Homework 6: Reading and understanding NLP research

CS 585, UMass Amherst, Fall 2016

due Friday Dec 9^{th} 11:55pm

A note on writing and grading

Unlike other assignments, you will be graded on the quality of your writing and exposition for this assignment. Being a good researcher or engineer requires technical writing skills: you need to be able to explain your ideas to others using clear prose and standard notation. This is also great practice for your final project report, which also requires these skills. (The bottom of the project page has a few writing tips. For useful stuff here).

A note on NLP venues

You are almost done with 585! If you want to keep learning about this stuff, check out papers from leading NLP conferences like NAACL, EMNLP, ACL and the journal TACL. For more on this, see this link.

Assignment

In this assignment you will pick one research paper from the following list to read.

- Bengio et al. (2003), A Neural Probabilistic Language Model, *JMLR*. (You can skip section 3 if you want. It concerns details for an efficient, parallel implementation of the model).
- Bamman and Smith (2015), Open Extraction of Fine-Grained Political Statements, *EMNLP*. (The paper presents two different models in 3.1 and 3.2; if you like, feel free to skip the details of one of them.)

You will probably have to read the paper a few times before you really understand it. In some cases, the paper may include material that is so unfamiliar you can't really make sense of it. That's ok: it's part of the research process. Getting a rough sense of what this material is talking about is a good first step.

As you read the paper, make a list of each symbol, its mathematical type, and a short description of what the symbol means.

A "type" describes a set of values that a symbol is allowed to refer to, like in programming: you can't fill a string variable in Java with an integer. A string variable is only allowed to refer to strings. We typically talk about mathematical types likes sequences, tuples, sets, strings, discrete values, integers, real numbers, and probability distributions over them. One very important distinction is to precisely distinguish between scalars, vectors, and matrices.

For example, if you were reading J & M chapter 7 and you saw equations 7.1 to 7.5 (pdf link) you might make the following rows in your table:

symbol	meaning	type
y	a class label; e.g. spam or not spam	discrete
$\hat{y}$	a predicted class label	discrete
x	the document; type is unclear from these equations	??
$P(y)$	the prior assigned probability for class y before seeing any evidence	discrete distribution
$f_i(c, x)$	the ith feature function with two inputs: proposed class c and document x	function, outputting 0 or 1
w_i	the parameter weight for feature i	real
Z	normalizing constant	real

Sometimes papers have this sort of table in them to help the reader understand it, like Figure 1 in this paper by Brendan. It's typically most helpful when there's lots of mathematical symbols. Well-written papers are careful to define symbols before they use them (so the table is not always strictly necessary). In any case, creating such a table for yourself helps understanding.

As you're reading, it's easiest to jot some of this down on paper.

Questions

Answer the questions below. We also supply rough guidelines for expected length of answer.

1. Which paper did you pick?
2. What notation do the authors use? Answer by typing up your symbol table from above. Limit yourself to no more than 20 rows in your table; choose the most important symbols if you have too many.
3. What problem are they trying to solve? (∼2 sentences)
4. Why is that problem important? (2-5 sentences)
5. How did the authors know they were successful? For example, in your structured perceptron assignment you know the model "worked" because you could measure the how accurately it tagged a given tweet with the correct part of speech. Can you identify a problem with the authors' measure of success? Make one argument why the measure of success is appropriate and one argument why the measure of success is not appropriate. (∼10 sentences)
6. How does the paper represent language? For instance, do they use a bag or words, a bag of bigrams, a constituent parse (like you found with the CYK algorithm)? Is this representation suitable for their task? Make one argument why it is suitable and one argument why it is not suitable. (∼5 sentences)
7. Academic papers cite other academic papers, creating a chain of related knowledge. These next two questions related to cited work.

- (A) Research papers are almost always an improvement, reaction or twist on other research ("prior work") that others have done before. A paper might advance the state-of-the-art in some established task like image captioning, or apply a technique normally used in one area in a totally new way. Of all of the works cited in your paper, what seems to be the most important cited **prior work**? Explain why that citation seems most important. (∼5 sentences)
- (B) Research papers almost always make use of tools, methods and algorithms that have been developed by others. For instance, a paper on summarization might make use of general-purpose optimization techniques, which can be used for lots of things, not just creating a summary. Of all of the works cited in your paper, what citation of a **tool, method or algorithm** seems most important? Explain why. (∼5 sentences)

8. Make a list of **five** terms that you do not understand from your paper. For each, do a bit of research and write a brief description of the term. (∼2 sentences each.)

   For example, if you were reading this paper, you might see the term "linear dynamical system" and, after reading a bit on wikipedia and elsewhere, make the following bullet point.

   - **Linear dynamical systems (LDS):** A dynamical system is a model that describes the location of something in space, dependent on time (e.g. location of pendulum after t seconds). Linear dynamical systems are systems where the relationship between time and space can be described using a linear function.

9. Pick one equation from the paper and explain what it means using a mix of prose and mathematical notation. (∼15 sentences.) Imagine you are writing to a fellow CS student who has a technical background but doesn't know anything about this paper.