SIGGRAPH Asia 2016



Visualizing the Asynchronous Discussion Forum Data with Topic Detection By Simon Y. K. Li, Gary K. W. Wong





Document version 1.0

Agenda

- 1. Project Background
- 2. Problem of Analyzing Discussion Forum
- 3. Prior Works for New Experiment Design
- 4. New Experiment Design and Results
 - 4.1 Improved model of text mining
 - 4.1.1 Data Visualization Forum Graph
 - 4.1.2 Data Visualization LDAvis
- 5. Discussion and Further Development
- 6. Conclusion







Project Background

Watch BBC Documentary

Film

For private study or research only.

The Education University of Hong Kong Library

Not for publication or further reproduction.

- In this project, a total 24 undergraduate students in the HKIEd from the General Education course called *"Technology, Entertainment and Mathematics"* have been sampled for this preliminary experiment.
- One of the course requirements was to complete a *reflective posting* on an online discussion forum in Schoology.
- They were asked to watch a BBC documentary film called "*Beautiful Equations*" and other *selected movies*.
- Afterwards they posted their reflections in the forum. Each student was also required to <u>comment on three self-selected peers</u>, which were extracted in our experiment as text files for analysis

Reflection





Project Background (...continued)

- A software tools called "Polaris" from Ohsawa Laboratory was used for mining text from the following sources.
- Sources of Data (text format):
 - Online reflective discussion forum, etc.
- Performed analysis using KeyGraph to generate the visual patterns to identify:
 - The formulation of key concepts from black nodes and links
 - chances (red nodes and links) for the purposes of decision making and planning in the associated areas above.







Project Background (...continued)



Two conferences papers were published before with the results analyzed using

Keygnaphcation University of Hong Kong Library

For private study or research only. Not for publication or further reproduction.



2. Problem of Analyzing Discussion Forum

- Problem 1
 - Teachers usually want to know how their students perform or what the students are thinking
 - However, it is difficult and time-consuming to read all online discussion forum threads in details to comprehend the information inside manually



• Problem 2

Not for publication or further reproduction.





3. Prior Works for New Experiment Design

- 1. Why using social interaction analysis for our Moodle discussion forum?
 - Coffrin, Corrin et al. (2014) proposed visualization methods to <u>realize the student</u> <u>engagement and performance</u> in massive open online course (MOOC) environment.
- 2. Why using a probabilistic topic model with clustering visualization approach?



Image Source: HSE Science Olympiad

2011			Inspired by
Duval (2011) learning analytics could facilitate by <u>collecting</u> , <u>analyzing</u> , and <u>displaying the traces</u> that learners left behind to improve learning.	2012 One of the well- developed learning analytics systems is called Gradient's Learning Analytics System <u>(GLASS)</u> according to Leony, Pardo, et al. (2012).	2015 - 2016 •Ezen-Can, Boyer et al. (2015) used the ideas of <u>clustering</u> to group discussion topics. •Atapattu, Falkner et al. (2016) raised the ideas	ideas, we se probabilist together w clustering approach i to visualize performanc
The Education University of Hong Kong Library private study or research only.	This system captures and visualizes the events of learning with a <u>dashboard</u> serving as a presentation layer to display important	using <u>topic-wise</u> classification of discussion threads on MOOC.	teachers ca understanc performanc by using a v
for publication or further reproducti	analytics figures.		

nspired by all these deas, we selected using probabilistic topic model ogether with a clustering visualization approach in this project to visualize the student performance so that the eachers can better understand the performance-related data by using a visual mean.

4. New Experiment Design and Results

- In the latest experiment environment, we deployed a <u>Moodle environment to host</u> <u>discussion forum</u> of reflective postings from students.
- The contents of discussion forum of the general education course extracted from this Moodle environment for social interaction analysis which performed by a tool called <u>Forum Graph</u> (Chan, 2013)
- The data was then exported to a collection of programs written in R with packages implementing Latent Dirichlet Allocation (LDA) (Blei, 2012).





Image Source: Keep Calm-o-Matic



4.1 Improved model of text mining

- LDA was used as a text mining model in our latest experiment for topic discovery based on generative statistical/probabilistic model. It assumes that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics.
- LDA aims at uncovering the hidden thematic structure in a collection of some document to help identify interesting and useful patterns. A topic is a multinomial distribution over many different ranked keywords of the corpus of some document. The levels of details provided for analysis can be made deeper.
- This approach is better than just relying on keygraphs, by using clustering co-occurrence of keywords to determine concepts as the relevant, rather



rm topics.



4.1.1 Data Visualization – Forum Graph

 The social interaction inside one of the course discussion forums was visualized using "node" and "edges" in which "nodes" refer to the student while "edges" refer to interaction performance. The general rules to comprehend the Forum Graph are stated as follows.



