Visualizing semantic space of online discourse: The Knowledge Forum case

Bodong Chen
University of Toronto
252 Bloor St W, Suite 9-118, IKIT
Toronto, Canada
bodong.chen@utoronto.ca

ABSTRACT

This poster presents an early experimentation of applying topic modeling and visualization techniques to analyze online discourse. In particular, Latent Dirichlet Allocation was used to convert discourse into a high-dimensional semantic space. To explore meaningful visualizations of the space, Locally Linear Embedding was performed reducing it to two-dimensional. Further, Time Series Analysis was applied to track evolution of topics in the space. This work will lead to new analytic tools for collaborative learning.

Categories and Subject Descriptors

I.2.7 [**Natural Language Processing**]: Text analysis; K.3.1 [**Computer Uses in Education**]: Collaborative learning

General Terms

Algorithms, Experimentation, Measurement

Keywords

Discourse Analysis, Text Mining, Semantic Analysis, LDA, Knowledge Building, Collaborative Learning

1. INTRODUCTION

Online discussion platforms have been widely applied to support collaborative learning from K-12 to tertiary education. As the usage has grown, so has the need for effective tools to interpret and facilitate online discussion. In this area has emerged a rich body of leaning analytics research that turns online discourse data into insights about collaborative learning. A variety of tools have since been developed to tackle important aspects of online discourse, ranging from social interactions to temporal patterns [1, 3]. Among these studies, analysis of textual information has been a constant focus, as written language dominates communication in most online dialogues. Text mining—the process of extracting interesting patterns from text documents [9]—has been identified as one of the most widely applied techniques in educational data mining [5]. Combined with visualization techniques, patterns revealed by text mining could be used to re-present online discussion in meaningful ways to promote collaboration and reflection [4].

Research of Knowledge Building (KB), a social constructivist pedagogy engaging learners to collectively advance knowledge through communal discourse [8], has maintained a long-standing interest in analyzing online discussion. Knowledge Forum [7], the most widely used KB environment, is implemented with a suite of assessment tools to interpret and facilitate KB discourse [10]. Because of KB's emphasis on communal discourse, the evolution of public community knowledge has been a key concern of KB analytics. The major analytic questions in this area include: What is the state of community knowledge? Is the discourse effective in advancing knowledge? Where is the community discourse heading? The present study represents a recent experimentation to analytically tackle these questions. Applying topic modeling and visualization techniques, it attempts to explore meaningful ways to represent online discourse to promote reflection and "metadiscourse." Results will contribute to an ongoing Knowledge Forum rebuild with new means of representing knowledge to facilitate idea improvement.

2. SUMMARY OF PRELIMINARY RESULTS

Building on prior work, the present study aims to apply text mining and visualization techniques to achieve the following goals: (1) to model semantic space of KB discourse focusing on its emergent topics or discussion themes, (2) to visualize interpreted semantic space and inspect roles of different variables (e.g., time, students, and views) in KB discourse, and (3) to track student participation in different discussion themes over time.

To achieve these goals, I experimented with a range of techniques using a well-documented Knowledge Forum dataset. This dataset was produced by a Grade 4 class when they were studying a science unit "Light." This dataset contains 308 *notes* from six *views*,[1] with names such as "How Light Travels," "Colours of Light," and "Shadows."

To achieve the first goal of modeling semantic space, *Latent Dirichlet Allocation* (LDA), a probabilistic topic-modeling technique [2], was applied on this dataset. A list of topics were identified, turning discourse into a high-dimensional semantic space represented by topics. For each topic, its

[1] In Knowledge Forum, students post ideas to their community in the form of *notes*, which can be conceptually organized in a *view* to serve idea improvement.

LAK '14, Mar 24-28 2014, Indianapolis, IN, USA
ACM 978-1-4503-2664-3/14/03.
http://dx.doi.org/10.1145/2567574.2567595

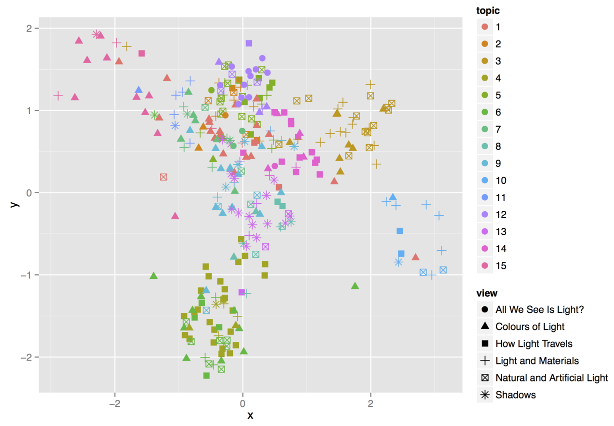

Figure 1: Visualizing notes by topics and views.

