

Relation Detection and Classification

Lecture 17: November 6, 2013

CS886-2 Natural Language Understanding
University of Waterloo

CS886 Lecture Slides (c) 2013 P. Poupart

1

Relations

- **Relation**: a way in which two or more entities are connected
- Examples
 - “Born in”:
 - Shakespeare was born in Stratford-upon-Avon
 - Stratford-upon-Avon is the city in which Shakespeare was born
 - “mother of”:
 - Janes is the mother of John
 - John’s mother is Jane
 - “supplier of”:
 - Automodular is a supplier of Ford
 - Automodular supplies auto parts to Ford
 - “cure of”:
 - Clarithromycin can cure certain types of ulcer

CS886 Lecture Slides (c) 2013 P. Poupart

2

Formalism

- Relation: **set of entity tuples**
- Example:
 - Consider a set of N companies $\{C_1, C_2, C_3, \dots, C_n\}$
 - Relation: “supplier of”
 - $\{ \langle C_1, C_4 \rangle, \langle C_2, C_4 \rangle, \langle C_3, C_5 \rangle \}$
- Entity types are unary relations
 - E.g. Location, person, organization, protein, disease

CS886 Lecture Slides (c) 2013 P. Poupart

3

Sentiment/Opinion

- Opinion: $\langle o, f, s, h, t \rangle$, where
 - Where o (object), f (feature of object), h (opinion holder), t (time), s (sentiment: +ve, -ve, or neutral)
- Three relations: +ve, -ve, neutral
- Tuples of 4 entities: $\langle o, f, h, t \rangle$

CS886 Lecture Slides (c) 2013 P. Poupart

4

Notes

- The same tuple may be part of different sets that define different relations
 - i.e., relations may have overlapping sets
- Different relations will generally be denoted by different sets, however they have the same set of tuples if the domain contains few entities
 - E.g. “born in” and “lives in” may be denoted by the same sets if all people are children that haven’t moved away from their place of birth

CS886 Lecture Slides (c) 2013 P. Poupart

5

Relation Detection and Classification

- Task: Identify tuples of entities in a document that are characterized by a certain relation
- Common Approaches
 - Supervised Learning
 - Bootstrapping

CS886 Lecture Slides (c) 2013 P. Poupart

6

Supervised Learning

- Data: annotated corpus
 - Entities already extracted
 - Tuples of entities characterized by the desired relation are already marked
- Train classifier that returns true or false for any candidate tuple
 - E.g., decision tree, support vector machine, naïve Bayes model, logistic regression, etc.

CS886 Lecture Slides (c) 2013 P. Poupart

7

Features

- Entity features
 - Entity types
 - Bag-of-words of entity mentions
 - Head word of each entity
- Relation features
 - Bag-of-words of text between entity mentions
 - # of words between the entity mentions
 - # of entities between the entity mentions
 - Syntactic chunk of words between entity mentions
 - Features of the dependency path between entities

CS886 Lecture Slides (c) 2013 P. Poupart

8

Bootstrapping

- Idea: start with a small set of data
 - Set of phrase templates
 - Set of entity tuples
- Alternate between two steps
 - Discover **new entity tuples** that fill existing templates
 - Discover **new templates** that embed existing entity tuples

CS886 Lecture Slides (c) 2013 P. Poupart

9

Semantic Drift

- When bootstrapping makes errors in one step it can lead to more errors in the next step and so on
 - No known solution
- Bootstrapping is reminiscent of expectation maximization
 - EM alternates between two steps
 - EM also suffers of semantic drift

CS886 Lecture Slides (c) 2013 P. Poupart

10