2nd Year Project – Project Phase 2
Finer-Grained Sentiment Analysis on Amazon Product Reviews

Barbara Plank

April 12, 2019

So far, we have focused on sentiment analysis systems using existing data and developed a predictive system to assign sentiment polarity (positive, negative) to an entire review text. In this second phase of the project, we will dig deeper into sentiment review texts, and focus on finer-grained sentiment analysis. This includes data annotation and engaging with the existing literature.

For instance, consider the following review text (from our *games* dataset used in phase 1):

> Both a fan of STAR WARS and GameBoy Advance, when I got this game I was enthralled, it's so awesome! The controls are good, and the levels are entertaining and expertley-done. For people who like these games, you'll find this one is great fun. The only complaint is that the graphics should have been a bit better. But apart from that it's a reallty fun and exciting game. This is one of the most impressive GBA releases. I highly reccomend this one to any gaming fan and SW enthusiast.GRAPHICS: 7/10SOUND: 9/10FUN FACTOR: 9/10

What do you observe in this review with respect to:

a) overall sentiment?

b) finer-grained sentiment, i.e., aspects? What are aspects here? What challenges do you see if you were to automatically detect these aspects and their sentiment polarity?

c) other observations from this review example that might make automatic sentiment detection difficult?

Discuss with your neighbor. Read further after having answered these questions.

Warm-up on annotator agreement: Calculate Cohen's kappa manually for the example provided in the `agreement` folder. Answer the question.

Finer-grained sentiment analysis

While the overall sentiment of the review example is *positive*, we see that the reviewer expresses sentiment with respect to certain *aspects* (e.g. controls, graphics) and that the sentiment polarity towards a certain aspect might be different compared to the overall sentiment.

The overall questions we are posing to ourselves in this phase are: *How can we dig deeper into product opinions? Can we identify what sentiment is expressed about aspects of products (by proxy)?*

There is an entire research field that is concerned with fine-grained sentiment analysis: **Aspect-based sentiment analysis (ABSA)**. It focuses on the extraction of fine-grained opinion phrases from review texts. For example, consider the following examples taken from [3]. Each example indicates for each *aspect* the corresponding *subjective phrase* and its sentiment polarity:

Some examples are given below, with aspects marked in blue and subjective phrases marked in red:

- I had no problems with the return.
  - *return* is a target of *no problems*. *no problems* is positive.
- The washer itself is great, the included hose is junk.
  - *washer* is a target of *great*, *hose* is a target of *junk*. *great* is positive, *junk* is negative.

Figure 1: Examples of aspect-based sentiment analysis [3]

Quoting the original paper: "The annotators were instructed to regard everything as an aspect that is part of a product or related to it and can influence the opinion about it, including the whole product itself."

ABSA is a non-trivial problem. Annotation for ABSA is complex and it takes considerable time (in fact, read [3] and the discussion therein about annotation time). Therefore, in this phase we will try to get to finer-grained sentiment analysis of aspects by proxy. Instead of identifying subjective phrases and their targets, we will focus at the *the sentence level*: break review texts down to the sentence level and identify a) which sentences carry evaluations (sentiment) about the product that is reviewed; b) assign sentiment polarity at the sentence level.

Once we know which sentences carry positive or negative (or neutral) sentiment for a product within a review, they can be further analyzed to extract the particular aspects that contribute to these sentiments. For example, the subsequent analysis could focus on the identification of certain words (e.g., nouns) and identify their closest modifiers (e.g., adjectives). This typically constitutes a very strong baseline for ABSA.

The following readings are highly recommended before/while doing this assignment:

Requested reading:

- Jurafsky and Martin chapter 4 (recap), 19 (optional)
- Klinger and Cimiano (2014) [3]
- Pang and Lee (2002) [2]
- Hu and Liu (2004) [4]

1 Assignment Phase 2

Your phase 2 group challenge is to create a fine-grained sentiment analysis dataset at the sentence level for a new Amazon product domain: baby products. You are asked to do *manual annotation* and evaluation of the *annotation quality*. You must also document all decisions that you took, and present your outcomes in the phase 2 presentations. You are expected to work on this in this lab and on Tuesday after Easter, and present your outcome in the group presentations on Friday, April 26. Notice: since this phase is shorter in comparison to the other phases, the group presentation should be **5 minutes**.

