IST736 Text Mining

HW1

An Evaluation of Sentiment Classification Tools

Artificial Intelligence (AI) has become a popular topic recently. Assume that you are a consultant at a public relations firm, and a client of your firm would like you to evaluate the current public sentiment toward AI in social media like Facebook and Twitter.

Since there are too many comments on social media, you can't manually collect and analyze them all. Fortunately you have discovered some free sentiment analysis tools, and now need to evaluate whether they are good enough to do sentiment analysis for your assigned task.

You need to collect a sample data set and choose two tools to compare their effectiveness in sentiment analysis. Write a report to describe (1) your sampling strategy and whether it would result in a representative sample of public sentiment toward AI, (2) data preparation and system evaluation process, and (3) your conclusion on whether these tools are suitable for your task.

A few examples of tools:

If you would like to do some programming for this assignment, here are a few popular tools you could consider:

1. NLTK’s built-in sentiment analysis tools. See sample code in <http://www.nltk.org/howto/sentiment.html>
2. VADER <https://github.com/cjhutto/vaderSentiment>

If you prefer GUI-based tools at this time, consider:

<http://sentistrength.wlv.ac.uk/>

<http://text-processing.com/demo/sentiment/>

Your report should

* Be formatted as a research paper that includes introduction, method, result, and conclusion sections.
* Up to 4 pages.
* One inch margin on all sides.
* 12pt Times New Roman or 11pt Arial.
* Attach your data file in csv format.

**Grading criteria:**

Report grading is similar to reviewing academic papers. Grading will focus on the following aspects:

1. The analytical methods are all used correctly, and the interpretation is convincing. (50%)
2. The report has clarity of presentation. It is well organized, clearly written, within page limit (25%)
3. The report provides sufficient information for others to replicate the analyses. (Necessary information includes but is not limited to problem definition, data description, description of analytical methods, processes, result interpretation, and conclusion). (25%)