

# Latent Dirichlet Allocation for Textual Student Feedback Analysis

Swapna GOTTIPATI\*, Venky SHANKARARAMAN, Jeff LIN Rongsheng  
*School of Information Systems, Singapore Management University, Singapore*  
\*swapnag,venks,jefflin@smu.edu.sg

**Abstract:** Education institutions collect feedback from students upon course completion and analyse it to improve curriculum design, delivery methodology and students' learning experience. A large part of feedback comes in the form textual comments, which pose a challenge in quantifying and deriving insights. In this paper, we present a novel approach of the Latent Dirichlet Allocation (LDA) model to address this difficulty in handling textual student feedback. The analysis of quantitative part of student feedback provides general ratings and helps to identify aspects of the teaching that are successful and those that can improve. The reasons for the failure or success, however, can only be deduced by analysing the textual comments from the students. In order to fully decipher the qualitative, textual feedback effectively and efficiently, researchers have attempted text mining techniques, which use natural language processing and machine learning algorithms to parse the text and extract the relevant insights. Our solution, using LDA models to discover the aspects or topics of the comments. We then employ sentiment mining techniques to classify the comments as positive or negative. To assess its performance, we applied our solution model on the data collected from teaching evaluations of Singapore Management University. Our experiments and evaluations show that LDA models perform better than clustering models in finding aspects from students' comments. In addition, the sentiment mining results indicate that classification method performs better than lexicon models. Also described in paper is the technical architecture of the tool along with some visuals of the interactive dashboard.

**Keywords:** Teaching evaluations, Quantitative feedback analysis tool, Topic extraction, Sentiment Mining, Latent Dirichlet Models, Classification.

## 1. Introduction

Students and their learning experience form a key aspect of achieving and sustaining the quality of higher education. Their feedback on the curriculum and instruction is, therefore, essential for improving the teaching and learning process (Dietz-Uhler & Hurn, 2013). In order to collect feedback, most education institutes employ popular methods implemented such as online and in-class surveys (Donovan, Mader, & Shinsky, 2010). Online surveys have an added benefit of keeping the repository of the scores and comments from the students over multiple courses and over many years. Such repository enables us to track the quality of the course, and aids the instructor in assessing the impact of the changes he or she makes to the course. At the same time, it also helps management to make data-driven assessments of teaching evaluation scores for faculty appraisal and renewal activities (Leckey, & Neill, 2001, Venky et al. 2017).

Student feedback is not a one-time process but a continuous cycle with several stages: collection, analysis, action and decisions, and monitor (Lockyer & Dawson, 2011). Student feedback surveys typically come in two forms: numerical ratings, and textual comments (or quantitative and qualitative forms). For the numerical part, several education institutes incorporate a rating scale such as Likert scale for capturing the student feedback. Such numerical ratings, while excellent in spotting what went wrong or which components of the course had issues, fail miserably in getting to the causes behind the shortcomings. The reasons and remedial measures can only be gleaned from the textual comments from the students. Hence, the combining the quantitative and qualitative feedbacks is critical in the curriculum and instructor improvements.

Analyzing the quantitative feedback is easy thanks to the ready availability of statistical tools and methods to run on numerical data. However, they fall short when it comes to in-depth analysis of qualitative data (Scott, Grebennikov, & Shah. 2008). Consequently, the analysis of the qualitative feedback, which is textual in nature, is a painstaking process. For instance, in a course with a large number of students, the faculty will not be able to manually analyze and derive insights from the qualitative course evaluations. On a macro level, analyzing the qualitative feedback from all the courses in a curriculum will be an impossibly tedious enterprise for the curriculum designers and managers to undertake. Yet, such an analysis is a necessary driver for effective decisions.

The quantitative feedback questions on teaching effectiveness are about preparation, clarity, encouragement, stimulation of interest, availability presentation skills, enthusiasm, fairness and concern for students (Nitin et al. 2015). The feedback questions on course experience are about learning experience, clarity of objectives, quality of feedback, value, skills, usefulness of projects/cases/assignments, and interactivity. The two open-ended qualitative feedback questions are worded as follows:

- Please give responsible feedback regarding what you liked/disliked about the instructor.
- Please give responsible feedback regarding what you liked/disliked about the course.
- The freeform textual answers from these students to these two questions typically consist of their sentiments about the learning and teaching aspects of the course and the instructor. The goal of the tool described in this paper is to extract the topics and sentiments based on the textual comments.

