Progress Report

Martin Pettersson, Natural Science DEEP
2014/2015

The Winograd Schema Challenge

- Instances of difficult pronoun resolution problems

- Requires understanding of implicit information such as belief, desire or intentions of characters

Provides *Twin Sentences*:

Lions hunt zebras because they are hungry.
Lions hunt zebras because they are delicious.

Intuition and Motivation

- Currently there is a set of WSC problems that are especially difficult to resolve due to sentiment issues.
- Most of them have some sort of conditional statement.
- It would be nice to explore models that may or may not improve the performance on these sentences.

10   Fox News is more popular in ratings than CNN since they are more interesting.

11   Fox News is more popular in ratings than CNN since they are boring.

74   Sega lost to Nintendo because they were superior.

75   Sega lost to Nintendo because they were inferior.

158   Obama beat John McCain, because he was the better candidate.

2   Spiderman defeated Magnito because he is a good guy.

3   Spiderman defeated Magnito because he is a bad guy.

8   Federer consistently beat Nadal since he was the better tennis player.

9   Federer consistently beat Nadal since he was the worse tennis player.

16   UPS provides much convenience to the customers with their drop off service since they make the shipping of packages extremely easy.

Intuition and Motivation

Sega lost to Nintendo because they were superior.

Sega lost to Nintendo because they were inferior.
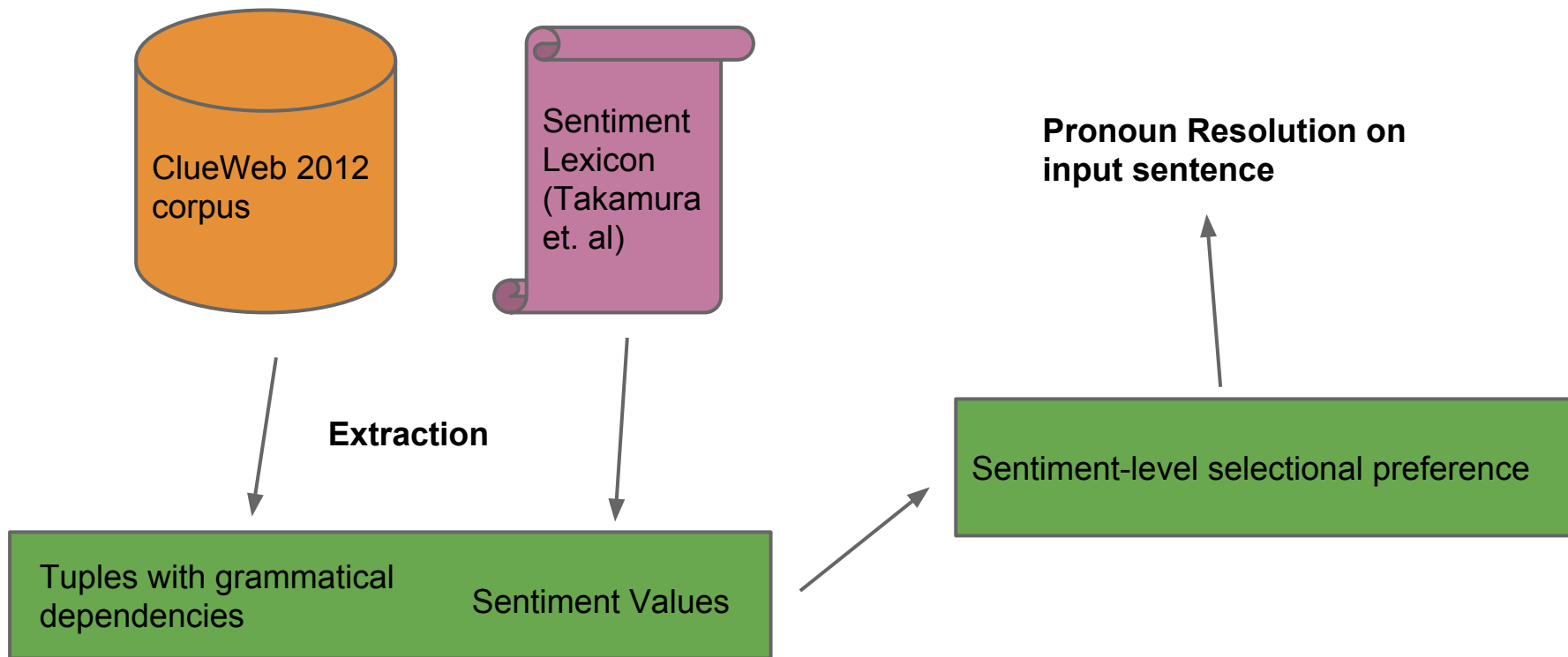
Intuition and Motivation

Sega **lost to** Nintendo because **they** were **superior**.

Sega **lost to** Nintendo because **they** were **inferior**.

Possible interesting research questions

- Can the emotional/sentimental value of words in a sentence determine who or what is talked about/referred to?
- Is there a relationship between sentimental value of grammatical dependencies and the sentimental value of a pronoun?
- *Follow up question*: If there is, can this knowledge be used to improve an existing coreference/pronoun resolution system?

