

Using SMERT to Identify Actionable Topics in Student Feedback

Theodore T. Allen^a, Sara M Vinson^a, Alah Raqab^a, and Yosef Allam^b

^aThe Ohio State University, 1971 Neil Avenue, 210 Baker Systems, Columbus, Ohio 43210

^bEmbry-Riddle Aeronautical University 600 S. Clyde Morris Blvd. Daytona Beach FL 32114

Email: allen.515@osu.edu

Abstract

In this article, we explore the application to course evaluation analysis of two recently introduced text clustering methods: Latent Dirichlet Allocation (LDA) and Subject Matter Expert Refined Topic (SMERT). The clustering methods transform free-style or ordinary text into cluster or “topic” definitions and estimated proportions. For example, one result was that 12% of the student words were concerns about the content length for the online videos. We briefly review the clustering methods and then focus on a real world case study. The case study is associated with students’ feedbacks from a required for third-year Industrial Engineering and is used to illustrate the practical value of the proposed methods. The main result shows that the proposed method provides feedbacks that have the greatest cumulative effect which allows us to focus attention on a few important feedbacks rather than less important ones.

Keywords: Text analytics; topic models; course evaluation; quality control

Introduction

Education produces large quantities of “free style” text that is potentially relevant to system improvement including evaluations, essays, and reports. However, this information is often looked at only once and then never used again, especially when a professor/lecturer is looking through a large set of class evaluations. There is simply too much information available for the professor/lecturer to properly utilize.

A tool to look through and analyze this text would greatly increase the usability of this educational information. Text analytics is a field that uses various techniques to model large amounts of text. Much research has been done elsewhere in text analytics, such as the creation of a program that generates shortened version or “summaries” from research articles¹. This study, however, focuses on using a recently proposed innovative text processing and clustering method called subject matter expert refined topic (SMERT) models².

SMERT models are intended to permit the derivation of the main topics within all records or “documents” and the estimation of the fraction of the words in the overall “corpus” on each topic. There is also a more widely known approach with the same objective called Latent Dirichlet Allocation (LDA). The purpose of this article is to help clarify what these two specific text analytics clustering methods, LDA and SMERT, can offer to facilitate the interpretation of student written evaluations. Technically, SMERT models include LDA models as a special case and are designed to permit the user to “tune” the topic or cluster definitions using subject matter expert (SME) knowledge.

Latent Dirichlet Allocation (LDA) Description

Blei, Ng and Jordan³ proposed a clustering method called Latent Dirichlet Allocation (LDA). Their primary purpose was to create helpful cluster or “topic” descriptions and associate the derived topics with parts of text documents. The clustering method begins after a natural language processing has transformed the words into numbers, e.g., all instances of “computer” might be labeled “27”. Because documents can contain words on multiple topics and the topics are undefined, the associated clustering problem was more difficult than that attempted by many other clustering methods such as linear discriminant analysis and support vector machines. These other methods generally rely on pre-tagged documents and the assumption that each document is in a single cluster. Latent Dirichlet Allocation (LDA) was also “generative” meaning that the model can itself create new documents or collections of documents called corpora.

Therefore, LDA is a generative probabilistic model for collections of discrete data such as text corpora. Here, we briefly review the LDA models attempting to keep the description as concise and understandable as possible. LDA is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words.

In our notation, the ordinary or “low-level” data are the words or word-phrases in “document” or record. Specifically, $w_{d,n}$ is the n^{th} word in document d . The constant “ N_d ” is the number of words in that document. The multinomial or cluster or topic assignment for the word $w_{d,n}$ is $z_{d,n}$. The T -dimensional random vector θ_d represents the probabilities a randomly selected word in document d assigned to each T topics or clusters. The parameter “WC” represents the number words in the dictionary. The WC-dimensional random vector ϕ_t represents the probability that randomly selected words are assigned to each pixel in the topic indexed by $t = 1 \dots T$. The posterior mean of for each word is generally used to define the topics. The prior parameters and are scalars in that all documents and all pixels are initially treated equally.

