Negation Detection using NooJ

Gaurish Thakkar, Nives Mikelic Preradovic, Jeremy Barnes

Faculty of Humanities and Social Sciences in Zagreb
Croatia

Language Technology Group, University of Oslo
Norway

The project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 812997.
Outline

- Task of Negation Detection
- How Did We Solve It
- What Were the Results
- Conclusion & Future Work
Negation Detection

• Negation can be implicit, as in “with this act, it will be his first and last movie”—it carries a negative sentiment, but no negative words are used.

• Negation can be explicit, as in “this is not good.”

• Negation can be morphological where it is either denoted by a prefix (“dis-”, “non-”) or a suffix (“-less”).
  • "It seems a singularly useless thing to steal," said Sherlock Holmes.
  • I was not sure whether I had left it here or in the Shipping Office.
Problem Statement

• Given a sentence $S = \{ t_0 \ldots t_n \}$ where $t_i$ is token, the aim is to identify negation-cue $N_c = \{ t_i \ldots t_k \}$ and negation-scope $N_s = \{ t_j \ldots t_l \}$.

• "Come, come, we are not so far wrong, after all, " said Holmes.
  • ("",'O'),('Come','O'),('','O'),('come','O'),('','O'),('we','B_scope'),('are','I_scope')
    ,("not","B_cue"),('so','B_scope'),('far','I_scope'),('wrong','I_scope'),('','O'),('after','O'),('all','O'),('','O'),('""','O'),('said','O'),('Holmes','O'),('.','O')
Related Work

• Koza W. et al. (2018)
  • Negated findings in radiological reports. (medical terminology dictionary+ grammar rules).

• Tanushi, Hideyuki, et al. (2013)
  • Compared 3 different approaches for negation detection namely NegEx, PyConTextNLP and SynNeg.

• Chapman et al. (2001)
  • Negex regular expression-based algorithm.
### Dataset

- The Conan Doyle neg-corpus

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>3644</td>
<td>787</td>
<td>1089</td>
</tr>
<tr>
<td><strong>Negated</strong></td>
<td>842</td>
<td>144</td>
<td>235</td>
</tr>
</tbody>
</table>

Number of sentences
Experimental Setup

• Low-resourced setting
  • We do not use the annotated train set for supervision.

• Use dev set for tuning and test for computing final score.

• Unannotated version of train set is tagged via NooJ grammar is used for training subsequent systems.
Overall Diagram

NooJ - Simple Grammar
- Detect negated sentences
- Detect neg-cues
- Detected neg-scope

Generative Model - Hidden Markov Model
- Linked Hidden Markov Model
- Naïve Bayes

Discriminative Model - BiLSTM
We study a patient who became quadriplegic after parenteral magnesium administration for pre eclampsia.
Models

