Ming, N.C., & Ming, V.L. (2013, June). Visualizing Topics, Time, and Grades in Online Class Discussions. Paper presented at the 10th International Conference on Computer Supported Collaborative Learning (CSCL). Madison, WI.

Visualizing Topics, Time, and Grades in Online Class Discussions

Norma C. Ming, Nexus Research & Policy Center, San Francisco CA 94105, Norma@NexusResearch.org;
Graduate School of Education, UC Berkeley, Berkeley CA 94720
Vivienne L. Ming, Socos LLC, Berkeley CA 94703, neuraltheory@socos.me;
Redwood Center for Theoretical Neuroscience, UC Berkeley, Berkeley CA 94720

Abstract: We present a series of visualizations of online discussions that combine topic modeling with other dimensions of the discussion contributions, to help faculty assess and improve learning from discussions. After applying probabilistic latent semantic analysis (pLSA) to calculate the relative conceptual distance between discussion posts, we projected posts or collections of posts into a two-dimensional space. By color-coding points according to their temporal position in the course or according to the author's final grade, we captured patterns in students' contributions that connect the topic modeling factors to more intuitively familiar characteristics. We consider how some possible qualitative features of the discussion may be represented in the topic space and outline future work to develop these tools further.

Introduction

As usage of class discussion forums has grown in both online and blended courses, so has the need for effective tools to monitor and interpret the activity in those forums. Just as they do in face-to-face environments, faculty must recognize teachable moments and facilitate effective interaction, but mediated by text-based, asynchronous communication. With this challenge also comes an opportunity: Using automated machine intelligence to mine the discussion record for key patterns may streamline the reading process, enabling faculty to focus on the most critical and valuable opportunities to intervene.

Existing applications of text mining to discussion forums have incorporated a variety of visualizations and features to guide users in navigating those discussions (*e.g.*, Awuor & Oboko, 2012; Kim, Shaw, Ravi, Tavano, Arromratana, & Sarda, 2008; Faridani, Bitton, Ryokai, & Goldberg, 2010). Yet much of this work focuses on the student or primary discussion participant as end user, rather than targeting the needs of an instructor seeking to facilitate a discussion to meet particular educational goals. In addition to obtaining a quick read on major themes and disagreements within a discussion, faculty need to assess students' understanding of key concepts, the quality of their participation, and their progress toward course goals. Addressing these disparate needs together in an integrated environment can help faculty keep students on track while also encouraging broader exploration. The work presented here offers a proof-of-concept using text mining to create visualizations linking formative and summative assessment to help faculty support productive online discussion.

Background

Text mining methods include numerous statistical techniques to identify patterns in a large body of text. Our focus here is on utilizing topic modeling to examine the semantic content of discussions, rather than incorporating syntactic, linguistic, stylistic, or sentiment analysis. Topic modeling analyzes a collection of documents to discover the topics discussed in those documents, as represented by a set of weighted terms (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). One type of topic model, probabilistic latent semantic analysis (pLSA), analyzes the probability of word co-occurrence in a given document, assuming Gaussian distributions of topics and word likelihoods (Hofmann, 1999). It treats each document as an unordered "bag of words" and infers a small set of latent factors which explain the distributions of words across documents. Each latent factor is a list of co-occurring words (a topic), and each document may be represented as a weighted combination of those factors (or topics).

For simplicity and proof of concept, we applied pLSA rather than one of the many more sophisticated topic models available (*e.g.*, latent Dirichlet allocation, LDA, and its variants). In previous related research, we used pLSA to create topic space visualizations depicting the relationship between students' posts and the instructor's posts, demonstrating the feasibility of the technique for revealing key patterns in their discussion interactions (Ming & Baumer, 2011). Other research successfully predicted students' course grades by applying pLSA and hierarchical LDA to their discussion posts (Ming & Ming, 2012). Here we connect discussion patterns with course grades to better illuminate important trends as the discussion unfolds, so that instructors may intervene to guide individual students or particular discussion threads. Future continuations of this work will explore how other models, algorithms, and visualizations may improve upon the results obtained here.

