Murthy, S. et al. (Eds.) (2018). Doctoral Student Consortium Proceedings. 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education.

Data-Driven Teaching Assessment in Inquiry-Based Learning by Topic Modeling

Hiroyuki Kuromiya[a*], Ichiro Hidaka[a] & Yoshiharu Yamamoto[a]
[a] *Graduate School of Education, University of Tokyo, Japan*
*kuromiya@p.u-tokyo.ac.jp

Abstract: Inquiry-based learning (IBL), wherein learning is driven by a student's inquiry, becomes popular in high school in many developed countries. IBL environments are also considered as a promising field of learning analytics or educational data mining. Until today, many researchers tried various machine learning methods to the process of IBL. However, student IBL outcomes written as a text have rarely been analyzed quantitatively. This paper aims to quantitatively assess the teaching of IBL via unsupervised machine learning method with natural language processing. Here, we propose a novel method for teaching assessment with topic modeling. Since educational assessment needs two kinds of information, i.e., what teachers want students to learn and how students change, we also plan to calculate the correlation between the topic and teacher's rating and the correlation between the topic and the posted year of the documents. In a preliminary experiment, we collected students' graduate theses over 31 years in a high school (N=3,328), we confirmed the tendency that a topic evaluated by teachers was also a topic which is popular among students. We think that the method of using the topic model applied to a large collection of output texts would be a good way to assess the teaching of IBL.

Keywords: Educational Assessment, Inquiry-based Learning, Topic Model, Educational Data Mining, Natural Language Processing

1. Introduction

Inquiry-based learning (IBL) is an environment wherein learning is driven by a process of inquiry owned by the student (CEEBL, 2018). Since the Programme for International Students Assessment (PISA) includes problem-based tasks, inquiry activities and related problem-based tasks belong to a main educational aim in all developed countries (Dostál, 2015).

IBL environments are also considered as a promising field of learning analytics or educational data mining. Until today, some researchers tried various machine learning methods to the process of IBL (Vahdat, 2017). However, text-style outcomes such as reflection have rarely been analyzed quantitatively on IBL because the data are unstructured and not quantitative style.

1.1 Motivation

Even though text as IBL outcomes have complex and qualitative form, we expected that recent machine learning techniques, especially unsupervised learnings could automatically summarize the data and provide useful feedback to teachers.

1.2 Literature Review

Beyond the context of IBL, the research applying natural language processing to student's learning logs are increasing. Taniguchi et al. (2017) and Nwanganga et al. (2015) analyzed student's ePortfolios and applied text-mining method. They counted the learning-related words or measuring emotions from the text. Other researchers analyzed the process of IBL in online course activities. Andrade et al. (2018) focused on talk and text in online inquiry-based learning and calculated similarity among them by using part-of-speech tags. He (2013) focused on the online questions and chat messages recorded by a live video streaming system. He applied clustering techniques to them and found major themes in student's question. With regards to offline IBL activities, Ito et al. (2016) focused on weekly reports of students and explored the association between learning-related words and student's self-evaluation score. However, there are few studies that used a stochastic language model such as topic modeling on the context of IBL.

2. Proposed research work

2.1 Preliminary Research Question

As a Ph.D. dissertation work, I will find novel and better way to utilize natural language processing techniques in IBL. In my first years, I want to focus on topic modeling and propose a new way to utilize it for student writings.

2.2 Expected Contribution

This study will expand the scope of learning analytics because currently qualitative data such as student reflection are not fully analyzed. Furthermore, this study will also contribute to evidence-based education. If teachers know learning log will return useful feedbacks, teachers have the incentive to accumulate data on their education.

3. Proposed Research Methodology

3.1 Tokenizing

We assume that we have a dataset like Table 1. Since Japanese has no separation symbol such as a "space" between words, sentences are to split into words by morphological analysis. At this point, high-frequency (greater than 3,000 times) or low-frequency (less than 10 times) words are excluded, and only nouns are extracted. Finally, we construct a word-document matrix for the next step.

Table 1: Example of the dataset

	Students' writing	**Year**	**Rating**
1	"I researched how computers work in our daily life. To investigate, I visited some websites …"	1999	3
2	"My motivation for research is that my father started having back pain last year. I started researching workers' health problems and these are the findings …"	2011	5
…	…	…	…
N	"Last year a law was published about people with disabilities. I will introduce it as easily as possible. First …"	2008	4

3.2 Extracting Topics from Student's Writing

To extract semantic structures from the text, we use Latent Dirichlet Allocation (LDA) (Blei et al., 2003), one of the most popular topic models in natural language processing. LDA is also known as a useful tool for educational data mining (Slater et al., 2017).