Table 1 shows some sample comments and the corresponding topics and sentiment. The first comment has two topics, “Faculty engagement” and “Faculty enthusiasm”. At the same time, the comment can be categorized as a positive sentiment. The third sentiment is about “Faculty fairness” and “Course content usefulness”.

Table 1. Sample feedback with topics and sentiment

<b>Feedback</b>	<b>Topics</b>	<b>Sentiment</b>
She is always willing to guide and is always passionate about it.	Faculty engagement, Faculty enthusiasm	Positive
The course is refreshing to the mind, freeing from the textbooks and notes.	Course learning experience	Positive
The prof should not focus too much attention on class participation as it only comprises of 10%.	Course delivery	Negative

The consolidated analysis of the comments is presented as a user-friendly visualization, enabling the faculty to absorb the information in short span of time and make informed decisions on the course delivery. In the case of the third comment, since the sentiment is negative, the faculty may take action on the “Class participation” component of the course to address the student’s concern.

Analyzing and consolidating hundreds of students’ comments manually is time-consuming and infeasible task. Therefore, it is essential to extract the knowledge (as shown in Table 1) and to present it in user-friendly visuals, as implemented in our feedback analysis tool. The insights from student feedback analysis play a major role, not only improving the curriculum model and course content, but also in discovering the gaps in current evaluation model. These insights can enable the faculty or the curriculum managers to make informed decisions. We applied the text analysis tool on feedback collected on 183 courses. The evaluations of tool show that LDA models and Textblob sentiment classification tools perform very well in extracting knowledge from the comments.

In section 2 of this paper, we present the background of LDA models and sentiment mining. We then present a literature survey on educational data mining in Section 3. Subsequently, in section 4, we describe our solution model and the tool. Finally, we describe the datasets, experiments and findings from key stages in the solution model in Section 5 and conclude with the future directions of this work in Section 6.

## 2. LDA and Sentiment Mining

*Latent Dirichlet Allocation:* LDA, also called as topic model, is an unsupervised probabilistic generative model. Given a set of documents and a number of topics as input, LDA automatically returns a relevant set of words probabilistically associated for each topic (Blei, Ng and Jordan M. I. 2003). LDA is based on the intuition that each document contains words from multiple topics; the proportion of each topic in each document is different, but the topics themselves are the same for all documents.

One of the advantages of topic models is that they are unsupervised algorithms, which removes the requirement to manually annotate the corpus, thereby reducing cost, and improving the objectivity of the analysis results. LDA performs well mainly due to its better handling of synonymy and polysemy because of the probabilistic association of words to topic. Therefore, in our study, we use LDA in a semi-supervised approach to discover and trace major topics from student comments. In order to handle large text size of aspect extraction, topic model based summarization models are the most user friendly visualizations of the large corpus.

*Sentiment Mining:* Sentiment mining aims at classifying the comments into positive or negative polarities using supervised or unsupervised methods (Pang, Lee, and Vaithyanathan. 2002). Sentiments are usually targeted on certain aspect of the context. For example, “workload too heavy and deadline between assignment and project are too near” is a comment with a negative sentiment towards the topic, “assignments”. The task of sentiment classification use lexicon methods that are based on sentiment words and phrases which are instrumental to sentiment analysis or classification techniques like SVM models (Bing, 2010). A list of such words and phrases is called a sentiment lexicon. The task of sentiment target detection aims at extracting the sentiment targets in the reviews using multiple heuristic techniques. Extracting suggestions is a sub task of sentiment analysis (Gottipati et al, 2018) which can aid the faculty to extract the suggestions for course improvements from teaching evaluations. In our solution approach, we use LDA models to extract the topics and assign the comments to the relevant topics.

## 3. Related Work

In this section, we focus on works related to LDA models and sentiment mining relevant to student feedback analysis.

*LDA in education data mining:* Early adoptions of LDA for educational data include the work of Haruechaiyasak and Damrongrat (2008), who recommended articles from Wikipedia by calculating the similarity measures among topic distributions of the articles. Wei et al. (2011) provided resource recommendation for users in the e-learning system based on contents and user log activities applying LDA models.

Ming and Ming (2012) applied hierarchical LDA to predict the grades of the students and showed that these analyses provide information that aids in student assessments. Zhang et al. (2012) applied LDA to online discussions of four Chinese classrooms to extract topics and display the temporal profiles of the topics. Sherin (2012) used LDA to extract fragments (categories) of ideas from student interviews. Southavilay et al. (2013) used LDA models to mine cloud data from Google Docs to gain insights on how learners' collaborative activities, ideas and concepts are developed during the process of writing. The above studies point to the promising potential of LDA to capture conceptual topics in education datasets.