Methods

We examined student data from the discussion forum of an introductory, undergraduate-level biology course in the online degree-granting program at a large, for-profit university. Students were typically expected to post at least two substantive responses to each of two discussion questions per week, throughout all five weeks of the course. While individual instructors were granted some freedom in selecting the specific questions and assignments, each course instance (class) was required to adhere to a standard course outline and schedule with regard to learning goals, topics covered, and texts used. For this project, we restricted our analysis to students' contributions to the main discussion forum alone rather than including instructors' comments; conversations in individual and group discussion forums; or students' work on assignments, quizzes, projects, and tests. Later analyses incorporating these additional data from students, instructors, and other normative sources may further enrich our picture of student knowledge as evident from text.

Our analysis focused on 17 distinct classes taught by four instructors previously studied in other related research (Ming & Baumer, 2011). We normalized grades to be between [0,1] and removed data from students who dropped out before earning a final grade, to avoid confounding the quantity or assigned topics of their text with more subtle features associated with grades. This yielded a final dataset of 9118 posts by 230 students taking the same course within approximately a one-year interval. Posts were tokenized using a novel method called phrasal pursuit, which learns statistically meaningful phrases of arbitrary length. Phrases which occur regularly in the student posts and improve the pLSA model likelihood (described below) are incorporated into the bag of words/phrases representation. It uses model fitness rather than pairing likelihood as its selection criteria. For this method, we produced a dictionary of 5495 phrases.

Using pLSA, a probabilistic, generative extension of standard LSA, concepts/topics emerge as generative factors inferred from the documents by maximizing the data likelihood via gradient methods. A post with multiple topics is represented as the additive combination of multiple factors, and each factor is, in turn, a representation of a specific correlation pattern between the 5495 phrases in our dictionary. For our application of pLSA we assumed 100 topics/concepts from the discussion. After training the model to uncover the latent topics/concepts in the student posts, we used it to visualize the student work in 2-dimensional concept spaces. Posts or collections of posts were "projected" into 100-dimensional concept space by inferring the pLSA factors present in the writing (*i.e.*, computing the non-zero factor coefficients by gradient descent). For all of the projected documents we then used local linear embedding (LLE) to find a 2-dimensional representation which maximally preserves the spatial relationship between documents in 100-dimensional space.

Additional qualitative analyses drew from prior case studies characterizing the interaction patterns and discussion quality in selected threads reflecting a range of facilitation styles (Ming & Baumer, 2011).

Results and Discussion

The results in Figure 1 show that pLSA-based topic modeling may be used to capture some of the semantic differences in the discussion posts of students who receive different grades. In this graph, each point represents all of the comments by one student, color-coded by the student's final grade in the course. The aggregated posts from each student were first "projected" into the 100-dimensional pLSA concept space, and then LLE was used to further reduce the representation of the student's writing down to two dimensions. The horizontal gradient showing grades increasing from left to right suggests a machine-detectable difference in the topics they discuss. The vertical dimension reveals that students receiving C's and lower appear to neglect certain topics, represented below the dotted line. Closer examination of individual students' posts and of the topics and terms that correspond to these two axes will be invaluable for helping to interpret what these differences mean and how an instructor might potentially intervene to address them.

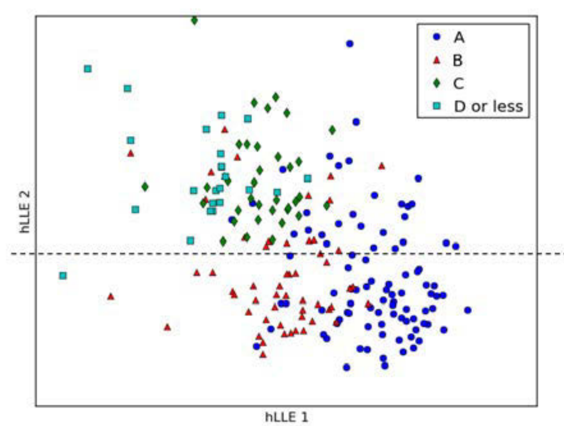

Figure 1. Topic space projection showing posting regions by individual students, color-coded by final grades. Each point corresponds to one student and represents all posts by that student in the main discussion forum.

