Aspect-Based Opinion Mining by Location from Consumer Reviews

Charlotte Chen chc190@pitt.edu
Dan Hui dth26@pitt.edu
David Neiman dsn9@pitt.edu
David Tsui dat83@pitt.edu

Abstract

We propose to investigate aspect-based opinion mining for Yelp reviews, filtering by location. This will enable us to draw conclusions about consumer insights and compare them by regions.

Introduction

Sentiment analysis is an application of text classification to determine whether a piece of writing is positive, negative, or neutral. By analyzing how people think and feel, we can gain considerable insight into consumer preferences and make predictions about future trends and expectations. Restaurant reviews are only one of many sources for product insights; consumers regularly use platforms such as Yelp, Amazon, and TripAdvisor to leave feedback and read other comments when considering whether to buy a product.

Sentiment analysis with aspect-based opinion mining uses explicit and implicit aspects to derive conclusions about data. Explicit aspects directly highlight specific targets in a sentence, while implicit aspect extraction relies on other information, implicit aspect clues (IACs), which are used to imply meaning about a specific target. The sentence "The price is too high" explicitly provides a negative opinion about the *price* aspect, while the sentence "This camera is sleek and very affordable" implicitly provides a positive opinion about the aspects *appearance* and *price*.[1] Using these features, we can produce a quantitative measurement of the writers' opinions.

Approach

We will be using the data set from the Yelp Data Challenge. Location is provided in several forms: neighborhoods, full addresses, city, and state. User reviews contain the number of stars given, as well as the number of votes given and the text feedback. These reviews will be grouped by city and state.

To compare consumer insights by regions, we will first group each set of opinionated texts by the specified areas. From the grouped regions, we proceed with sentiment analysis on texts using explicit and implicit aspect extraction to find the relationship each aspect has with the review. Then, frequent aspect extraction can be used to identify the common opinions across the consumer reviews in a given region. In this way, correlations can be found in common aspects with reviews given a location.

---

[1] http://www.aclweb.org/anthology/W14-5905

Several possible corpi for aspect extraction are provided on the SemEval web site (http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools), which include restaurant reviews with polarity using both category (positive, negative, neutral), and quantitative measures. One corpus data set for implicit aspect extraction is located at http://www.gelbukh.com/resources/implicit-aspect-extraction-corpus/, which contains several sets of product reviews.

One technique we will use for both explicit and implicit aspect extraction is association rule mining. By using the apriori algorithm, we can generate frequent itemsets and compute the frequency of a relation in the data set. This will allow us to search the relationships in our training data to see if given aspects and opinionated words are frequently correlated with certain feelings.

Additionally, Zhang and Liu discuss how opinion words that modify frequent aspects can also modify infrequent aspects, such that we can apply these opinionated words to a generalized instance. They use the example of having *picture* as a frequent aspect and using the sentence "The picture is absolutely amazing" to infer meaning in the sentence "The software is amazing".[2] In this manner, we can categorize sentences that are similar to our training data using new nouns or combinations/ordering of words, as well as by searching for known relationships among words.

Evaluation

In evaluating our explicit and implicit aspect extraction algorithms, we hope to correctly categorize 60% of a test data set as positive, negative, or neutral. Several published approaches have produced between 70-90% success both in precision and recall. Explicit aspect extraction has been widely studied and we hope to develop an algorithm that will meet this goal. Implicit aspect extraction poses a more difficult problem. Thus, we hope to use this technique to augment our results from the explicit aspect extraction by 10% or more.

To evaluate the accuracy of our results, we plan on using Amazon's Mechanical Turk to have human labeling of a given opinionated text to several aspect choices. For example, a text will be given the choices of "price", "service", "atmosphere", etc. as options for the text and matching polarities. We will compare our algorithm's predictions with the human labels for accuracy.

Milestones

By March 18th, we plan to have finished data pre-processing. By the end of March, we will have finished our work on association rule mining, and will have started to work on implicit aspect extraction through commonalities in words and sentence structure. We will work on the implicit aspect extraction algorithm until April 15th, at which point we will finish collecting and analyzing the data. We will present on April 28th and submit our report by April 29th. The group members will share responsibilities equally.

---

[2] https://www.cs.uic.edu/~lzhang3/paper/ZhangLiu-AEEE.pdf

References

Bing Liu, Lei Zhang. Aspect and Entity Extraction for Opinion Mining. Available at https://www.cs.uic.edu/~lzhang3/paper/ZhangLiu-AEEE.pdf

Bo Pang, Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. Available at http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf

Eric Cambria, Alexander Gelbukh, Chen Gui, Lun-Wei Ku, Soujanya Poria. 2014. A Rule-Based Approach to Aspect Extraction from Product Reviews. *Proceedings of the Second Workshop on Natural Language Processing*. Available at http://www.aclweb.org/anthology/W14-5905