*Sentiment analysis in in education data mining:* Early works of sentiment analysis of student feedback using data mining approach include Altrabsheh et al. (2014) who devised a system to analyze sentiments in real time to provide real-time intervention in the classroom. They combined Support Vector Machines and Naïve Bayes. Rashid et al. used generalized sequential pattern mining and association rule mining to analyze opinion words from student feedback (Rashid et al. 2013). Nitin et al. (2015) combined clustering and sentiment classification models to extract the topics and the sentiments from the student's feedback. The limitation with clustering model is that each comment is only assigned to single cluster and the topics coherence quality is low.

In our solution, we propose LDA models to address this limitation and the comments will be assigned to more than one cluster. At the same time, we also use classification based approach for the sentiment mining task. We explain the details of our solution model in the next section.

## 4. Solution Model

### 4.1 Solution Design

In this section, we describe the concept solution model of our tool and then the details of the tool implementation. Figure 1 shows the concept solution model of the textual feedback analysis tool.

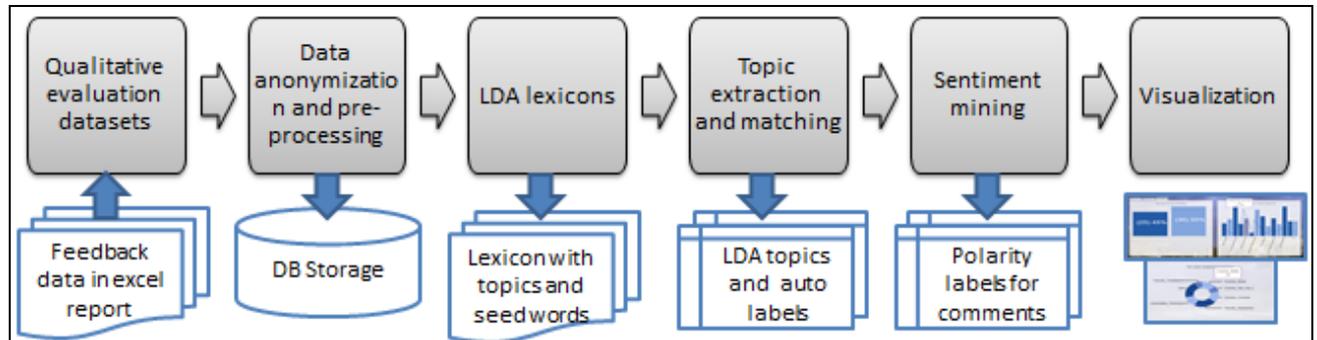


Figure 1. Concept solution model for LDA based textual feedback analysis tool

*Qualitative evaluation datasets:* The input for our solution approach is the quantitative data from all courses in the University. In this stage, we designed and implemented the database for the tool.

*Data anonymization and pre-processing:* The data consists of faculty names, course names and course ID's which are sensitive and should be maintained confidential. Hence the faculty names and course names are anonymized. The main challenge we faced in this dataset is the faculty names with in the comments.

*Data Anonymization:* For example, “Prof John is very enthusiastic about the course”. To handle the anonymization in the textual comments, we combined dictionary and Named Entity Recognition techniques (Bing 2010). NER tools are capable of extracting the person names from the text. The person names are then replaced by “xxxx”. Once the data has been anonymized, we moved to the next stage of lexicons for LDA models. The cleaned data is stored in the database.

Sentence tokenizing aids in segmenting the comment as each comment may have multiple sentences. Students' usually comment on more than one topic and comments may contain multiple sentences. For example, “Prof works hard for the class. She ensures that we learn from each class”, has two topics, “Faculty preparation” and “Faculty concern”. Therefore, to capture multiple topics in the textual feedback, all the comments are tokenized into sentences. Further, stopwords are removed and documents are converted into to word matrix as a part of pre-processing process (Bing 2010).

*LDA lexicon:* LDA lexicons are used for semi-supervision of topic extraction and auto labeling (Ma. et al. 2013). In this stage, we generated a lexicon for nine topics such as instructor engagement, instructor preparation, instructor feedback, course value, course assignments, course skills, etc. These topics serve two purposes; generate better quality of clusters of comments and determine the optimum number of the clusters for each course.