Examining the discussion by individual posts rather than aggregated by student offers additional insight into some of the differences associated with course grades. As before, the axes in these figures were chosen for maximal separation rather than inherent meaning, so shorter distances between points reflect greater similarity between posts. Figure 2a shows that discussion posts by students earning grades of D or less are clustered in the center of the graph with little differentiation among them. As course grades increase, the associated comments travel farther away from the center, suggesting that these posts are exploring more specific topics. These results are consistent with our earlier finding, upon applying hierarchical LDA to data from a different course, that students earning higher grades discuss more specialized topics, while students earning lower grades discuss more general topics (Ming & Ming, 2012).

It is worth noting that this representation provides only a simple two-dimensional projection of the data. The results from the two figures combined indicate that there are multiple dimensions along which discussion comments and course grades covary, with comments moving toward the lower right corner in Figure 1 and diverging from the center in Figure 2a as course grades increase. The additional structure evident in Figure 2a further reinforces that there are specific directions in which higher-earning students' comments are moving, possibly corresponding to instructors' questions and comments or other aspects of the course structure. This suggests that students earning higher grades are not simply discussing topics with greater depth or specificity, but are addressing particular concepts that may be worthwhile.

One possible interpretation of Figure 2a's central cluster of posts by students earning low grades is that those students gradually stopped participating later in the course, and that region may contain the most basic concepts from the beginning of the course. However, coding the discussion comments according to course week reveals that the different weeks of the course correspond to the separate branches of the graph, as shown in Figure 2b. This could reflect the shift in topics as designed in the course outline, or the particular discussion questions that were asked. The different patterns produced by the two color coding schemes suggest that students earning lower grades tend to remain in the central, most general region even when the course topics invite more specific comments by students earning higher grades.

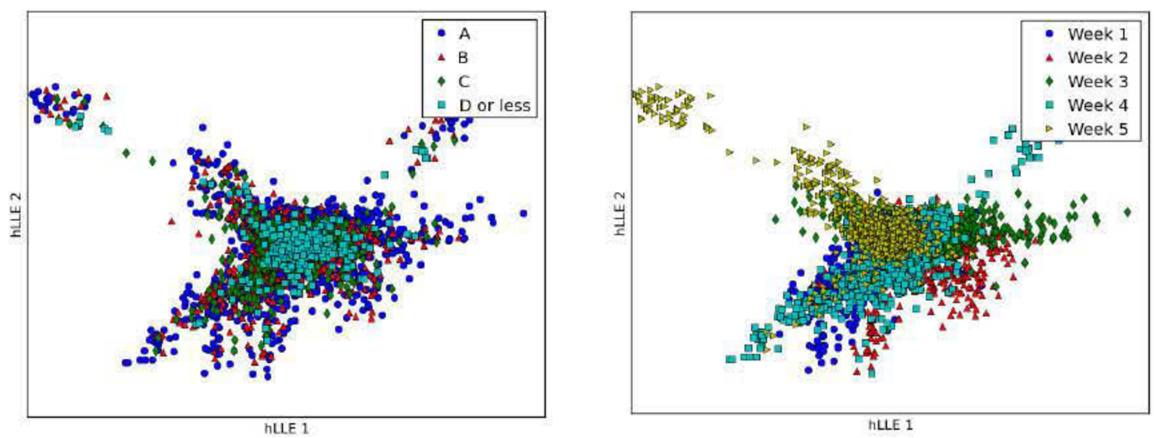

Figure 2. Topic space projection showing individual posts. The graph on the left (a) is color-coded by the final course grades earned by the post author; the graph on the right (b) is color-coded by course week.

Closer reading of individual comments and discussion threads suggests several worthwhile dimensions to examine for their potential correspondence with the axes shown in these topic space projections. While posts that are more distant from the center show more specificity and tend to come from students earning higher grades, those posts do not necessarily all contain comparable quality or educational value, regardless of the final grades earned by the students who authored them. Specificity is a correlate but not a guarantee of higher grades.

Greater specificity could potentially reflect a particular instantiation of a concept, a previously underemphasized step in a causal chain, a related fact, a personal anecdote, or a discussion of implications. One question described how disruptive evolution could lead to speciation and asked students to provide a new example of the same phenomenon. Here, conceptual distance may reflect the extent of overlap with the initial example and other students' examples in the type of organism, the feature mentioned (*e.g.*, animals' coloring), or the causal mechanism (camouflage and predation). Insofar as these dimensions vary in their importance, they also reveal that increased hierarchical depth may not always correspond to deeper understanding.