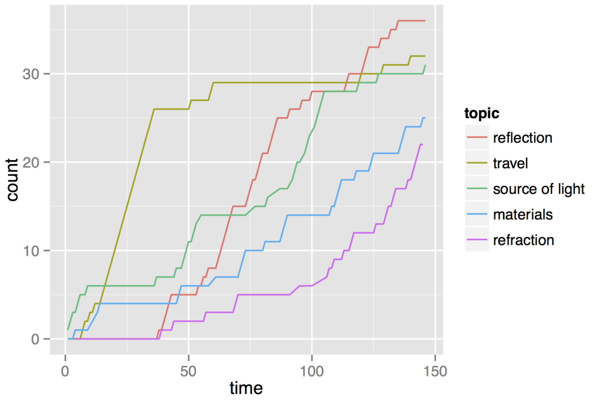

Figure 2: Evolution of the top five semantic topics.

probabilistic associations with *terms* are computed, making it possible for interpretation. For example, one topic was mostly associated with word stems—“mirror,” “look,” “angl,” and “bounc”—and so was clearly about “reflection.” At the same time, each *note*’s probabilistic memberships of topics were also calculated, so popularity of a topic could be approximated by the count of associated notes.

To visually represent topic modeling results, I further reduce the modeled high-dimensional space into two dimensions. *Locally Linear Embedding*, which is thought to be superior to multidimensional scaling [6], was applied to achieve this goal. *Notes* could then be visualized by different variables, such as topics, views, timestamps, and students, to explore usefulness of different visualizations. For instance, Figure 1 maps notes by topics and views. Given further assistance with interpretation, users could use this visualization to identify interested topics in each view and connect semantically related but physically isolated ideas across views.

Finally, I applied time series analysis to trace development of topics. By tracking participation in each topic across time, knowledge development of different topics becomes visible. Figure 2 presents the results of the top five topics. According to this figure, different topics had their “rises and falls.” For example, the beginning of the discourse mostly focused on “how light travels,” while “refraction” and “reflection” dominated the end. “Sources of light” and “light and materials” were discussed throughout the unit, while the other three topics had their own “golden times.”

In summary, the present study seeks to advance current learning analytics of online discourse by providing new means of interpreting and visualizing knowledge. It will inform development of new Knowledge Forum analytic tools emphasizing development of community knowledge. Broadly, it contribute to the important area of group-level, discourse-centric analytics focusing on the social aspect of learning.

3. REFERENCES

[1] A. Bakharia and S. Dawson. SNAPP. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK ’11*, page 168, New York, New York, USA, 2011. ACM Press.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, Mar. 2003.

[3] G. Dyke, R. Kumar, H. Ai, and C. Rosé. Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. In J. van Aalst, K. Thompson, M. J. Jacobson, and P. Reimann, editors, *The future of learning: Proceedings of the 10th international conference of the learning sciences (ICLS 2012) - Volume 1, Full Papers*, pages 363–370. ISLS, Sydney, Australia, 2012.

[4] J. Oshima, R. Oshima, and Y. Matsuzawa. Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, June 2012.

[5] C. Romero and S. Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135–146, July 2007.

[6] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science (New York, N.Y.)*, 290(5500):2323–6, Dec. 2000.

[7] M. Scardamalia. CSILE/Knowledge Forum. In A. Kovalchick and K. Dawson, editors, *Education and technology: An encyclopedia*, pages 183–192. ABC-CLIO, Santa Barbara, CA, 2004.

[8] M. Scardamalia and C. Bereiter. Knowledge building. In J. W. Guthrie, editor, *Encyclopedia of education*, volume 17, pages 1370–1373. Macmillan Reference, New York, NY, 2 edition, 2003.

[9] A.-H. Tan. Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, pages 65–70, 1999.

[10] C. Teplovs, Z. Donoahue, M. Scardamalia, and D. Philip. Tools for Concurrent, Embedded, and Transformative Assessment of Knowledge Building Processes and Progress. In *Proceedings of the 8th iternational conference on Computer supported collaborative learning*, pages 721–723, New Brunswick, New Jersey, USA, 2007. International Society of the Learning Sciences.