1. Data exploration. Identify and download the data set that is available for your group on our github page `project-phase2` Examine the dataset. Is the dataset pre-processed (tokenized?). How many reviews are in the data? What is the meta-data which is available? (e.g. what is `asin`? tip: you can use it on the Amazon online catalogue to find back the original product, which might be helpful for the annotation)

2. Assignment:

   - Annotation guidelines: Develop a set of annotation guidelines to categorize sentences into product-related sentiment polarity statements (positive, negative, maybe neutral). Decide how you handle other sentences, i.e., those that are part of the review but do not express sentiment about an aspect of the product under review (e.g., other products, other things you might find).
   - Annotation round 1: pair up in two people within your group. Select 2 reviews each, split them up into sentences, and annotate the sentiment polarity at the sentence level (individually). Compare your results. Calculate agreement statistics (kappa scores). What do you observe? Discuss in the group. Revise your annotation guidelines if necessary.
   - Annotation round 2: annotate another 2 reviews each. Compare your annotations, calculate agreement statistics (kappa score). What do you observe?
   - Once you have reached a sufficient level, split up the rest of the reviews amongst yourselves, and annotate them individually. Make sure you can relate each sentence back to its original review.

- Summarize your fine-grained sentiment analysis dataset.
- Present your findings in the group presentation 2. Include a slide on future work which contains concrete research questions (and overall direction) you would like to tackle/explore in phase 3 of your project (it has to be related to aspects we discussed during phase 1 and 2).
- (Optional/bonus): Extract topics that can be found in the product-relevant opinion sentences (positive, negative).

3. Discuss and present your solutions to the following considerations:
   - Sentence segmentation. How did you segment your reviews into sentences? Develop/use a sentence segmenter to do so - what are possible considerations?
   - How did you decide to annotate sentences which are not about an aspect of a product?
   - How did you handle difficult cases? (e.g., how do you handle examples like the second one in Figure 1? What impact might your decision have on a system that automatically detects the sentence-based polarity?)
   - How did you refine your guidelines, what were important considerations during the different annotation phases?
   - Note: We will collect all annotations from all groups and in phase 3 evaluate your system on the data you have collectively annotated. Start thinking now: which machine learning system would you implement for sentence-level sentiment classification on these annotated reviews?
4. Other considerations:
   - Group repository. Make sure you use your group repository throughout the entire project phase. Logging by group members will be important.
   - Read the Klinger and Cimiano (2014) paper [3] (individually)
   - Read the Hu and Liu (2004) paper [3] (individually).
   - Group work. Discuss methods or ideas from these two papers that could be useful for your own project on sentiment analysis. Discuss limitations of the papers and outline possible solutions.

2 The hand-in (and oral presentation) and deadlines

You must hand in the following:

1. A link to the group project github repository
2. The code snippets (here Python notebooks are perfectly fine)
3. The project group presentation slides (as **pdf**)

Deadline: You will have to upload your group presentation slides by **Friday, April 26, 2019** at 08:00 CET, which is the day in which your group will present the project.

Submission instructions: Upload the code and pdf to LearnIT. Make sure your slides include a link to the group github repository.

Your **5 (! shorter) minute oral presentation** should correspond to the structure of a general scientific presentation on fine-grained annotation of Amazon product reviews, their challenges and solutions.

References

[1] Jurfasky and Martin, In Preparation. *Speech and Language Processing (3rd ed. draft)*. Available at https://web.stanford.edu/~jurafsky/slp3/

[2] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. *Thumbs up? Sentiment classification using machine learning techniques*. In EMNLP 2002. Available at: http://www.cs.cornell.edu/home/llee/papers/sentiment.home.html *Received the Test-of-Time Award at NAACL 2018* https://naacl2018.wordpress.com/2018/03/22/test-of-time-award-papers/

[3] Roman Klinger and Philipp Cimiano. *The USAGE Review Corpus for Fine Grained Multi Lingual Opinion Analysis*. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). http://www.lrec-conf.org/proceedings/lrec2014/summaries/85.html

[4] Minqing Hu and Bing Liu (2004). *Mining and Summarizing Customer Reviews*. In KDD. https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf