HW1 – Intro NLP – Due 9/17 – Benjamin Panny

Section 1.3

In our term-document matrix, the rows are word vectors of D dimensions. Do you think that's enough to represent the meaning of words?

It is not enough to represent the meaning of words, only a fragment of their meaning in the context of Shakespeare's plays, which is based on which documents they occur in. For purposes other than document identification, this is not very useful.

Which vector space (term-document or term-context) produce similar words that make more sense than others and why do you think that is the case?

Term-context seems more similar because it contains more names in the top 10 most similar words. This makes sense because names may be more likely to occur next to the same words across Shakespear's plays (if the index word is "juliet"). Terms that occur in the same documents can be very dissimilar if they do not occur close together within the documents.

Consider any decisions you made in the prior sections when implementing your functions, such as whether you allowed a target word to co-occur with itself as a context word, and which window size you chose for the term-context matrix. How might any decisions you make impact our results now?

A context_window_size of 1 degrades the usefulness of the term-context matrix. A more realistic context_window_size, that is closer to the size of a predicate (half a sentence or more), will probably capture more information about the context of any particular word, making the top-10 similar words appear more intuitively similar.

Section 1.4

Redo the analysis in section 1.3 with ranked words and compare using tf-idf and PPMI with unweighted term-document and term-context matrices.

The results between the term-document and tf-idf matrices are more similar than the results between term-context and PPMI matrices.

Discuss findings from comparing approaches. Do some approaches appear to work better than others, i.e produce better synonyms? Do any interesting patterns emerge?

This is difficult to tell with a context_window_size of 1, but using the index word "enemy" did seem to yield some synonyms in PPMI, as did the word "fall", yielding "befall" and "fallen". A pattern I did not observe in any other approaches in my manual sweep over some index words.

How does weighting with tf-idf compare to using the unweighted term-document matrix? How does weighting with PPMI compare with using the unweighted term-context matrix? How does term-context/PPMI compare to term-document/TF-IDF? Include results and discussion.

The results between the term-document and tf-idf matrices are more similar than the results between term-context and PPMI matrices.

The 10 most similar words to "juliet" using cosine-similarity on term-document frequency matrix are:

1: procures; 0.9611235625180706

2: benedicite; 0.9611235625180706

3: ghostly; 0.9380710616293597

4: capulets; 0.8748623439660793

5: mercutio; 0.8748623439660793

6: capulet; 0.8748623439660792

7: pump; 0.8748623439660792

8: laura; 0.8748623439660792

9: pitcher; 0.8748623439660792

10: behoveful; 0.8748623439660792

The 10 most similar words to "juliet" using cosine-similarity on tf-idf matrix are:

1: procures; 0.9611235625180706

2: benedicite; 0.9611235625180706

3: ghostly; 0.9380710616293598

4: pump; 0.8748623439660792

5: mercutio; 0.8748623439660792

6: households; 0.8748623439660792

7: blubbering; 0.8748623439660792

8: stinted; 0.8748623439660792

9: duellist; 0.8748623439660792

10: switch; 0.8748623439660792

The 10 most similar words to "juliet" using cosine-similarity on term-context frequency matrix are:

1: am; 1

2: warwick; 0.7863009277715228

3: maria; 0.7813181188793948

4: helena; 0.7788038614889767

5: cloten; 0.7687813263801225

6: clown; 0.7673992667724928

7: touchstone; 0.7671713498011156

8: gloucester; 0.7663929420831365

9: jaquenetta; 0.7625294119579562

10: diomedes; 0.7592329642096558

The 10 most similar words to "juliet" using cosine-similarity on PPMI matrix are:

1: throe; 0.1784176591434079

2: hist; 0.1760453035422327

3: ursula; 0.15724404397747238

4: silvia; 0.15320819864340685

5: benumbed; 0.15156266689461984

6: pe; 0.14629131220181457

7: unrecuring; 0.14621967997890617

8: opposeless; 0.146182789702897

9: banditti; 0.1460769309986898

10: inseparable; 0.14349439839154998

Part 2

With that PPMI-weighted term-context matrix, find the vectors for identity labels in the provided list. Look at the top associated words (by PPMI) for at least 4 identity labels of your choice. Do you see any that may reflect social stereotypes? It is helpful to compare the top PMI words for certain identity terms with other related ones (such as men compared with women). Discuss and provide selected results in the report.

I chose the words “woman”, “black”, “mental”, and “disabled”

“Woman” does not appear to reflect any social stereotypes, though “dress” is in the top 15 PPMI. neither does “black”. “black” seems largely associated with other colors and articles of clothing. “Mental” is associated with what is intuitively associated with “mental”, such as emotions and clinical terms such as psychiatry and experimental. “Disabled” has associations with associated terms, such as “disabilities”, “amputees”, and “strengths”

Comparing “opposite” labels didn’t help much.

Qualitative analysis: For at least 4 different identity labels, dig into the contexts that leads to high PMI association with other words, especially for any words that show social bias if you found that. 1st-order similarity: find specific examples from the dataset where an identity label occurs with a top-associated term that shows some social bias or does not. This might not occur; if not, you can look at

2nd-order similarity in which the two words occur with similar context words. Sample the contexts/documents in which the 2 words occur separately. Do they occur with the same set of context words? This can also be examined by looking at the vectors for the identity term and the highly associated other term in the term-context matrix. Do these share high values in certain dimensions that correspond to certain context words? Provide selected results and discuss findings in the report. Do you see evidence for representational harms (see below) learned by a bag-of-words model of this SNLI corpus? If so, which type do you see? Provide examples that support your conclusions. If you don't find any potential harms, provide examples of what you examined and how you interpreted those associations.

"Black" very often occurs with colors as an adjective for objects, pets, or articles of clothing. As such, articles of clothing is similar to "black" because articles of clothing will often have other contexts with different colors, as the word "black" does. I also see sentences with "woman" and "men" describing people wearing kinds of clothes, explaining dress being associated with "woman". Looking at "dress", it mostly occurs in sentences with "girl" and woman". The only sentences with the word "amputee" contain male or neutral pronouns. This could possibly bias a representation to "represent" that women can't be amputees.