LDA assumes that each word is generated from categorical distribution (which means topic or topic-word distribution φ), and each document has a different topic mixture (which means document-topic distribution θ). LDA estimates θ and φ. By estimating φ the topic is determined, and by estimating θ the topics contained in the documents are revealed. After applying LDA, the dataset reveals data shown in Table 2. The number of topics K is arbitrary for the analysis.

Aside from the number of topics, LDA has two hyper-parameters α and β. α is a parameter of Dirichlet prior to the document–topic distribution, and β is a parameter of topic–word distribution. Following Steyvers and Griffiths (2007), we set $\alpha = 50/K$ and $\beta = 0.1$. The parameters in LDA are estimated by Gibbs' sampling with 2,000 iterations (Griffiths & Steyvers, 2004). We use the "topic models" package (Hornik & Grün, 2011) on R to implement the sampling. Each topic is named by considering top 10 relevant words in the relevant topic. Relevant words are ranked by equation (1). In the equation, w means a word, k means a particular topic, and p(w) means the generation probability of word w.

$$\log \; (\;,\;) = \frac{(\;|\;)}{(\;)} \qquad (1)$$

Table 2: Dataset converted by LDA

	Topic 1	Topic 2	…	Topic K-1	Topic K	Year	Rating
1	40%	30%	…	1%	0%	1999	3
2	1%	9%	…	4%	70%	2011	5
…	…	…	…	…	…	…	…
N	1%	5%	…	50%	2%	2008	4

3.3 Calculating Correlations and Drawing a Scatter Plot

Educational Assessment is used to decide what teachers want their students to learn and ensuring that students learn it (Suskie, 2018). To investigate which topic teachers evaluated and which topic had been popular across years among students, we calculate a correlation between topic allocation and ratings and a correlation between topic allocation and years. A scatter plot can be drawn for the topics using the correlation with years (how they were popular) as the abscissa and the correlation with ratings (how they were evaluated) as the ordinate. We can say that this plot represents the tendency of coincidence between teachers' intentions and students' learning.

3.4 Calculating Coincidence between Teachers and Students

Finally, we examine the correspondence between topics evaluated by teachers and popular topics among students. To confirm this, we calculate a correlation coefficient in the scatter plot drawn above.

4. Preliminary Experiment

4.1 Data

We used 3,328 student research abstracts with teachers' rating. The data were obtained in a secondary school in Tokyo. In the studied school, students investigated, as their theses work, questions they formulated themselves from their interests. This is identical to open IBL (Bansi and Bell, 2008) because the inquiries are not given but formulated by the students.

Data collection ranges from the year 1984 to 2014 and the rating ranges from 1 to 5. Total number of characters in the dataset is about 4 million containing 2,729 nouns. The mean length of the abstracts was 1,203 characters, and the standard deviation was 236. Most papers were written in Japanese, but three papers quoted English sentences. The English text was ignored due to language unification.

4.2 Results

LDA extracted 50 topics from the dataset. Due to limited space, some of the topics are listed in Table 3.

We calculated correlations of each topic with teachers' ratings and with years. The results are plotted in Figure 2. Each point represents a topic, some of which are plotted with the names. Further, we calculated the correlation coefficient from the scatter plot, obtaining r=0.56.

Table 3. Samples of extracted topics

Topic name	**Relevant words (original in Japanese)**
Crime	Incident, Crime, Youth, Society, Bullying, Police, Cause, …
Participation Activities	Activity, Participation, Volunteering, Group, High School, International, …
Investigation	Investigation, Inquiry, Results, Reference, Interview, Conclusion, …
Community	Region, Resident, Government, Sightseeing, Popularity, …
Information	Information, Usage, Internet, Spread, Need, Solution, A Lot, …
Graduation	Graduation, Impression, Teacher, Achievement, First, One, Last, …

