Assignment 6

L245

Due Wednesday, April 5

1. Select two similar types of articles in the *Indiana Daily Student* (`http://www.idsnews.com/`), but written by different authors. Importantly, these should be articles within the same genre (two editorials, two sports reports, etc.)

   (a) Give the title, author's name, and date of each article.

   (b) Describe at least three stylistic differences between the two authors.

   (c) Could these differences be detected automatically? How?

2. I want you to read certain parts of a paper by Koppel, Schler, & Argamon (2009) (*Computational Methods in Authorship Attribution*, `http://onlinelibrary.wiley.com/doi/10.1002/asi.20961/epdf` available on Canvas under *Files*) and answer the following questions:

   (a) **Introduction** (p. 9): Describe in your own words the difference between the *profiling*, *needle-in-a-haystack*, and *verification* problems.

   (b) **Profiling** (p. 15–17): Of the four types of profiling tasks mentioned: i) which seems most difficult (and why)? ii) how do the most important features differ for the different tasks?

   (c) Pick one of the four profiling tasks (*Gender*, *Age*, *Native Language*, *Personality*) and some set of users on a social media or user-generated content site (Facebook, Twitter, Reddit, etc.). Do the features mentioned in the paper seem to work for the users you look at? If so, provide examples and explain. If not (and even if so), describe other features you see that might distinguish the types of users (e.g., males from females).

   (d) **Bonus:** Read the section on **Authorship Verification** and describe in your own words the intuition behind the *unmasking* method, in particular the use of degradation curves (i.e., pretend you have to describe this to someone who isn't technically-inclined).

3. Do question #3 from chapter 5 of the textbook (p. 151–152).

4. Do question #4 from chapter 5 of the textbook (p. 152).

5. **[Moved to next assignment:]** (Based on a previous assignment by Jason Baldridge):

   Assume you have a corpus of 4200 tweets about broccoli that have been labeled for whether they are positive or negative in their sentiment. 3247 of them are negative. We find the word *hate* in 643 negative messages and in 28 positive messages.

   (a) What is $P(negative)$ according to the above dataset?

   (b) What is $P(hate|negative)$ according to the above dataset?

   (c) What is $P(positive)$ according to the above dataset?

   (d) What is $P(hate|positive)$ according to the above dataset?

   (e) Using the ham-to-spam ratio as a model, what is the positive-to-negative ratio for the word *hate*?