*Topic extraction and lexicon matching:* In this stage, we first run the LDA model with number of topics set to fifteen. We then use similarity matching algorithm to map the randomly generated topics to the LDA lexicon to finalize the number of clusters and cluster the comments. At this stage, the tool generates the probability distribution of topics for each comment and we set the threshold to extract the top topics for each comment.

*Sentiment mining:* In this stage, the comments are subjected to a sentiment polarity computing algorithm. Based on the polarity score, the tokenized sentences will be classified as either positive or negative sentiment (Bing 2010). We tested this stage using two popular techniques; Polarity analyzer (Polarity Analyser) and Textblob (Loria, 2014). Polarity analyzer comes with words list that are pre-labelled with sentiments. The basic approach is that we first split the sentences into words and then look up each word and its antonyms (using wordnet.synsets, Christiane, 1998)

in the labelled words list.

Text blob algorithm makes use of Naïve Bayes classifier trained on the movie reviews dataset. Each word is scored and the cumulative score of the sentence will indicate whether the sentiment is positive (if greater than or equal to zero) or negative (if less than zero). In our preliminary experiments, we observed that the basic Textblob tool has limitations on the contrasting conjunctions and suggestive words (Ramanand, et al. 2010). Therefore, we further modified Textblob to consider contrasting conjunctions such as “although”, “despite” etc., and suggestive words such as “could”, “hope” etc.

*Visualizations:* In this stage, the results from LDA and sentiment mining stages are combined to generate interactive dashboard. The consolidated results are displayed using graphs and tables. These graphs are interactive in nature so that the faculty can do deeper analysis on the topic or sentiment in each cluster of comments.

#### 4.2 Tool Description

Our tool is built using python base web framework, Django (Carl, 2010). Figure 2 shows the high level architecture of our tool. Our tool is a web application that supports scalability, multiple user access, database features and authentication protocol.

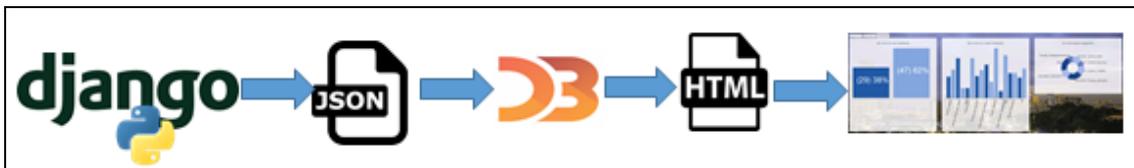


Figure 2: Overview of the software architecture

We setup Django with user authentication and it conveniently comes with administrator access. Python is majorly used for developing sentiment, opinion and topic modelling algorithms. JSON and D3 are used for visualization task which enable a better performance and various interactive charts. D3 is Javascript based library (Michael et al. 2010) that creates an interactive chart or graph from our JSON structure data. The script is incorporated into html content and interpreted by the web browsers.

## 5. Experiments and Results

### 5.1 Datasets

Table 2 shows the data statistics of the comments collected from our teaching evaluation tool. The data is collected for over four years and for all three academic terms. The data is suitable not only for individual course analysis but also the comparison analysis.

Table 2: Data Statistics

Data	Statistics
# Courses	183
# Faculty	334
# Comments	153302
# Years	4 years, 12 terms (2013-2017)

### 5.2 Findings

In this section, we present the evaluation of the topics extractions and sentiment mining tasks. We also present sample visualizations of the tool.

We manually grouped sixteen quantitative feedback questions into nine prevalent topics or aspects. This is to generate more meaningful topics for the qualitative comments. The LDA lexicon

topics are Faculty Interaction, Faculty Engagement, Faculty Feedback and Approachableness, Faculty Fairness and Preparation, Faculty Presentation, Course Content, Course Skills, Course Value Usefulness and Challenges, and Course Projects and Assignments.

### 5.2.1 Topic Evaluations

The LDA model topics are mapped with the LDA lexicon and those groups with matching scores below the threshold are labelled as “Unclear” category. We evaluated the LDA model and simple Cosine (Bing. 2010) based clusters for the topic extraction from the comments. Table 3 shows the results of LDA models and Cosine based clustering of comments.

