Ziqi Tang, Yutong Wang
Professor: Jiebo Luo
CSC 440: Data Mining

November, 03, 2020
Project Proposal

Do top school students tend to criticize professors more?: Mining and Summarizing comments on RateMyProfessor.com

Introduction

College students nowadays have very different styles in making their academic plans compare with days before the increase in internet use. In order to make sure they select the fittable courses, college students choose to use rating websites such as ratemyprofessors.com, PassCollege.com, Reviewum.com and etc [3]. Among all college professor rating websites, ratemyprofessors.com is the most popular one. Ratemyprofessors.com is a large rating website for professor ratings, course ratings and school ratings. There are more than 19 million ratings, 1.7 million professors and 7500 schools on the website (ratemyprofessors.com). Students who took the courses before can rate the professor on overall course quality, course difficulty, the grade received, comments keyword hashtags and etc. based on their experience in the course. Meanwhile, students who want to select a certain course by a specific professor can check the former ratings and comments to decide whether they think it is a suitable course for them.

Problem statement

Students may have different preferences for the courses and professors when choosing a course. For instance, one student may want to have a more challenging class where he/she learns a lot during the course, and another student may want to choose an easy course or a generous professor who gives an easy A for the grade. Similarly, rating standards may vary based on different perspectives as well. Students may give a higher rate to professors because of teaching quality, the hardness of exams, personality, charisma, grading leniency etc. In this project, we want to study the factors that affect the ratings of professors and discover the knowledge behind the ratings. We want to find the answers to the following questions: 1. What factors affect the ratings? 2. Do these factors vary among different schools? 3. What is the relationship between the factors and school ranks (based on us news ranking 2020).

Data acquisition

We could utilize the open resource scraper written by Rodantny and Karthik Raman. These two scrapers are able to scrape data from all professors of a single University. The scraped data includes both aggregate information as well as detailed review scores. The scrapers require the ID of the interested universities, we will need to manually look up the IDs from RateMyProfessor.com, or write another scraper to collect the IDs if there are too many universities we want to scrape. Figure 1 is a screenshot of what a rating page and an individual rating look like for one professor. We circle the information that we may interest in scraping.

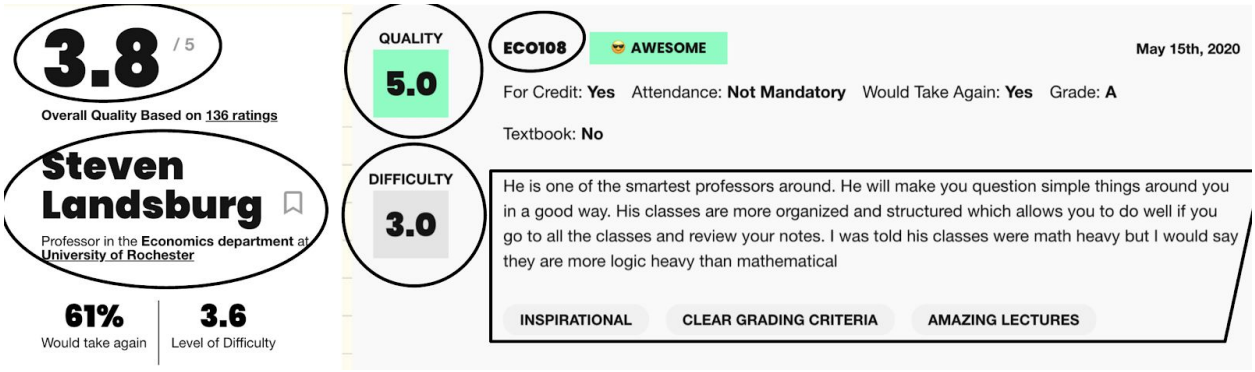

Figure 1: a screenshot of rating page in ratemyprofessors.com of a professor from U of Rochester

Algorithm ideas

We have found some papers about abstracting information from customer reviews, some of the findings can possibly be applied in our project. Liu and Hu [5] mentioned in their paper that they used NLP to identify adjective words in the customer reviews; these words normally are used to express opinions, so they call them opinion words. A bootstrapping technique was proposed to determine the semantic orientation using WordNet [4]. Liu and Hu performed Apriori to abstract frequent patterns with length less than 3 words. Then they performed Compactness Pruning and Redundancy Pruning. By doing so, we could get rid of meaningless feature phrases. Etzioni [6] proposed some methods to improve the performance of unsupervised information extraction. He used pointwise mutual information(PMI) to indicate the likelihood that words and phrases belong to the same category. Abulaish pointed out that the distribution of an overwhelming majority of reviews posted in online markets is bimodal [1], which means the scores, or ratings of the reviews may not convey a lot of information. He emphasized the importance of the reviews and proposed an opinion mining system, where he created a subjectivity/objectivity analyzer to filter out more objective comments and leave subjective ones. We may use this idea the opposite way, where we leave objectivity to show how likely students are to criticize the professors.

Reference

[1] Abulaish M., Jahiruddin, Doja M.N., Ahmad T. (2009) Feature and Opinion Mining for Customer Review Summarization. In: Chaudhury S., Mitra S., Murthy C.A., Sastry P.S., Pal S.K. (eds) Pattern Recognition and Machine Intelligence. PReMI 2009. Lecture Notes in Computer Science, vol 5909. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-11164-8_35

[2] Arbaugh, J. B. (2001). How Instructor Immediacy Behaviors Affect Student Satisfaction and Learning in Web-Based Courses. *Business Communication Quarterly*, *64*(4), 42–54. https://doi.org/10.1177/108056990106400405

[3] James Otto, Douglas A. Sanford Jr & Douglas N. Ross (2008) Does ratemyprofessor.com really rate my professor?, Assessment & Evaluation in Higher Education, 33:4, 355-368, DOI: 10.1080/02602930701293405

[4] Miller, G. A. et al. “Introduction to WordNet: An On-line Lexical Database.” International Journal of Lexicography 3 (1990): 235-244.

[5] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '04). Association for Computing Machinery, New York, NY, USA, 168–177. https://doi.org/10.1145/1014052.1014073

[6] Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S.Weld, Alexander Yates 2004. Unsupervised named-entity extraction from the Web: An experimental study. Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA https://doi.org/10.1016/j.artint.2005.03.001

[7] Silva, K. M., Silva, F. J., Quinn, M. A., Draper, J. N., Cover, K. R., & Munoff, A. A. (2008). Rate my professor: Online evaluations of psychology instructors. *Teaching of Psychology, 35*(2), 71–80. https://doi.org/10.1080/00986280801978434