Discovering Unknown Reflection: Exploring Sentimental Intentions of Online Teaching Evaluation in Adult Technology Education

Yankun He, Kenan Xiao, and Yuewei Shi
Auburn University

Abstract: Sentiment analysis (SA) is the process of identifying and classifying users' opinion from a piece of text into different sentiments. Student's evaluation of teaching is one of the common and necessary measures to assess the teaching quality of course instructors in a college setting. Traditionally, student evaluations are administered towards the end of the semester using a paper-based survey. However, recently online evaluations, such as RateMyProfessor.com (RMP) are becoming popular. This study collected 490 comments of professors in the Computer Science and Software Engineering department at Auburn University that meet the selection criteria from RMP. The research utilized a textBlob Python package to analyze sentiments of students' comments from different class standings.

Keywords: sentiment analysis, students' evaluation, RateMyProfessor

Student evaluations of teaching (SET) is one of the joint and necessary measures to a teaching quality of course instructors in a college setting. Besides, SET is also regarded as a significant consideration for promotion, tenure, and merit at most higher education institutions (Otto et al., 2008). Coladarci and Kornfield (2007) pointed out that SET can be an essential source of information for improving teaching and informing personal decisions when instruments are appropriately conducted, and the resulting data are thoughtfully considered. Different evaluation designs exist to conduct SET, and the most often SET survey instrument consists of plenty of fixed-ended questions on which students evaluate their instructors (Brockx et al., 2012). Some researchers used the scores on fixed-ended questions to generate a statistical report. For example, Diseth (2007) used evaluation scores conducted by 248 undergraduate psychology students to analyze the relationship between their evaluation perception and academic achievement. These fixed-ended questions contain both standardized questions rated on a Likert scale and open-ended questions, which complement standardized questions. Open-ended questions, which capture students' opinions, are not covered by the standardized questions and give greater freedom of expression (Baddam et al., 2019).

Traditionally, SET is administered by paper-based surveys, and students are usually asked to finish the evaluation on their instructor at the end of the semester. In recent years, online evaluation websites are growing up and becoming famous. Students comment about their educational experience in online forums or teacher review sites such as RateMyProfessors (RMP), Uloop, and Teacher Complains. These websites allow students to anonymously rate their professors and share their educational experience with great freedom. More importantly, these online comments open to the public. These feedbacks yield valuable insights for university administrators and help students clarify which universities to attend or courses to take (Abdelrazeq et al., 2015).

Sentiment analysis (SA) identifies and classifies users' opinions or attitudes from a piece of text, such as positive, negative, or neutral sentiments and different emotions like delighted, dismal, excited, or angry. SA plays a significant role in many areas like consumer feedbacks, social media monitoring, product analysis and so forth. Moreover, in the education community, SA can be applied to analyze the text written by students, whether in formal course surveys or informal comments from online platforms (Rani & Kumar, 2017). In this way, researchers have an approach to determine students' interest and preference in a class and figure out, which areas should be improved through corrective actions or methods.

This study uses the sentiment analysis method to analyze students' sentiments in different class standing (freshman year, sophomore year, junior year, and senior year). We collected 620 students' comments from RMP of the Computer Science and Software Engineering Department at Auburn University by using the Python Scrapy package. 490 comments met the selection criteria and contained complete information, including the professors' names, students' comments, the rate scores, etc. Moreover, for the sentiment analysis, we used the textBlob package, which takes in a piece of text, and returns the polarity and subjectivity of the text.