Table 3: Topic extraction: Sample LDA Vs Cosine based clusters from the dataset

<b>Feedback</b>	<b>LDA</b>	<b>Cosine Similarity clusters</b>
The course is somewhat useful as communication is an extremely important skill.	Course Skills, Course Values/ Use/ Challenges	Course Skills
very approachable and helpful. good materials and practices. The prof is very good explanation about the course and strong interest in the course The xxxx has really good knowledge about the course.	Course Project Assignment Cases, Course Values/ Use/ Challenges, Faculty Feedback, Faculty Interaction/ Engagement	Course Values/ Use/ Challenges

From table 3, we observe that the LDA models are capable of providing multiple topics and more relevant topics for the comments whereas cosine clusters assign single topic to the comment. The second comment is not tokenized into multiple sentences by the tool as the “.” is not grammatically correct. However, the results show that LDA models are useful for labelling comments to multiple topics where the clustering approach will be able to label to one topic.

### 5.2.2 Sentiment Classification Evaluations

Sentiment classification results are evaluated using standard data mining measures; precision, recall, accuracy and F-Score. We manually annotated 434 sentences from the comments and tested on all the three different tools described in Section 4. Table 4 shows the sentiment classification tool performance.

Table 4: Sentiment Classification Results

<b>Sentiment Tool</b>	<b>Recall</b>	<b>Precision</b>	<b>Accuracy</b>	<b>F-Score</b>
Polarity Analyser	93.10%	66.39%	67.51%	77.51%
Textblob	96.17%	67.47%	69.82%	79.30%
Textblob-Improved	90.80%	92.58%	90.09%	91.68%

From Table 4, we observe that performance of standard Polarity Analyser and Textblob is very close where Textblob has slightly better performance than Polarity Analyser, 1.8% higher F-Score. Compared to both the tools, improved-Textblob has better performance which handles both suggestive words and conjunction words. F-Score is 91.68% which is 12.38% higher than the standard Textblob tool.

### 5.2.3 Discussions

We further performed gap analysis to analyze the false positives and false negatives in the results. Table 5 shows sample comments and, the actual and predicted sentiment labels.

Table 5. Sample comments and sentiment classification by improved Textblob tool

<b>Comment</b>	<b>Actual</b>	<b>Predicted</b>
more support and help is needed from instructors and professors	Negative	Negative
although the course is a bit dry i see the value in it in the technology industry	Positive	Positive
course exposed me to a lot of new concepts which are very interesting and challenging	Positive	Positive
the instructor was monotonous uninspiring and seemed unsure about certain topics	<b>Negative</b>	<b>Positive</b>
he is very scarstic	<b>Negative</b>	<b>Positive</b>
it is definitely a course all SIS students should go through to understand the basic about real world computing issues	<b>Positive</b>	<b>Negative</b>

Table 5 shows the incorrect classifications in bolded font. Our analysis shows that some of the possible reasons for the incorrect predictions are missing sentiment words in the training dataset (uninspiring, unsure etc.) and spelling errors (“scarstic”). The other possible reason is that students express comments in the form of suggestions as shown in the fifth example comment, “it is definitely a course all SIS students should go.”. To improve the quality of the tool, it is important to identify the suggestions and segregate them as the suggestions and not as sentiments.

## 6. Conclusion

Student qualitative feedback analysis is critical for improving the teaching and learning process. In this paper, we described LDA based tool for the textual analysis and visualizations for faculty, curriculum managers and course designers. Students suggestions can be noisy is extracting the sentiments and hence filtering the suggestions will be useful in improving the performance of the tool. Further, we are working on the correlation analysis between qualitative and quantitative feedback.

## Acknowledgements

This research was supported by the Singapore Ministry of Education, Tertiary Education Research Fund under the research grant reference number MOE2016-2-TR44. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore Ministry of Education. This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore under its International Research Centers in Singapore Funding Initiative.

## References

- Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Learning sentiment from students’ feedback for real-time interventions in classrooms. In *Adaptive and Intelligent Systems* (pp. 40-49). Springer.
- Bing Liu. (2010) Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, 2nd ed, 2010.
- B. Pang, L. Lee, and S. Vaithyanathan. (2002) Thumbs up? Sentiment Classification using Machine Learning Techniques. In *EMNLP 2002*
- Blei, D. M., Ng A. Y., and Jordan M. I. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3 (March, 2003), 993-1022.
- Carl Burch. 2010. Django, a web framework using Python: tutorial presentation. *J. Comput. Sci. Coll.* 25, 5 (May 2010), 154-155.