With these parameter definitions, LDA is defined by the following joint posterior distribution function proportional condition:

$$P(w, z, \theta, \phi | \alpha, \beta) = P(\theta | \alpha) P(z | \theta) P(\phi | \beta) P(w | z, \phi) \\ \propto \prod_{d=1}^D Dir(\theta_d | \alpha) \times \prod_{d=1}^D \prod_{n=1}^{N_d} Mult(z_{d,n} | \theta_d) \times \prod_{t=1}^T Dir(\phi_t | \beta) \times \prod_{d=1}^D \prod_{n=1}^{N_d} Mult(w_{d,n} | \phi_{z_{d,n}})$$

where $Dir(\theta | \alpha) = \frac{1}{B(\alpha)} \prod_{t=1}^{|\alpha|} \theta^{\alpha_t - 1}$

and $Mult(x | \theta, n) = \frac{n!}{\prod_{k=1}^K x_k!} \prod_{k=1}^K \theta_k^{x_k} \rightarrow Mult(x | \theta, 1) = \prod_{k=1}^K \theta_k^{x_k}$.

Generally, LDA is the most cited method for clustering unstructured text perhaps because of its relative simplicity and because the resulting cluster or topic definitions are often interpretable⁴. Subject Matter Expert Refined Topic (SMERT) Description

SMERT models are used for probabilistic clustering of text records in one field databases. It is believed that topic models are popular partly because they are simpler and, therefore, more

observable than alternative models, which might include expert systems having thousands of ad hoc, case-specific rules³. Yet, the only ability to manipulate topics models comes through the prior parameters α and β . Specifically, shrinkage adjustments to α and β only control the degree of posterior uniformity in the document-topic and topic-word probabilities, respectively. We propose the subject matter expert refined topics (SMERT) model in Figure 1 to make LDA topic model more directable. The left-hand-side is identical to LDA with field-specific dictionaries with multinomial response data, w . The right-hand-side is new and begins with the arrow from \emptyset to x . This portion introduces binomially distributed response data, $x_{t,c}$ for $t = 1, \dots, T$ and $c = 1, \dots, WC$. Then $x_{t,c}$ represent the number of times for a given topic, t , word c is selected in $N_{t,c}$ in the model is controlled by the user and many combinations of topics t and words c can have $N_{t,c} = 0$. For more details about the interpreted equation between SMERT and LDA, see (Xiong and Allen)².

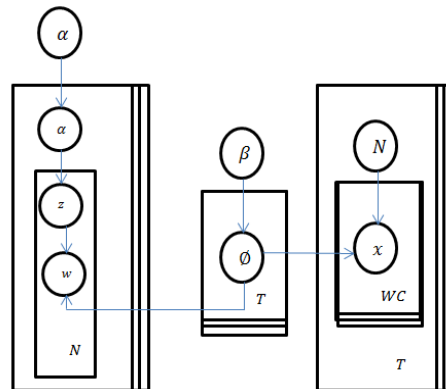


Figure 1: SMERT Model with LDA

Case Study Description

The case study was conducted in a class that was required for third-year Industrial Engineering undergraduate students. The sample included 33 students; 11 were distributed into the traditional laboratory setting, and the remaining 22 enrolled in the blended learning laboratory model. Both sections used the same required textbook, rubric, and syllabi while covering the same course material.

The Traditional Model

The traditional lab involved an instructor standing in front of students who had computers in front of them. The instructor gave a brief presentation, usually less than fifteen minutes, to introduce the laboratory objectives and the example problems covered. Then, the instructor led the students, who followed each command, to create example simulation models from the textbook.

The Blended Learning Model

The blended learning laboratory model involved three components:

1. Live instruction through the web. Students learned through virtual instruction, which allowed them to take the lab from a remote location or in a laboratory. The goals of this component were to orient the students to the ARENA software.