Likewise, despite their specificity, interesting related facts and personal anecdotes may be relevant or instructive in some cases but not in others. How closely they match other posts in the discussion offers no guarantee of their educational value, since they may spawn off-topic digressions or nuanced explorations of key concepts. Deciding when a conversation about identical twins shifts from pondering fundamental questions of nature-*vs.*-nurture to sympathizing about childrearing challenges demands deeper reading to identify distinguishing features.

Causal explanations and implications may be more likely to include topically relevant and illuminating discussion that advances student understanding. For example, articulating whether and how acquired traits might be inherited can prompt further exploration of how natural selection works. Similarly, identifying the energy source that drives capillary action in plants can help elucidate the underlying mechanism. Contemplating whether some populations might no longer be evolving or how hydrogen bonding affects water movement can also encourage deeper reflection on the biological and chemical processes involved. In these situations, addressing non-normative ideas may still be beneficial, by helping students to consider their ramifications and re-evaluate their own thinking.

These possibilities highlight the need for expert human judgment when intervening. Distinguishing among and mapping them to the dimensions on the topic space requires further work to link the qualitative analysis to the quantitative features from the hierarchical topic modeling. In particular, locating normative ideas within the topic space will provide valuable anchor points for tracking discussion trajectories. Continued analysis of how key concepts and misconceptions map on to the patterns evident here will further clarify the relationship between the topic modeling factors, students' grades, and their response to instructor intervention.

Conclusion and Implications

This work demonstrates the potential for applying topic modeling to generate insightful visualizations of student discussions that connect features of individual comments with later course performance. The selections shown here provide just a few early illustrations of how topic modeling can help to reveal the depth or sophistication of the concepts being discussed, as well as their correspondence to the desired topics and learning goals as designed in the course. At its most primitive, such work can analyze discussion post content to flag students who are likely to score poorly at the end of the course. More interesting applications can offer clues to content-specific reasons underlying those outcomes and suggest actions to influence learning trajectories. Creating a map of the topic space that includes normative sources (*e.g.*, the textbook, other required reading materials, instructor's lecture notes and comments) would help enable such identification. With such a guide, instructors then could quickly recognize when a particular student might be neglecting to consider some key concept, or when a discussion thread might be mired in a confusing misconception.

As a tool to help faculty monitor online discussions more effectively, such visualizations harness the power of machine intelligence to serve up potentially useful information for further evaluation and action by human intelligence. During their development they will need to be evaluated not just for their ability to capture important characteristics of student knowledge, but for their ability to convey useful and actionable information in an understandable manner. Future research will help determine how well tools based on these analytical techniques correspond to and augment instructors' professional knowledge.

References

Awuor, Y., & Oboko, R. (2012). Automatic assessment of online discussions using text mining. *International Journal of Machine Learning Applications, 1*(1). doi:10.4102/ijmla.v1i1.2

Donoho, D.L., & Grimes, C. (2003) Hessian eigenmaps: New locally-linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences, 100*(10), 5591-5596.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.

Faridani, S., Bitton, E., Ryokai, K., & Goldberg, K. (2010). Opinion Space: A scalable tool for browsing online comments. In *Proceedings of the ACM International Conference on Computer Human Interaction (CHI)*. Atlanta GA: ACM.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Berkeley, California: ACM, pp. 50-57. doi:10.1145/312624.312649

Kim, J., Shaw, E., Ravi, S., Tavano, E., Arromratana, A., & Sarda, P. (2008). Scaffolding of on-line discussions with past discussions: An analysis and pilot study of PedaBot. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems.*

Ming, N.C., & Baumer, E.P.S. (2011). Using text mining to characterize online discussion facilitation. *Journal of Asynchronous Learning Networks, 15*(2).

Ming, N.C., & Ming, V.L. (2012a, September). Automated predictive assessment from unstructured student writing. Paper presented at the First International Conference on Data Analytics, Barcelona, Spain.

Ming, V.L., & Ming, N.C. (2012b, December). Inferring conceptual knowledge from unstructured student writing. Paper presented at Neural Information Processing Systems, Lake Tahoe, Nevada.