Size of Node

Biggest size of a node identifies most responsive students.

Size and direction of Edge

Thick edge can be understood as a strong relationship between nodes. Also, the arrow of edge demonstrates which node are passively receiving some messages or who actively replying other's discussion.



4.1.1 Data Visualization – Forum Graph (Cont'd)

- Scenario 1: teachers want to identify how student perform in discussion and does the students act as a source or receiver in discussion
- The forum graph provides a "hover" function for users
- We can hover on one node and the graph indicates the response of the specific students
- It can illustrate how they think of their study, as the graph can identify the student's interaction with different discussion threads and investigate the performance group of students by using forum graph.



The Education University of Hong Kong Library

For private study or research only. Not for publication or further reproduction.





4.2 Data Visualization – LDAvis

- LDAvis is capable of providing the key term relevance in fixed size of topic models (TMs) in which LDAvis sufficiently visualizes the correlation of term among TMs and provides an interactive platform for users to select specific terms to reveal its related distribution of TMs
- In our experiment, the topic model parameter (k) has been initiated to 20 while setting up 5,000 iterations of (G) to execute the likelihood of MCMC (Markov Chain Monte Carlo) algorithm in LDAvis. The LDAvis graph, which contained the analysis results, was generated as following:



4.2 Data Visualization – LDAvis (Cont'd)

Basic Concepts of LDAvis:



4.2 Data Visualization – LDAvis (Cont'd)

- Scenario 2: teachers want to evaluate the relevant terms in Topic 2
- At the top left corner, we can select "Topic" for reviewing the corresponding topic in detail.
- Also, some specific topics were selected to review by pressing "Previous topic" or "Next topic" button.





4.2 Data Visualization – LDAvis (Cont'd)

- Scenario 3: teachers want to know the terms of "numbers" occurred in which topic.
- Hover on the "numbers" and the related topic groups will highlight as red circle
- "Topic 15" and "Topic 19" are not mutually exclusive but they are highly correlated, in which, the terms "numbers" is distributed in both "Topic 15" and "Topic 19". Besides, "Topic 19" performs as the subset of



ate study of research only

Not for publica Image Source: Deloitte University Press

Forpr

- We are going to invite	Continuous Enhancement		
serving teachers to test drive the methods we	- Another important	Further Development	
present in this paper by importing their students' works into the text mining and visualization to reveal	visualization can be network analysis using igraph and other R packages (Katya,	- LDAvis do not come with a meaningful name as a topic label.	
 any potential findings. Feedback would be collected to improve and 	2014). This especially further supplement the LDAvis.	 It is difficult to understand the topic the relevant keywords are grouped to. 	
further develop the methods based on field testing, which is the next few stages of our	- Using network analysis for the spotted topics periodically (e.g. weekly) may further help spot those	- Therefore, a taxonomy approach will be deployed to help resolve this particular problem so that	
experiment.	changes.	the relevant keywords can be further classified into a meaningful topic name	
		instead of just using topic numbers.	

香港教育大學 The Education University of Hong Kong

6. Conclusion

- By using the Forum Graph, academic administrators can perform social network analysis to understand the interaction among students and teachers to identify the frequent contributors and passive observers. The Forum Graph can probably help teachers better understand the participations of their students in the forum discussion.
- The LDAvis visualization tool also provides a huge potential to help teachers understand and research the existing and growing topics of discussion and probably discover accentual findings.
- LDAVis would help teachers spot whether their students can meet regular learning objectives and some unexpected learning outcomes by spotting some themes being closely located with themselves or being separately scattered as outliners.

These can be regarded as some major improvements to our experiment in learning analytics research.





Reference

- 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
- Andy, C. 2013, Reports: Forum Graph., Moodle. <u>https://moodle.org/plugins/report_forumgraph</u>
- Blei, D. M. 2012. Probabilistic Topic Models. Communications of the ACM 55(4):77–84
- 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
- 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.
- Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: a learning analytics approach. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (pp. 146-150).
- Katya, O., 2014. Network Analysis with R. <u>http://kateto.net/network-visualization</u>
- 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.
- Li, S. & Wong, G. 2016, Educational Data Mining using Chance Discovery from Discussion Board. In Proceedings of GCCCE'16, The Hong Kong University of Education (pp. 712-715).
- Ohsawa, Y., Benson, N. E., & Yachida, M., 1998. KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor. In Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on (pp. 12-18). IEEE.
- Sievert, C., and Shirley, K. E. 2014. LDAvis: A Method for Visualizing and Interpreting Topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces 63–70.
- Wong, G & Li, S. 2016, Academic Performance Prediction Using Chance Discovery from Online Discussion

Forums. In Proceedings of COMPSAC'16, IEEE (pp. 706-711)

of Hong Kong Library

End of Presentation – Thank you

