

Named Entity Recognition

Lecture 12: October 18, 2013

CS886-2 Natural Language Understanding
University of Waterloo

CS886 Lecture Slides (c) 2013 P. Poupart

1

Entities and Relations

- The essence of a document can often be captured by the entities and relations that are mentioned
- **Entity**: object, person, organization, date, etc.
 - Most things denoted by a noun phrase or pronoun
- **Relation**: property that links one or several entities
 - Most things denoted by an adjective, verb or adverb

CS886-2 Lecture Slides (c) 2013 P. Poupart

2

Named Entities

- Among all entities, named entities are often the most important ones for
 - Text summarization
 - Question answering
 - Information retrieval
 - Sentiment analysis
- Definition: **subset of entities referred by a “rigid designator”**
- Rigid designator: expression that always refers to the same thing in all possible worlds

CS886-2 Lecture Slides (c) 2013 P. Poupart

3

Named Entity Recognition (NER)

- Task:
 - Identify named entities
 - Classify named entities
- Classes:
 - Common: **Person, organization, location, quantity, time, money, percentage, etc.**
 - Biology: genes, proteins, molecules, etc.
 - Fine grained: all Wikipedia concepts (one concept per Wikipedia page)

CS886 Lecture Slides (c) 2013 P. Poupart

4

News NER example

Kofi Atta Annan is a Ghanaian diplomat who served as the seventh Secretary General of the United Nations from January 1, 1997, to January 1, 2007, serving two five-year terms. Annan was the co-recipient of the Nobel Peace Prize in October 2001.

Kofi Annan was born on April 8, 1938, to Victoria and Henry Reginald Annan in Kumasi, Ghana. He is a twin, an occurrence that is regarded as special in Ghanaian culture. Efua Atta, his twin sister, shares the same middle name, which means 'twin'. As with most Akan names, his first name indicates the day of the week he was born: 'Kofi' denotes a boy born on a Friday. The name Annan can indicate that a child was the fourth in the family, but in his family it was simply a name which Annan inherited from his parents.

In 1962, Annan started working as a Budget Officer for the World Health Organization, an agency of the United Nations. From 1974 to 1976, he was the Director of Tourism in Ghana. Annan then returned to work for the United Nations as an Assistant Secretary General in three consecutive positions.

Person
Location
Organization
Date
Nationality
Title

Biomedical NER example

ABNER v1.5

File Annotation Preferences Misc

Source Text

Analysis of myeloid-associated genes in human hematopoietic progenitor cells. Bello-Fernandez et al. Exp Hematol. 1997 Oct;25(11):1158-66.

The distribution of myeloid lineage-associated cytokine receptors and lysosomal proteins was analyzed in human CD34+ cord blood cell (CB) subsets at different stages of myeloid commitment by reverse-transcriptase polymerase chain reaction (RT-PCR). The highly specific granulomonocyte-associated lysosomal proteins myeloperoxidase (MPO) and lysozyme (LZ), as well as the transcription factor PU.1, were already detectable in the most immature CD34+Thy-1+ subset. Messenger RNA (mRNA) levels for the granulocyte-colony stimulating factor (G-CSF)

Annotated Text

Analysis of **myeloid-associated genes** in **human hematopoietic progenitor cells**. Bello-Fernandez et al. Exp Hematol. 1997 Oct; 25 (11) : 1158-66 .

The distribution of **myeloid lineage-associated cytokine receptors** and **lysosomal proteins** was analyzed in **human CD34+ cord blood cell (CB) subsets** at different stages of myeloid commitment by reverse-transcriptase polymerase chain reaction (RT-PCR) . The highly specific **granulomonocyte-associated lysosomal proteins myeloperoxidase (MPO)** and **lysozyme (LZ)** , as well as the **transcription factor PU.1** , were already detectable in the most **immature CD34+ Thy-1+ subset** . **Messenger RNA (mRNA)** levels for the **granulocyte-colony stimulating factor (G-CSF)**

Entity Recognition Tools

Annotate! protein DNA RNA cell line cell type

Classification

- Approach: classify each word (phrase) with an entity type
- Supervised learning:
 - Train with corpus of labeled text (labels are entity types)
- Semi-supervised learning:
 - Train with some labeled texts and large corpus of unlabeled texts

CS886 Lecture Slides (c) 2013 P. Poupart

7

Independent Classifiers

- Classify each word in isolation
 - Naïve Bayes model
 - Logistic regression
 - Decision tree
 - Support vector machine

CS886 Lecture Slides (c) 2013 P. Poupart

8

Correlated Classifiers

- Jointly classify all words while taking into account correlations between some labels
 - Hidden Markov Model
 - Conditional Random Field
- Adjacent words (phrases) often have correlated labels
- Identical words often have the same label

CS886 Lecture Slides (c) 2013 P. Poupart

9

Naïve Bayes Model

- Picture

CS886 Lecture Slides (c) 2013 P. Poupart

10

Features

- Features are more important than the model itself
 - Results: very sensitive to the choice of features
- Feature: anything that can be computed by a program based on the text
- Common features:
 - Word, previous word, next word (more words do not seem to help)
 - Prefixes and suffixes
 - Shape
 - Combination of features
 - Part-of-speech tags
 - Gazetteer

CS886 Lecture Slides (c) 2013 P. Poupart

11

Common Features

- Word, previous word, next word
- Prefixes and suffixes
- Shape

CS886 Lecture Slides (c) 2013 P. Poupart

12

Common Features

- Part-of-speech tags
- Gazetteer
- Combination of features

CS886 Lecture Slides (c) 2013 P. Poupart

13

Training

- Generative training: maximum likelihood
 - $\theta^* = \operatorname{argmax} \Pr(\text{class}, \text{features} | \theta)$
 - Closed form solution: relative frequency counts
 - Fast, but inaccurate
- Discriminative training: conditional maximum likelihood
 - $\theta^* = \operatorname{argmax} \Pr(\text{class} | \text{features}, \theta)$
 - No closed form solution: iterative technique such as gradient ascent
 - Slow but more accurate (optimize the right objective)

CS886 Lecture Slides (c) 2013 P. Poupart

14

Example

Derivation

Logistic Regression

- Alternative to Naïve Bayes model
 - Different parameterization, but often equivalent to discriminative naïve Bayes learning
- Idea: joint distribution is proportional to the exponential of a weighted sum of the features

$$\Pr(\text{class}_i | \text{Features}, w) \propto e^{\sum_j w_{ij} \phi_j(\text{words})}$$

Example

Discriminative Training

- Maximize conditional likelihood

$$w^* = \operatorname{argmax}_w \Pr(\text{class} | \text{Features}, w)$$

- No closed form solution: iterative technique
 - E.g. Gradient ascent

Joint Classification

- Joint classification allows us to take into account correlations between some labels
 - Adjacent words often have correlated entity types
 - Identical words often have the same entity type
- Approaches:
 - **Naïve Bayes extension**: Hidden Markov Model
 - **Logistic regression extension**: conditional random field