2. Self-paced instruction supported by PowerPoint and Camtasia. The project team utilized Carmen, OSU's learning management system, to post video-based lab materials and help monitor and assess a student's progress in the course. This learning management tool allowed the student the flexibility to master the course material at their own pace and increase the level of accountability since it is not always clear that every student is achieving the laboratory objectives in the traditional setting.
3. Classroom lecture time. The blended course maintained traditional face-to-face lecture. Students had the option to ask questions and to interact with the instructor and other students. This component integrates the lecture and the laboratory portions of the course, since students from both models attended.

A detailed description of the case study can be found in Allen, Artis, Afful-Dadzie, and Allam⁵.

Results

An important part of the output from both Latent Dirichlet Allocation (LDA) and subject matter expert refined topic (SMERT) models is the topic definitions. These are the cluster definitions. If they are interpretable, then the topic probabilities can be interpreted as the proportion of the words in the overall corpus on each interpretable topic. The table below shows the topic definitions after 10,000 collapsed Gibbs sampling iterations requiring approximately 0.5 minutes on an AMD Phenom II 2.6 GHz processor. The topic definitions are themselves the probabilities of all of the words in the dictionary relating to each topic. It is fairly common to define the topics by the words having the highest probability in each topic. The table below shows the top 10 words in all the topics from the LDA run for our student feedback case study.

Inspecting the Latent Dirichlet Allocation results, we construct definitions of the topics attempting to make them into interpretable sentences. This follows the approach in Allen and Xiong (2011). In the process of doing this, we find some words in each topic that do not fit in with the meaning. For example, in the first topic we imagine that students are complaining about the length of the online videos. If this is true, then it seems that the word "(Like Most) lab" would be on a different topic. Further, some of the related words would have been misallocated to that topic such that the true fraction of the words on this topic might be smaller.

Subject matter expert refined topic (SMERT) models are a recent invention designed to allow the user to edit the topic definitions so that they are relatively interpretable and accord with, hopefully, accurate information in the mind of the local subject matter expert (SME). In the teaching case study, the SME was one of the authors who was also a student in this class. Based on the SME's expert judgment, the initial LDA topics were interpreted in the table below. Also, the words to remove from each topic (if any) were provided. The words to be removed from each topic are called in Allen and Xiong (2013)² "zapping" words. The SME's zapping words are shown in the table. For example, the SME interpreted topic 1 to be complaining about the length of the videos. Therefore, the "outline" praising word in the "like_most" field was considered to be on a different topic.

Table 1. Latent Dirichlet Allocation topic generations given by the top 10 words in the dictionary for each topic ordered by estimated probability. For example, in cluster or topic #1, the most likely word is “(Mode) online” meaning that his topic related to the online mode of teaching.

Topic Description (Top Words)	Probability
(Mode) onlin (Video_Length) video (Video_Length) long (Like_Most) lab (Explain) lab (Like_Most) onlin (Audio_Feedback) sound (Improvement) video (Video_Ease) certain (Like_Least) lab	15.9%
(Lab_Changes) lab (Improvement) lab (Video_Feedback) video (Video_Ease) follow (Improvement) onlin (Like_Least) us (Narration_Content) time (Narration_Content) video (Explain) felt (Ability_Follow) video	12.7%
(Mode) onlin (Lab_Exam) b (Theory_Exam) b (Final_Grade) b (Project) b (Lab_Changes) modul (Lab_Changes) question (Lab_Changes) learn (Lab_Changes) lab (Lab_Changes) class	12.4%
(Final_Grade) b (Mode) tradit (Project) b (Theory_Exam) c (Lab_Exam) b (Like_Most) arena (Lab_Exam) c (Like_Most) learn (Lab_Changes) lab (Lab_Changes) class (Like_Least) long (Like_Least) time (Lab_Changes) arena (Lab_Changes) us (Like_Least) onlin (Explain) us (Like_Least) learn (Like_Least) took (Ability_Follow) follow (Visual_Cues) video	10.9%
(Like_Least) lab (Like_Least) instructor (Lab_Changes) homework (Lab_Changes) just (Lab_Changes) learn (Lab_Changes) class (Like_Least) didnt (Like_Least) class (Theory_Exam) c (Like_Most) learn	10.0%
(Lab_Changes) make (Lab_Changes) onlin (Like_Most) simul (Like_Most) interest (Like_Most) learn (Like_Least) complet (Explain) inform (Lab_Changes) student (Theory_Exam) b (Improvement_Details) uncheck	8.5%
(Theory_Exam) b (Video_Ease) video (Lab_Changes) lab (Lab_Changes) lectur (Lab_Changes) better (Video_Feedback) issu (Narration_Feedback) good (Mode) onlin (Video_Ease) follow (Lab_Changes) arena	8.5%
(Improvement_Details) lab (Lab_Changes) onlin (Improvement_Details) student (Explain) na (Audio_Feedback) good (Mode) tradit (Like_Most) complet (Ability_Follow) fine (Video_Detail) video (Video_Ease) work	7.2%
(Video_Detail) video (Video_Detail) long (Narration_Feedback) video (Visual_Cues) make (Video_Ease) video (Like_Most) lab (Like_Most) time (Audio_Feedback) good (Ability_Follow) video (Video_Feedback) issu	7.2%
(Ability_Follow) video (Video_Feedback) issu	6.7%