Literature Review

Several previous researchers focused on the bias evaluation of online SET and the validity of online SET as a measure of student learning. Online faculty rating sites are disputed because data from these websites are characterized by biases such as instructors' teaching styles, personal charisma, sense of humor, and grading leniency (Liaw & Goh 2003). In fact, these biases are not as valuable as a measure for either faculty performance or student learning. Feeley (2002) found that RMP ratings were affected by a halo effect so that it cannot reflect student learning. In this situation, instructors who have decent appearances or show leniency towards students' grades tend to achieve overall positive ratings. Moreover, online ratings may be entered by anyone, and at any time, they may be affected by emotion. Some students may have potential bias when rating their professors online (Otto et al., 2008). However, some researchers also stated that it is possible that online ratings may not be biased. Students who post ratings and comments on the websites did have experience with the professors. From this point of view, online ratings may be representative. Furthermore, Hardy (2003) pointed out that if some students give biased ratings, these comments are balanced between positive ones and negative ones.

Online student ratings could improve both student learning and instructor performance. Students can read previous rating information to choose instructors who are best suited to their learning preferences. For instructors, they can learn from online comments to improve their teaching performance. It is no doubt that these improvements are dependent on the validity of the rating as a measure of student learning. Previous studies created several models to examine the validity of SET. The research findings hold that learning is positively associated with instructors' clarity and instructors' helpfulness, while course difficulty is not linearly associated with student learning. Students could not learn best when courses are too difficult or too easy, and they will have better learning outcomes when the difficulty level is moderate and between these extremes (Centra, 2003; Harrison et al., 2004). Otta et al. (2008) used those models to examine whether the ratings reflect a halo effect or student learning. Their study showed that statistical support for RMP ratings reflect student learning.

In recent years, researchers have begun to apply sentiment analysis (SA) to the education areas. Ortigosa et al. (2014) proposed a SA approach for the e-learning environment using a combined method of Spanish lexical based and machine learning techniques. Rani and Kumar (2017) used natural language processing and machine learning to analyze student feedback, collected from both course surveys and online comments. Through these methods, they identified sentiment polarity and the emotions expressed. Baddam et al. (2019) analyzed sentiments from different class standing in the business department using text mining. These researchers helped university administrators and instructors to address problem areas in teaching and learning.

Methodology

This study aims to use the sentiment analysis method to analyze students' sentiments in different class standing. For the data collection step, we collected 620 students' comments from RMP of the Computer Science and Software Engineering Department at Auburn University using the Python Scrapy package. Four hundred ninety comments met the selection criteria and contained complete information: professors, courses, comments, quality scores, and difficulty scores. We used a textBlob package for the sentiment analysis, which takes in a piece of text and returned the text's polarity and subjectivity.

Results

Overall, students from different class standings show moderately positive sentiments about their professors. The negative sentiments about the professors reach the peak during sophomore year. In terms of students in their third and fourth years, they show relatively satisfaction with their professors and give positive comments. The results also show that graduate students mostly leave positive comments on their professors (see Table 1.)

Table 1. *Sentiment Results*

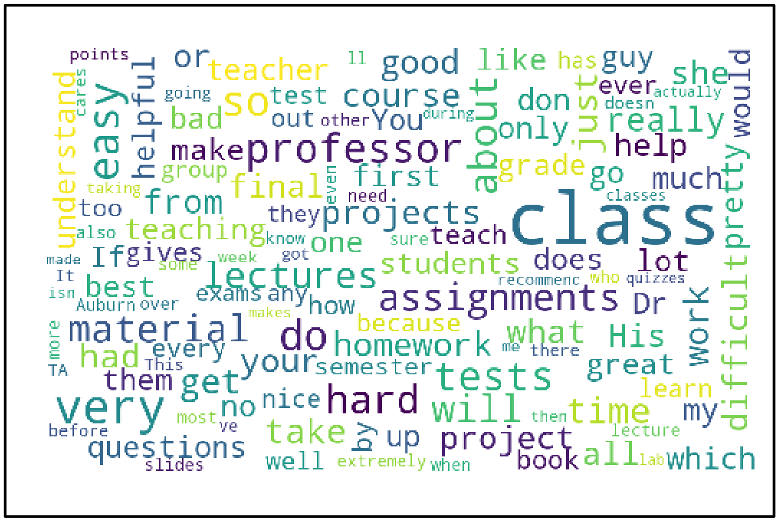
Category	Subcategory	Very negative	Moderately negative	Moderately positive	Very positive
Class Standing	Freshman	2.50%	38.75%	53.13%	5.63%
	Sophomore	1.79%	42.86%	48.21%	7.14%
	Junior	0.00%	25.87%	68.53%	5.94%
	Senior	0.00%	18.29%	69.51%	12.20%
	Graduate	4.44%	28.89%	57.78%	8.89%

Figure 1 demonstrates the word cloud that includes the most frequently used words that might reveal sentiments in these comments.

