WORD-SENSE DISAMBIGUATION
WITH SELECTIONAL PREFERENCE

Yoonjung Choi (yjchoi@cs.pitt.edu)
Introduction

- A word can have one or more meaning.
  - E.g., *carry*
    - S: (v) carry (win in an election) "The senator carried his home state"
    - S: (v) carry (keep up with financial support) "The Federal Government carried the province for many years"
    - S: (v) carry (capture after a fight) "The troops carried the town after a brief fight"
    - ...

- **Word-sense Disambiguation (WSD)** is a problem to identify which sense (i.e. meaning) of a word is used in a sentence.
Introduction

- Predicates often have a preference for particular arguments.

  *The vocalist* *sings* *a ballad.*

  vs.

  *The exception* *sings* *a tomato.*

- This inclination of predicates to select for particular arguments is known as *selectional preference.*
Method

- Latent Dirichlet Allocation (LDA)
  - A generative model that discovers similarities in data using latent variables.

- Intuition: We assume that each sense is associated with a distribution over semantic classes (“topics”) and these classes are shared across senses.

- Modeling document-term co-occurrence
  → Modeling sense-argument co-occurrence
Method

• The model produces arguments as follows:
  1. For each verb sense $s$, draw a multinomial distribution $\Theta_s$ over argument classes from a Dirichlet distribution with parameters $\alpha$.
  2. For each argument class $z$, draw a multinomial distribution $\Phi_z$ over argument from a Dirichlet distribution with parameters $\beta$.
  3. To generate an argument for $s$, draw an argument class $z$ from $\Theta_s$ and then draw an argument $n$ from $\Phi_z$. 
Why LDA?

- LDA naturally consider the class-based selectional preference without a pre-defined set of classes.

- LDA naturally handle ambiguous arguments since they are able to assign different topics to the same phrase in different contexts.

- Once a topic distribution has been learned over a set of training relations, one can efficiently apply inference to unseen relations.
Data

- SensEval
  - The purpose of SensEval is to evaluate the strengths and weaknesses of WSD program with respect to different words, different varieties of language, and different languages.

- SensEval-2
  - TrainData contains 3,794 instances.
    - # terms : 131
    - # senses: 452
  - TestData contains 1,908 instances.
    - # terms: 101
    - # senses: 353
Data

• Sample Data

<instance id="promise.0528">
<answer instance="promise.0528" senseid="537584"/>
<context>

Before she had a chance to answer he had gone into the fields. The forecasts <head>promised</head> several days of hot weather and because he had help in the house Moran decided to cut all the meadows.

</context>
</instance>
Experiment

• **MALLET**
  • It provides efficient, sampling-based implementations of LDA. Also, it includes an extremely fast and highly scalable implementation of Gibbs sampling.

• For each term, the different model is built.
  • # topics: 100
  • # iterations: 1,000
Results

• Measurement
  • Recall: percentage of right answers on all instances in the test set.
  • Precision: percentage of right answers in the set of answered instances.

• Results of participants in SensEval-2
  • F-score Range: 0.141~0.642
  • Average F-score: 0.434

• Our Results
  • Precision, Recall, and F-score: 0.461
Conclusion

- LDA has several advantages for selection preference system.

- Even though our model didn’t show the best performance, it is better than the average.

- It shows that considering selectional preference is helpful to WSD system.

- If we combine other features/models, we can get better WSD performance.
Reference

THANK YOU 😊