter archeology of learning analytics and knowledge conferences. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (Poughkeepsie, New York, US, March 16-20, 2015). ACM, New York, NY, 340-349.

[29] Kandula, S., Curtis, D., Hill, B., & Zeng-Treitler, Q. 2011. Use of topic modeling for recommending relevant education material to diabetic patients. In *AMIA Annual Symposium Proceedings* (Washington, DC, US, October 22-26, 2011). American Medical Informatics Association, 674–682.

[30] Law, M. H., & Jain, A. K. 2003. Cluster validity by bootstrapping partitions. Technical Report MSU-CSE-03-5, Department of Computer Science and Engineering, Michigan State University, Michigan, USA 2003

[31] Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., and Blei. D. 2009. Reading tea leaves: How humans interpret topic models. In Advances in Neural Information Processing Systems 22, Y. D. Bengio, J. L. Schuurmans, C. K. I. Williams, and A. Culotta, Eds. MIT Press, 288–296.

[32] Greene, D., O'Callaghan, D., & Cunningham, P. 2014. How many topics? Stability analysis for topic models. In *Proceeding of Machine Learning and Knowledge Discovery in Databases* - European Conference, ECML PKDD '14 (Nancy, France, September 15-19, 2014). Springer Berlin Heidelberg, 498-513.

[33] Lange, T., Roth, V., Braun, M. L., & Buhmann, J. M. 2004. Stability-based validation of clustering solutions. *Neural Computation*. 16, 6(2004), 1299-1323.

[34] Thelwall, M., & Buckley, K. 2013. Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the American Society for Information Science and Technology*. 64, 8(2013), 1608–1617.

[35] Ortigosa, A., Martín, J. M., & Carro, R. M. 2014. Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*. 31 (2014), 527-541.

[36] Sundar, P. P. 2013. A Comparative Study for Predicting Student's Academic Performance Using Bayesian Network Classifiers. *IOSR Journal of Engineering (IOSRJEN) e-ISSN*, (2013), 2250-3021