Discussion

This study had two limitations. First, the RMP's online comments were only collected from the Computer Science and Software Engineering Department at Auburn University, so the data was relatively small. Secondly, the dataset lacked details about students' information since a secondary source provided it. In this case, we identified students' class standing by courses rather than their actual class standings.

Figure 1. *Words Cloud of Graduate Student Comments About Professors*



In the future study, we will design a model to examine the new version of RMP's validity. In previous studies, researchers used three models to examine the validity of online SET. They found that learning is positively associated with instructors' clarity and instructors' helpfulness, while course difficulty is not linearly associated with student learning. However, the RMP website had modified the evaluation scales after 2016. They changed helpful and clarity scales into overall quality and added the question: "would you take again?". We should focus on building up a new model to adapt to the new version of RMP for further study. Moreover, we will also design a study to explore the reasons that cause sentiment change among different class standings. Besides, we will provide a path for adult educators to discover the sentiments of learners and improve instructors' teaching effectiveness.

References

Abdelrazeq, A., Janßen, D., Tummel, C., Jeschke, S., & Richert, A. (2016). Sentiment analysis of social media for evaluating universities. In S. Jeschke, Isenhardt, I., Hees, F., & Henning, K, *Automation, Communication and Cybernetics in Science and Engineering 2015/2016* (pp. 233-251). Springer.

Aung, K. Z., & Myo, N. N. (2017, May 24-26). Sentiment analysis of students' comment using lexicon-based approach. *Proceeding of the IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. https://ieeexplore.ieee.org/document/7959985

Baddam, S., Bingi, P., & Shuva, S. (2019). Student evaluation of teaching in business education: Discovering student sentiments using text mining techniques. *The e-Journal of Business Education & Scholarship of Teaching, 13*(3), 1-13.

Brockx, B., Van Roy, K., & Mortelmans, D. (2012). The student as a commentator: students' comments in student evaluations of teaching. *Procedia-Social and Behavioral Sciences, 2012*(69), 1122-1133.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*(5), 495-518.

Coladarci, T., & Kornfield, I. (2007). RateMyProfessors.com versus formal in-class student evaluations of teaching. *Practical Assessment, Research, and Evaluation, 12*(1), 1-15.

Diseth, Å. (2007). Students' evaluation of teaching approaches to learning, and academic achievement. *Scandinavian Journal of Educational Research, 51*(2), 185-204.

Hardy, N. (2003). Online ratings: Fact and fiction. *New Directions for Teaching and Learning, 2003*(96), 31-38.

Harrison, P. D., Douglas, D. K., & Burdsal, C. A. (2004). The relative merits of different types of overall evaluations of teaching effectiveness. *Research in Higher Education, 45*(3), 311-323.

Hugh Feeley, T. (2002). Evidence of halo effects in student evaluations of communication instruction. *Communication Education, 51*(3), 225-236.

Liaw, S. H., & Goh, K. L. (2003). Evidence and control of biases in student evaluations of teaching. *International Journal of Educational Management, 71*(1), 37-43.

Rani, S., & Kumar, P. (2017). A sentiment analysis system to improve teaching and learning. *Computer, 50*(5), 36-43.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior, 31*(1)*,* 527-541.

Otto, J., Sanford Jr, D. A., & Ross, D. N. (2008). Does RateMyProfessor.com really rate my professor? *Assessment & Evaluation in Higher Education, 33*(4), 355-368.