Dependency Parsing (continued)
Lecture 20: November 15, 2013

CS886-2 Natural Language Understanding

University of Waterloo

Dependency Parser

- Training (parameter estimation)
  1. Estimate parameters of scoring function
  2. Estimate parameters of sequence tagger
- Testing (parsing)
  1. Compute edge weights for a sentence
  2. Find maximum spanning tree
  3. Find best labels for each edge

Scoring Function Estimation

- Estimate parameters of scoring function based on labeled corpus
- Probabilistic scoring function:

$$\Pr(edges|words,\lambda) \propto \exp\left(\sum_i \lambda_i f_i(words, edges)\right)$$

where f_i is a feature and λ_i is its weight

- Maximum conditional likelihood:

$$\lambda^* = argmax_\lambda \Pr(edges|words,\lambda)$$

Scoring Function Features

- Feature: any function that can be computed based on the text
- Presence/absence of an edge
  - Pair of words
  - Words between the pair of words
  - Pair of part-of-speech tags
  - Distance between two words
  - Gender, number, person attributes
  - Etc.

Sequence Tagger Estimation

- Estimate parameters of edge tags based on labeled corpus
- Probabilistic scoring function:

$$\Pr(labels|words, edges, \gamma) \propto \exp\left(\sum_i \gamma_i f_i(words, edges, labels)\right)$$

where f_i is a feature and γ_i is its weight

- Maximum conditional likelihood:

$$\gamma^* = argmax_\gamma \Pr(labels|words, edges, \gamma)$$

Sequence Tagger Model

- Tree-based conditional random field

Sequence Tagger Features

- Feature: any function that can be computed based on the text and the graph
- Single label
  - Pair of words linked by the corresponding edge
  - Words between the pair of words
  - Pair of part-of-speech tags
- Pair of labels

Dependency Parser

- Training (parameter estimation)
  1. Estimate parameters of scoring function
  2. Estimate parameters of sequence tagger
- Testing (parsing)
  1. Compute edge weights for a sentence
  2. Find maximum spanning tree
  3. Find best labels for each edge

Edge Weights

- To find an MST, we need a weight for each edge

$$w(edge) = \sum_{\{i|f_i \; depends \; on \; edge\}} \lambda_i f_i(edge, words)$$

Maximum Spanning Tree (MST)

- MST finds the tree that maximizes $\sum_{edge} w(edge)$
- Scoring function in terms of edge weights:

$$\Pr(edges|words) \propto \exp\left(\sum_{edge} w(edge)\right)$$

- Hence, MST finds tree that maximizes the scoring fn

Edge Labeling

- Compute best labeling
  $labels^* = argmax_{labels} \Pr(labels | edges, words, \gamma^*)$
- Algorithms:
  - Viterbi (dynamic programming)
    - First order model
  - Gibbs sampling
    - Long range dependencies