A First Simple Solution

ClueWeb 2012 corpus

Sentiment Lexicon (Takamura et. al)

Pronoun Resolution on input sentence

Extraction

Tuples with grammatical dependencies

Sentiment Values

Sentiment-level selectional preference

Knowledge Extraction

- Collapsed Stanford dependencies from ClueWeb, made a text file for each dependency. All tuples were lemmatized.
- Sentiment Values from the *Takamura et. al* sentiment lexicon.
- Constructed a file where each tuple got linked with its corresponding sentiment values.

Sentiment-level Selectional Preference

- From the extracted knowledge we are now able to perform the following query:

  sp(governor, dependant, dependency)

- *Example:*

sp(defeat, *, nsubj) will return all sentiment values of the dependants of the verb *defeat* with dependency *nsubj*.

Pronoun Resolution

1. Input sentence is parsed using the Stanford CoreNLP collapsed dependency parser.
2. The Sentiment-level selectional preference model will run on each antecedent candidate.
3. The results of each antecedent candidate are compared to that of the pronoun.

Step-by-step explanation of a positive example

Input:

Obama beat **John McCain**, because [he]+ was the better candidate

Step-by-step explanation of a positive example

Input:

Obama beat **John McCain**, because [he]+ was the better candidate

Parse and lemmatization:

nsubj(beat,obama), dobj(beat,mccain)

Step-by-step explanation of a positive example

Input:

Obama beat **John McCain**, because [he]+ was the better candidate

Parse and lemmatization:

nsubj(beat,obama), dobj(beat,mccain)

Sentiment-level selectional preference:

*sp(*beat,*,nsubj) → 10   *sp*(beat,*,dobj) → -8

Step-by-step explanation of a positive example

Input:

Obama beat **John McCain**, because [he]+ was the better candidate

Parse and lemmatization:

nsubj(beat,obama), dobj(beat,mccain)

Sentiment-level selectional preference:

*sp(*beat,*,nsubj) → 10   *sp*(beat,*,dobj) → -8

Comparison with pronoun sentiment:

Obama is chosen because of the positive value, which is correct.

...and a negative example

Input:

Sega lost to **Nintendo**, because [they]- were inferior.

Parse and lemmatization:

nsubj(lose,sega), prep_to(lose,nintendo)

Sentiment-level selectional preference:

*sp(*lose,*,nsubj) → 79   *sp*(lose,*,prep_to) → -14

Comparison with pronoun sentiment:

Nintendo is chosen because of the positive value, which is incorrect.

How the values are calculated

*sp(*beat,*,nsubj) → 10

How the values are calculated

*sp(*beat,*,nsubj) → **10** , sample output:**

       beat chef vbd nn 0.00347406

       beat black vbd nnp -0.0495369

       beat oklahoma vbd nnp 0.00230084

       beat maker vb nn 0.0169773

       beat soldier vbp nns -0.0305791

       beat reading vbd nns 0.0296309

       ...

Result of the Tuple Extraction

Dependency	dobj	nsubj	prep_after	prep_for	prep_in	prep_of	prep_than	prep_to	prep_with	prepc_to
ClueWeb tuples	212484	339505	1698	36323	49373	71844	1613	29711	24382	820
Tuples With Sentiment	130529	89340	857	24770	37130	53565	630	19294	16452	227

Experimental Setup

- *Test set:* 28 Sentences from WSC devset200 that have sentiment issues.
- Annotated the pronouns manually with a sentiment value. (just + or -).
- Parsed the sentences using Stanford CoreNLP collapsed dependecies parser.
- I ran the sentiment-level selectional preference model on the dependencies of the candidate antecedents and compared the values with my annotations.

Experimental Results

Correct: 15

Incorrect: 6

Non-resolvable: 7

Results from Error Analysis

- Stanford CoreNLP does not lemmatize comparative adjectives. This was an issue in 5 of the non-resolvable cases.
- Two cases of parsing errors. (dependencies missing)
- There is no error handling for negations. If we reverse sentiment polarity for words like *not* or *but* the results would probably be better.
- In some cases the sentiment values look strange.
- One case of a world-knowledge error.

Interesting World Knowledge Error

Input:

[The Simpsons]+ have lasted 23 seasons while The Sarah Connor Chronicles lasted two seasons, since [they]+ have been wildly successful

The Simpsons have lasted 23 seasons while [The Sarah Connor Chronicles]- lasted two seasons, since [they]- did not provide a very good show

- Only resolvable by the knowledge that 23 seasons are better than 2 seasons (for televisions series).

Critique of the Work

- Not much related work to compare this project to.
- There are a lot of papers on sentiment analysis and pronoun resolution, but none of these mention sentiment-level selectional preference.
- Only one sentiment lexicon used.
- Only a subset of ClueWeb was used (~180MB)
- No real heuristic has been used to determine the sentiment of the pronoun. I only used my own manual annotations (and they are somewhat subjective).

Future Plans

- Implement a heuristic for determining the sentiment value of the pronoun.
- Experiment with different sentiment lexicons (Warriner et al, SentiWordNet)
- Make an automatic rule-based system and compare it to a classifier-based approach.
- Compare the results to an existing coreference resolution system and give some convincing answers to my research questions.

Questions!