Story Cloze Task: UW NLP System

Roy Schwartz, Maarten Sap, Yannis Konstas, Leila Zilles, Yejin Choi and Noah A. Smith
LSDSem 2017
Outline

• System overview
  • Language modeling
  • Writing style

• Results

• Discussion
Background

<table>
<thead>
<tr>
<th>Story Prefix</th>
<th>Endings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe went to college for art. He graduated with a degree in painting.</td>
<td>Then he got hired.</td>
</tr>
<tr>
<td>He couldn't find a job. He then responded to an ad in the paper.</td>
<td>Joe hated pizza.</td>
</tr>
</tbody>
</table>
Approach 1: Language Modeling

\[ e^* = \text{argmax}_{e \in \{e_1, e_2\}} p_{lm}(e | \text{prefix}) \]
Approach 1.1: Language Modeling\

\[ e^* = \arg\max_{e \in \{e_1, e_2\}} \frac{p_{lm}(e|\text{prefix})}{p_{lm}(e)} \]
Approach 2.0: Style

- Intuition: authors use different **style** when asked to write *right* vs. *wrong* story ending

- We train a style-based classifier to make this distinction

- Features are computed using **story endings only**
  - Without considering the story prefix
Combined Model

• A logistic regression classifier

• Features:
  
  • LM features: $p_{lm}(e|\text{prefix})$, $p_{lm}(e)$, $\frac{p_{lm}(e|\text{prefix})}{p_{lm}(e)}$
    
    • An LSTM RNNLM trained on the ROC story corpus

  • Style features: sentence length, character 4-grams, word 1-5-grams
    
    • Features computed without access to the story prefixes

• Model is trained and tuned on the story cloze development set
Results

\[ p_{lm}(e|\text{prefix}) \]

DSSM

LexVec

\[ \frac{p_{lm}(e|\text{prefix})}{p_{lm}(e)} \]

Style

LM + Style

Story Cloze Task: UW NLP System @ Schwartz et al.
Our LM expression is proportional to pointwise mutual information:

\[
\log \left( \frac{p(e|\text{prefix})}{p(e)} \right) = \log \left( \frac{p(e, \text{prefix})}{p(e)p(\text{prefix})} \right) = PMI(e, \text{prefix})
\]
### Most Heavily Weighted Style Features

<table>
<thead>
<tr>
<th></th>
<th>Freq.</th>
<th></th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Right</strong></td>
<td></td>
<td><strong>Wrong</strong></td>
<td></td>
</tr>
<tr>
<td>‘ed .’</td>
<td>6.5%</td>
<td>START NNP</td>
<td>54.8%</td>
</tr>
<tr>
<td>‘and ’</td>
<td>13.6%</td>
<td>NN .</td>
<td>47.5%</td>
</tr>
<tr>
<td>JJ</td>
<td>45.8%</td>
<td>NN NN .</td>
<td>5.1%</td>
</tr>
<tr>
<td>to VB</td>
<td>20.1%</td>
<td>VBG</td>
<td>10.1%</td>
</tr>
<tr>
<td>‘d th’</td>
<td>10.9%</td>
<td>START NNP VBD</td>
<td>41.9%</td>
</tr>
</tbody>
</table>
Discussion: Style

• different writing tasks  different writing style
  (mental state?)

• Common sense induction is hard
  • What are our models learning?
  • It is important to reach the ceiling of simple “dumb” approaches
  • The added value of our RNNLM indicates that it is learning something beyond shallow features

• Schwartz et al., 2017, The Effect of Different Writing Tasks on Linguistic Style: A Case Study of the ROC Story Cloze Task
Summary

- \[ p_{lm}(e | \text{prefix}) \]
- \[ p_{lm}(e) \]

• Style features that ignore the story prefix get large performance gains

• A combined approach yields new state-of-the-art results – 75.2%

Thank you!

Roy Schwartz  roysch@cs.washington.edu  http://homes.cs.washington.edu/~roysch/