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

Story Prefix	Endings
Joe went to college for art. He graduated with a degree in painting. He couldn't find a job. He then responded to an ad in the paper.	Then he got hired.
	Joe hated pizza.

Approach 1: Language Modeling

$$e^* = \underset{e \in \{e_1, e_2\}}{\operatorname{argmax}} \, p_{lm}(e|\text{prefix})$$

Approach 1.1: Language Modeling⁺

$$e^* = \underset{e \in \{e_1, e_2\}}{\operatorname{argmax}} \frac{p_{lm}(e|\text{prefix})}{\boldsymbol{p_{lm}(e)}}$$

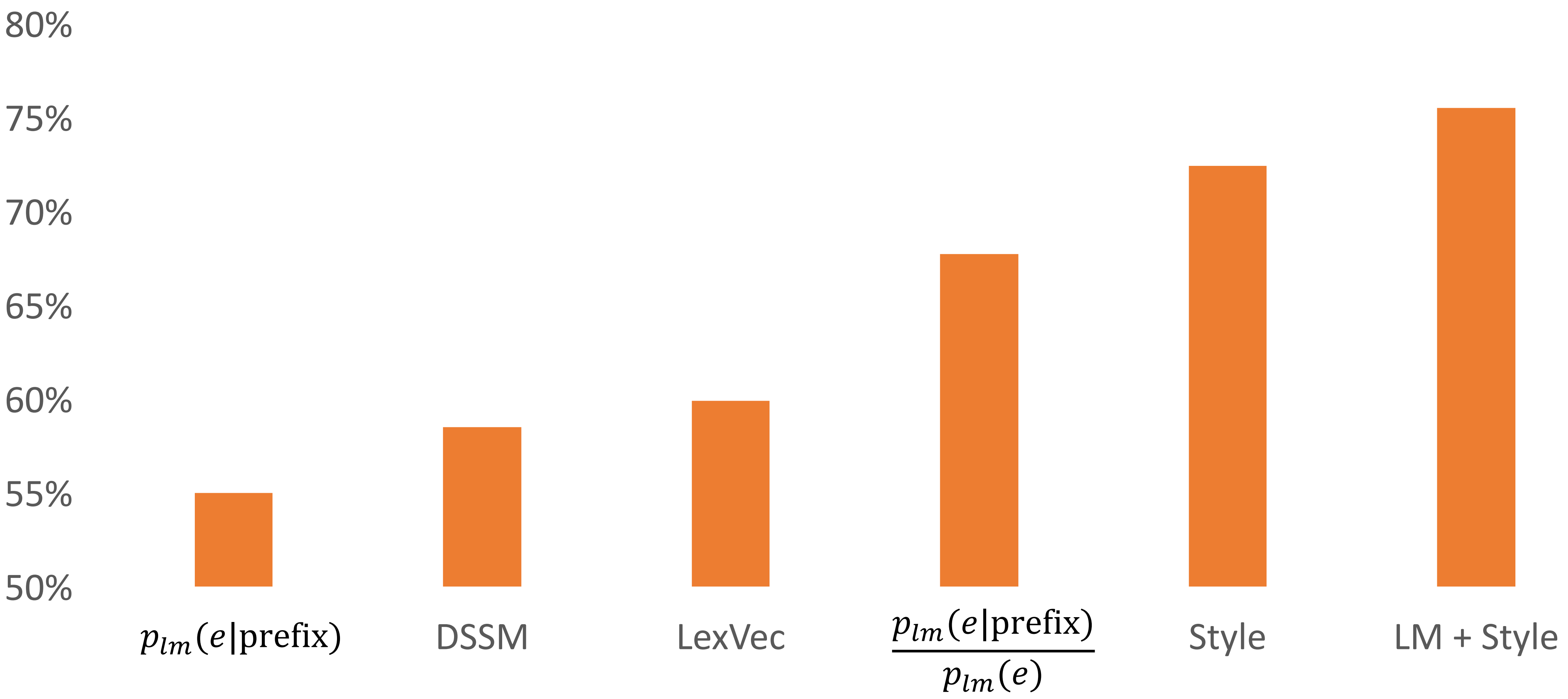
Approach 2.0: Style

- 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



80%
75%
70%
65%
60%
55%
50%
$p_{lm}(e|\text{prefix})$
DSSM
LexVec
$\frac{p_{lm}(e|\text{prefix})}{p_{lm}(e)}$
Style
LM + Style

Discussion: Language Modeling+

- 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

Right	Freq.	***Wrong***	Freq.
‘ed .’	6.5%	START NNP	54.8%
‘and ’	13.6%	NN .	47.5%
JJ	45.8%	NN NN .	5.1%
to VB	20.1%	VBG	10.1%
‘d th’	10.9%	START NNP VBD	41.9%

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

- $\frac{p_{lm}(e|\text{prefix})}{\boldsymbol{p_{lm}(e)}}$

- 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/