Rerunning for 10,000 iterations of SMERT required approximately 8.7 minutes on an AMD Phenom II 2.6 GHz processor. The results shifted the topic probabilities as indicated in the table. If the results are to be believed, only 11.2% of the words in the corpus were complaining about the labs being too long. The most common complaint was not about the content but instead about the pacing of the lab (topic 2 was 12.7%). We believe that this occurred because some praising words had, in the LDA model, been mixed in with the complaining words wrongly. By removing these words, the complaining percentage relating to the first topic is (likely) more accurate.

Table 2. The manually refined topic definitions and the subject matter expert (SME) generated “zapping words” which are to be removed from each topic definition based on SME judgment. At the right are the topic proportions from LDA and also from SMERT indicating a reallocation as the zapping words are effectively eliminated from the topics.

Topic	Description From LDA Results: Student...	Zapping Word	LDA	SMERT
1	Believed the online videos were too long content-wise.	like_most-online	15.9%	11.2%
2	Thought the videos were really slow to follow.		12.7%	12.7%
3	Felt the online modules needed better ties with in-class materials.		12.4%	10.7%
4	Liked ARENA and that they were learning a simulation software.		10.9%	11.4%
5	Student thought that ARENA takes too long by itself.		10.0%	8.7%
6	Did not like how the lab instructor taught.	lab_changes-homework	8.5%	6.5%
7	Believed all be online or the online should be more available.	like_most-interest	8.5%	10.7%
8	Thought that the lab lecture content could be improved.		7.2%	9.5%
9	Wanted to complete the modules whenever they wanted.		7.2%	10.2%
10	Thought videos were good or made miscellaneous comments.	visual_cues-make	6.7%	8.5%

Re-sorting the topics using the SME definitions, we derive the Pareto chart shown below. As the SMERT technology matures, creating such summaries will require less effort from SMEs and computing times. It seems possible that within the next two years, creating such Pareto charts summarizing feedback for a large number of university courses will become inexpensive in cost and time. As one of us was the instructor of record for the course in question, we can report that having numbers (while possibly inaccurate) creates a greater sense of urgency for improving the specific issues mentioned. The overall SEI scores for the related course were above average for our unit. Without having a summary focusing the negative aspects of the feedback, improvement of that course would likely not happen. Now, an improvement effort to address the feedback is being formulated.

Conclusions

Course evaluation provides tangible evidence that departments are putting resources into courses that benefit students. Also, evaluations help the direction of scarce resources to support courses that achieve objectives. In this paper, we proposed the applications of two recently proposed text clustering or analysis methods to aid in the interpretation of free style text student feedback. In our case study, the Latent Dirichlet Allocation method aided the derivation of reasonably understandable topics or clusters in the data. Yet, some of those clusters included words that did not make sense to our subject matter expert (SME). The subject matter expert refined topic (SME) used feedback from the SME in terms of words to be “zapped” or eliminated in the various topics. Putting in this “high-level” data, the resulting topic definitions made more sense to our SME and were more actionable. Also, the parts of the corpus of student feedback associated with the zapped words were reassigned to other topics shifting the topic proportions. This (hopefully) made these proportions more accurate because the model used SME knowledge.