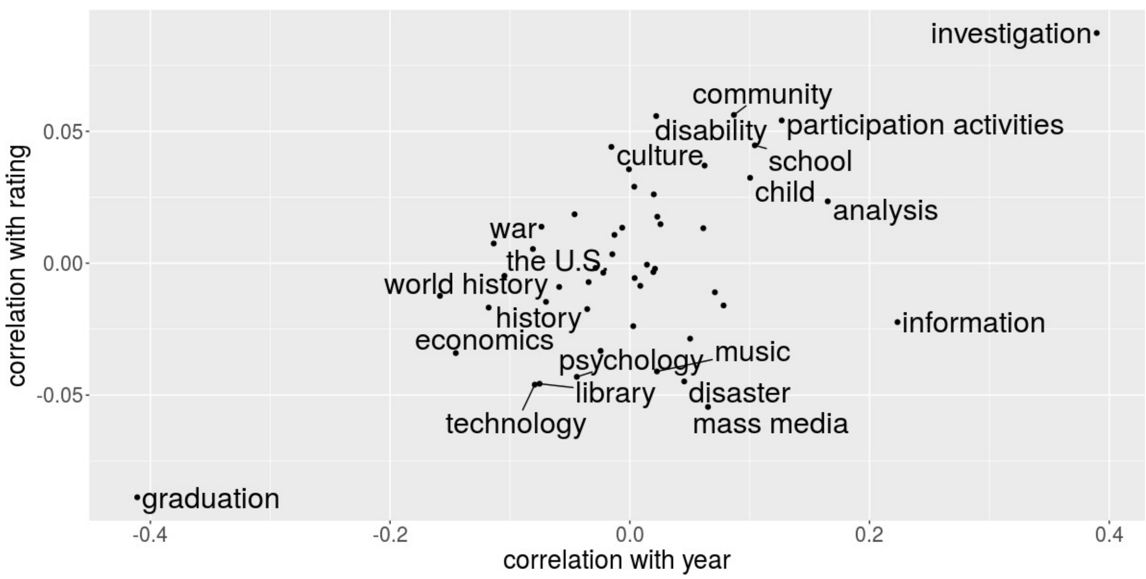

Figure 2. Correlation of topic allocation with year (abscissa) and with teacher's rating (ordinate) for each topic.

5. Conclusion

From the degree of coincidence value of r=.56, we confirmed that the topics evaluated by teachers and topics popular among students overlapped. This implies the validity of teaching of IBL can be quantitatively captured by LDA. It means that the method of using the topic model applied to a large collection of output texts would be a good way to assess the teaching of IBL. However, while the correlation of the correlations was relatively high (r > .5), the correlation between topics and years and the one between topics and ratings were not high (for most topics, |r| < .2 and |r| < .05, respectively). We should address the issue in the future.

References

Andrade, A., Georgen, C., & Stucker, M. (2017). Exploring a Text-Mining Approach as Rapid Prototyping Tool for Formative Assessments in Inquiry-Based Online Learning. Philadelphia, PA: International Society of the Learning Sciences..

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and children*, *46*(2), 26.

Centre for Excellence in Enquiry-Based Learning (2010, June 30) What is Inquiry Based Learning (EBL)? University of Manchester. Retrieved from http://www.ceebl.manchester.ac.uk/ebl/

Dostál, J., & Gregar, J. (2015). *Inquiry-based instruction: Concept, essence, importance and contribution*. Univerzita Palackého v Olomouci.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, *5*, 15991-16005.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National academy of Sciences, 101(suppl 1), 5228-5235.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. Computers in Human Behavior, 29(1), 90-102.

Hornik, K., & Grün, B. (2011). topicmodels: An R package for fitting topic models. Journal of Statistical Software, 40(13), 1-30.

Ito, K., Kizuka, A., & Oba, M. (2016, July). A Trial Utilization of Weekly Reports to Evaluate Learning for System Development PBLs. In 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI) (pp. 1064-1067). IEEE.

Nwanganga, F., Aguiar, E., Ambrose, G. A., Goodrich, V., & Chawla, N. V. (2015, March). Qualitatively exploring electronic portfolios: a text mining approach to measuring student emotion as an early warning indicator. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (pp. 422-423). ACM.

Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, *42*(1), 85-106.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, *427*(7), 424-440.

Suskie, L. (2018). Assessing student learning: A common sense guide. John Wiley & Sons.

Taniguchi, Y., Suehiro, D., Shimada, A., & Ogata, H. (2017, July). Revealing Hidden Impression Topics in Students' Journals Based on Nonnegative Matrix Factorization. In Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on (pp. 298-300). IEEE.

Vahdat, M. M. (2017). *Learning analytics and educational data mining for inquiry-based learning* (Doctoral dissertation, Technische Universiteit Eindhoven).