- Choochart Haruechaiyasak and Chaianun Damrongrat. 2008. *Article recommendation based on a topic model for wikipedia selection for schools*. 5362:339–342
- Christiane Fellbaum (1998, ed.) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Dietz-Uhler, B., & Hurn, E. J. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*.
- Donovan, J., Mader, C. E., & Shinsky, J. (2010). Constructive student feedback: online vs. traditional course evaluations. *Journal of Interactive Online Learning*.
- Haruechaiyasak C and Damrongrat C. 2008. Article Recommendation Based on a Topic Model for Wikipedia Selection for Schools. In Proceedings of the 11th International Conference on Asian Digital Libraries: Universal and Ubiquitous Access to Information (ICADL 08), George Buchanan, Masood Masoodian, and Sally Jo Cunningham (Eds.). Springer-Verlag, Berlin, Heidelberg, 339-342. DOI=
- Nitin Gokran Ila, Venky Shankararaman and Swapna Gottipati "Analyzing Educational Comments for Topics and Sentiments: A Text Analytics Approach" In *Proceedings of 45th Annual Frontiers in Education Conference*, October 2015, El Paso, Texas.
- Leckey, J., & Neill, N. (2001). Quantifying quality: the importance of student feedback. *Quality in Higher Education*, 7(1), 19-32.
- Lockyer, L. & Dawson, S. (2011). "Learning Designs and Learning Analytics". LAK'11, *Learning Analytics and Knowledge*, February 2011, Banff, AB, Canada.
- Loria, S. (2014), *Textblob: Simplified Text Processing*. <http://textblob.readthedocs.org/en/dev>
- Ma, Baizhang & Zhang, Dongsong & Yan, Zhijun & Kim, Taeha. (2013). An LDA and synonym lexicon based approach to product feature extraction from online consumer product reviews. *Journal of Electronic Commerce Research*. 14. 304-314.
- Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 2301-2309.
- Ming, N., & Ming, V. (2012). Predicting student outcomes from unstructured data. Proceedings of the 20th Conference on User Modeling, Adaptation, and Personalization.
- Polarity Analyser from  
[https://github.com/sayonetech/text-polarity-analyser/blob/master/polarity\\_analyser.py](https://github.com/sayonetech/text-polarity-analyser/blob/master/polarity_analyser.py)
- Ramanand, J., Bhavsar, K., Pedanekar, N. (2010) Wishful thinking: finding suggestions and 'buy' wishes from product reviews. In: *Proceedings of the NAACL HLT 2010 Workshop* pp. 54–61.
- Rashid, A., Asif, S., Butt, N. A., & Ashraf, I. (2013). Feature Level Opinion Mining of Educational Student Feedback Data using Sequential Pattern Mining and Association Rule Mining. *International Journal of Computer Applications*, 81(10), 31-38.
- Scott, G., Grebennikov, L., & Shah, M. (2008). Using qualitative data to prove and improve quality in Australian higher education. *Evidence based Decision Making: Scholarship and Practice*, 97.
- Sherin, B. (2012). Computing student science conceptions with Latent Dirichlet Allocation. Paper presented at the 10th International Conference of the Learning Sciences.
- Southavilay Vilaythong, Kalina Yacef, Peter Reimann, and Rafael A. Calvo. 2013. Analysis of collaborative writing processes using revision maps and probabilistic topic models. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13), ACM, New York, NY, USA, 38-47.
- Gottipati Swapna, Shankararaman Venky, Jeff Rongsheng Lin. "Text analytics approach to extract course improvement suggestions from students' feedback". In *Journal Research and Practice in Technology Enhanced Learning*. 2018.
- Venky Shankararaman, Swapna Gottipati, Sandy Gan."A Conceptual Framework for Analyzing Students' Feedback" *IEEE Frontiers in Education Conference (FIE)* (FIE 2017, Oct), Indianapolis.
- Vilaythong Southavilay, Kalina Yacef, Peter Reimann, and Rafael A. Calvo. 2013. Analysis of collaborative writing processes using revision maps and probabilistic topic models. In *Proceedings (LAK '13)*, NY, USA, 38-47.
- Wei Kuang, Nianlong Luo, and Zilei Sun. 2011. Resource recommendation based on topic model for educational system. *Information Technology and Artificial Intelligence Conference (ITAIC)*, 2011 2:370–374, Aug
- Zhang, Y., Law, N., Li, Y., and Huang, R. (2012). Automatic extraction of interpretable topics from online discourse. In *International Conference of the Learning Sciences (ICLS) 2012*, Volume 1, (p.443-450).