Peer Review - Project Two

Anonymous

April 4, 2016

Peer-assessment

First I'd give some general feedback, and then I will quickly give feedback on each of your sections.

General Remarks

First of all, it's a nicely structured report, the sections are short and to the point, and everything seems to be covered.

You tend to use both I and We in your report, I find that highly distracting, I recommend you pick one or the other. There's some misspellings and grammatical errors in your report, for example, in **1. INTRODUCTION** the sentence *"In order label the sentiment"* is grammatically incorrect. You also vary between past and present tense (use / used), I recommend you do additional proofreading for the final hand-in.

I do not recommend including specific code into the main part of your report, how a tool functions may change with a new release, and it's generally not relevant for the average reader, even if the reader is a fellow student. Making a reference to the tools you're using will allow the reader to research specifics if need be. If you still need to show code, include them as an appendix, or make the scripts evident in the .tar.gz, which accompany the report.

Introduction

It introduces the reader to the problem you're trying to solve nicely. However, I dislike your use of references in this section, as it should be sufficient to have a reference after mentioning the dataset for the first time, and not after each of your examples, having this many references to the same source in this short a period is a messy. You have a reference to SentiStrength, but your reference is simply the word "SentiStrength", adding a hyper-link, author and/or paper is strongly recommended, I think you forgot this.

METHODOLOGY

SentiStrength

A quick introduction to SentiStrength could be nice, what is it? Why was is chosen versus other tools? Strengths and Weaknesses of SentiStrength?

I like your choice of weighing negative sentiment stronger, as it's rarer. You mention Maite Taboada here, include a reference so a curious reader and research further.

You talk about having reviews without annotations, I cold do with a short introduction to the dataset, who generated it? How large is it? Is this ignored part a relevant part? Remember the reader can be anyone with a scientific background, not necessarily just a fellow student or teacher.

Likewise, why do you need to re-annotate the data? Was it not sufficiently annotated?

In the introduction you said you would use 3-fold, and now you're using 10-fold. Why? SentiStrength supports 3-fold as well as 10-fold. I'm not satisfied with the reason given (better results? how so? Is our dataset large enough to give a sufficiently large sample size when 10-folding? What do the teachers ask for?)

You mention term weight at the very end of the section. I would have liked a better introduction to the method at work, with terms having weights. Who supplies the original weight? Can new words be introduced? It's not impossible to think that the forum from which our data is mined uses certain words to carry some weight, how can we teach these to SentiStrength?

Stanford

Just to be clear, Stanford also has 0 and 5 for sentiment values.

Skipping reviews which are too long is of course a valid option, however, as far as I can tell from your dataset file, you are now down to 3300 of the original 4900 reviews, just be careful you don't get a too small dataset. Also, reviews which are longer than 150 characters may be very relevant, as they are very likely the actual reviews, while texts which are smaller often are feedback to the actual reviews. Maybe another solution can be determined? For example. splitting long reviews, as the overall sentiment shouldn't change.

I like how you explain the general idea with a 3-fold approach, so much so that I recommend you move it to the start of the report, and maybe make a section (terminology) for terms like 3-fold etc which may not be given for every reader.

Findings

You value some of your results as "good"/"pretty good". What is good? Compared to what? (Stanford claims 85.4

If you don't make a reference to your tables, some readers may ignore them. Guide the reader, by making references to the correct table when needed, also Table 3 is

positioned as if it's part of the conclusion, it should appear in the Findings section.

You claim an average value of 72% in your SentiStrengh folds, I suggest you make a table holding these values, as you've done with Stanford CoreNLP.

Table 2: "Accuracy on the whole LEGO dataset using the training models from above". From above is a very weak reference, are you referring to text, or the table? Use \ref \label methods in latex to show the connection.

61% being not satisfying may need a bit of elaboration. How much is enough for it to be satisfying?

Conclusion

You mention making improvements, that's very general, which improvements? Which sections of your project would benefit from being improved?

What would be a satisfying result?

References

In reference [1] you spell lego with lower case, everywhere else LEGO is in capital font. [2] SentiStrength is hardly a reference.

In Recap

Overall a good first draft for a final report. I didn't get lost at any point, and I'm convinced the author has a good understanding of the methods described. I'm however missing some deeper evaluations and thoughts on the results, "pretty good" is not enough, I would love to see some elaborations on why you are/not satisfied with your results?

I did not execute your code while reviewing your assignment, as I assume it works, however, while skimming the code, I would've liked some comments and the amount of disabled code doesn't make the code any easier to understand. Is the disabled code irrelevant?