The resulting application aided in the subjective evaluation of the scope of the problem. For example, we found that only 11.2% of the words in the corpus were complaining about the labs being too long. The most common complaint was not about the content but instead about the pacing of the lab (topic 2 was 12.7%). Finally by using the SME definitions, we derive the Pareto chart to identify those feedbacks that have the greatest cumulative effect, and thus screen out the less significant effect in an analysis. The Pareto chart allows us to focus attention on a few important feedbacks rather than less important feedbacks⁶. Based on Pareto chart ordering, we can focus the improvement activities in the appropriate specific direction.

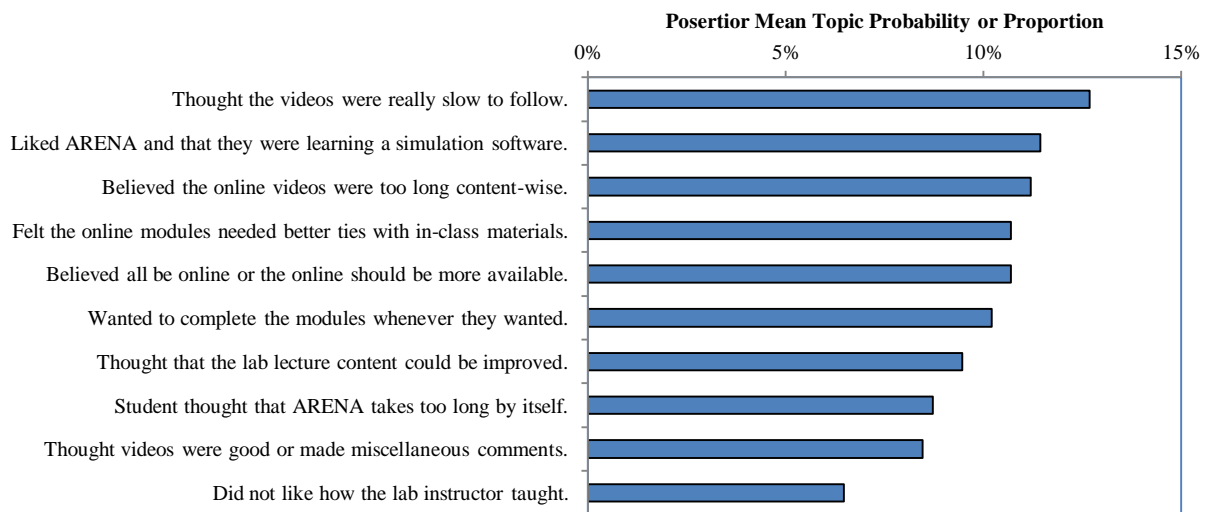


Figure 2. The final summary Pareto which was, without too much computer time or subject matter expert (SME) effort, generated to summarize the student comments.

References

- [1] C. Dunne, B. Shneiderman, R. Gove, J. Klavans, and B. Dorr, "Rapid understanding of scientific paper collections: integrating statistics, text analysis, and visualization," University of Maryland, Human-Computer Interaction Lab Tech Report HCIL-2011, 2011.
- [2] Xiong H, Allen T. Multi-Field Subject Matter Expert Refined Topic Models for Text Modeling, Integrated Systems Engineering Technical Report 2013.
- [3] Blei, DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3. 2003, 993-1022.
- [4] Griffiths TL, Steyvers M. Finding scientific topics. *Proceedings of the National Academy of Sciences*. 2004, 101 (suppl. 1), 5228-5235.
- [5] Allen, T. T., S. Artis, A. Afful-Dadzie, and Y. Allam. Case Study Application of Blended Learning for Delivering Discrete Event Simulation Laboratories.
- [6] Charantimath, P. M. (2009), Total Quality Management, 3rd edition, Pearson Education.