(a) Classical HMM

(b) Linked HMM

Naive Bayes

x₁

x₂

x₃

sentence

word

vec

x₁

x₂

x₃

x₄

look-up layer

forward LSTM

h₁

vec

h₁

vec

h₂

vec

h₂

vec

h₃

vec

h₃

vec

h₄

vec

h₄

BiLSTM

CRF layer

tag

P₁

tag

P₂

tag

P₃

tag

P₄

LSTM output
Step 1: Nooj Grammar
Step 2: Labelling Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Token Acc.</th>
<th>Token Votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>BigramNegationCueNT</td>
<td>21</td>
<td>17</td>
<td>406</td>
<td>1.0000</td>
<td>38</td>
</tr>
<tr>
<td>CommonFalsePositives</td>
<td>0</td>
<td>0</td>
<td>427</td>
<td>1.0000</td>
<td>10</td>
</tr>
<tr>
<td>CommonTruePositivesImplicitSCue</td>
<td>140</td>
<td>5</td>
<td>287</td>
<td>0.9793</td>
<td>145</td>
</tr>
<tr>
<td>CommonTruePositivesImplicitScope</td>
<td>1</td>
<td>10</td>
<td>426</td>
<td>1.0000</td>
<td>11</td>
</tr>
<tr>
<td>CueConstituencyKeywords</td>
<td>254</td>
<td>158</td>
<td>173</td>
<td>0.7948</td>
<td>1272</td>
</tr>
<tr>
<td>Non-EntityPunctuation</td>
<td>0</td>
<td>0</td>
<td>427</td>
<td>0.9511</td>
<td>348</td>
</tr>
<tr>
<td>NoojLabels</td>
<td>276</td>
<td>88</td>
<td>151</td>
<td>0.8033</td>
<td>2278</td>
</tr>
<tr>
<td>TrigramNegationCue</td>
<td>1</td>
<td>0</td>
<td>426</td>
<td>1.0000</td>
<td>3</td>
</tr>
</tbody>
</table>
Results
1. How many lines were detected using the simple Nooj Grammar?

<table>
<thead>
<tr>
<th></th>
<th>Detected</th>
<th>Total</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>700</td>
<td>842</td>
<td>0.83</td>
</tr>
<tr>
<td>Dev</td>
<td>113</td>
<td>144</td>
<td>0.78</td>
</tr>
<tr>
<td>Test</td>
<td>202</td>
<td>235</td>
<td>0.85</td>
</tr>
</tbody>
</table>
2. What is the Precision, Recall and F1-score on applying the grammar for detecting the negation cue and scope?

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>0.76</td>
<td>0.74</td>
<td>0.75</td>
<td>0.81</td>
</tr>
<tr>
<td>Dev</td>
<td>0.76</td>
<td>0.73</td>
<td>0.74</td>
<td>0.77</td>
</tr>
<tr>
<td>Test</td>
<td>0.75</td>
<td>0.73</td>
<td>0.74</td>
<td>0.82</td>
</tr>
</tbody>
</table>

*All score are Micro Avg*
### Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NooJ</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NooJ+LinkedHMM+BiLSTM</td>
<td>0.75</td>
<td>0.68</td>
<td>0.71</td>
</tr>
<tr>
<td>NooJ+LF+LinkedHMM+BiLSTM</td>
<td>0.76</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td>NooJ+LF+HMM+BiLSTM</td>
<td>0.75</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td>NooJ+LF+NB+BiLSTM</td>
<td>0.53</td>
<td>0.74</td>
<td>0.62</td>
</tr>
<tr>
<td>LF+LinkedHMM+BiLSTM</td>
<td>0.55</td>
<td>0.71</td>
<td>0.62</td>
</tr>
</tbody>
</table>

**Step 2: Generative Model**

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NooJ</td>
<td>0.75</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td>NooJ+LinkedHMM+BiLSTM</td>
<td>0.75</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td>NooJ+LF+LinkedHMM+BiLSTM</td>
<td>0.76</td>
<td>0.75</td>
<td>0.76</td>
</tr>
<tr>
<td>NooJ+LF+HMM+BiLSTM</td>
<td>0.75</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td>NooJ+LF+NB+BiLSTM</td>
<td>0.67</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>LF+LinkedHMM+BiLSTM</td>
<td>0.50</td>
<td>0.74</td>
<td>0.60</td>
</tr>
</tbody>
</table>

**Step 3: Discriminative Model**
Error Analysis

• A lot of false positives are introduced e.g. but, although

• Current setup fails to capture complex sentences
  • # Mr. Sherlock Holmes, who was usually very late in the mornings, save upon those not infrequent occasions when he was up all night, was seated at the breakfast table.
  • # ``Now, tell me, Dr. Mortimer -- and this is important -- the marks which you saw were on the path and not on the grass?"
Conclusion & Future Work

• Simple grammar graphs do perform well.
• NooJ can be leveraged as a labelling function for training NLP systems.
• Improve grammar
  • Handle false positives
  • to handle long-range dependencies.
• Repeat the experiment with other languages with low-resources.
• Utilize the setup for other NLP tasks.
References


Questions ?