Natural Language Processing 1

Lecture 2: Language models and part-of-speech tagging

Katia Shutova

ILLC
University of Amsterdam

28 October 2020

Outline.

Question 1: Ambiguity

What kinds of ambiguity manifest themselves in the following newspaper headlines?

Iraqi head seeks arms

Stolen painting found by tree

Teacher strikes idle kids

Question 2: Language modelling – perplexity

- ▶ Wall Street Journal corpus
- ▶ Train on 38 million words
- ▶ test on 1.5 million words

N-gram Order	Unigram	Bigram	Trigram
Perplexity	962	170	109

Question 2: Language modelling – perplexity

N-gram Order	Unigram	Bigram	Trigram
Perplexity	962	170	109

The table shows a clear inverse correlation between *n*-gram order and perplexity.

- ▶ Why does perplexity decrease as *n* increases?
- ▶ Do you expect this decrease in perplexity to continue with higher values of *n*? Why?

Question(s) 3: Discussion

1. What sources of information can you use to PoS-tag unknown words?

2. How are n-gram language modelling and PoS tagging methods affected by the differences in genre? Which of the two tasks is more sensitive to this difference and why?

3. For the challenges and applications listed in lecture 1, which problems can n-gram language modelling and PoS tagging solve?

Question 4: PoS tagging

NN1	singular noun	NN2	plural noun
PNP	personal pronoun	VM0	modal auxiliary verb
VVB	base form of verb	VVI	infinitive form of verb

Two possible analyses of *They can fish.*

- They_PNP can_VM0 fish_VVI ._PUN
- They_PNP can_VVB fish_NN2 ._PUN

List the probabilities involved in each of the analyses. Which of the analysis will be most likely chosen, and why? (Assuming the model was trained on a balanced corpus.)

Question 5: Examples of PoS tagging errors

Turkey will keep for several days in a fridge.

Turkey_NP0 will_VM0 keep_VVI for_PRP several_DT0
days_NN2 in_PRP a_AT0 fridge_NN1

We have hope that the next year will be peaceful.

We_PNP have_VHB hope_VVB that_CJT the_AT0 next_ORD
year_NN1 will_VM0 be_VBI peaceful_AJ0

Why does the model make these errors?

Question 5: Examples of PoS tagging errors

Turkey will keep for several days in a fridge.

Turkey_NP0 will_VM0 keep_VVI for_PRP several_DT0
days_NN2 in_PRP a_AT0 fridge_NN1

We have hope that the next year will be peaceful.

We_PNP have_VHB hope_VVB that_CJT the_AT0 next_ORD
year_NN1 will_VM0 be_VBI peaceful_AJ0

Why does the model make these errors?

Acknowledgement

Some slides were adapted from Ann Copestake, Dan Jurafsky and Tejaswini Deoskar