Statistical Natural Language Processing

Lecture 10: Named Entity Recognition

Dr. Momtazi
Amirkabir University of Technology

Outline

1. Named Entity Recognition
2. MaxEnt Classification
3. Sequential Modeling
4. Evaluation

Outline

Introduction

- Identify and classify names in text

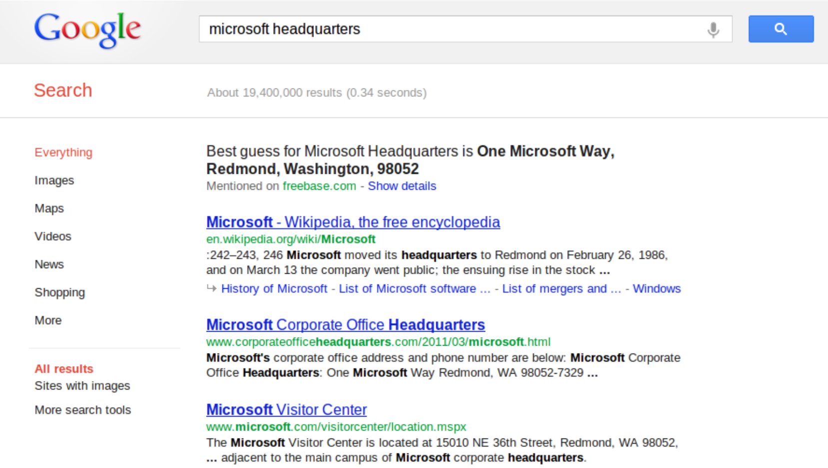
Motivation

- Factual information and knowledge are normally expressed by named entities
  - Who, Whom, Where, When, ...
- Question answering systems are looking for named entities to answer users' questions
- Named entity recognition is the core of the information extraction systems

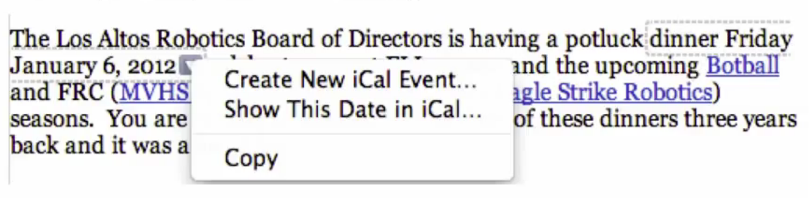
Applications

- Finding the important information of an event from an invitation
  - Date, Time, Location, Host, Contact person
- Finding the main information of a company from its reports
  - Founder, Board members, Headquarters, Profits
- Finding medical information from medical literature
  - Drugs, Genes, Interaction products
- Finding the target of sentiments
  - Products, Celebrities

Applications



Applications



Named Entity Recognition (NER)

- Finding named entities in a text
- Classifying them to the corresponding classes

"Steven Paul Jobs, co-founder of Apple Inc, was born in California."

" $\underbrace{\textit{Steven Paul Jobs}}$, *co-founder of* $\underbrace{\textit{Apple Inc}}$, *was born in* $\underbrace{\textit{California}}$."

" $\underbrace{\textit{Steven Paul Jobs}}_{\textit{PER}}$, *co-founder of* $\underbrace{\textit{Apple Inc}}_{\textit{ORG}}$, *was born in* $\underbrace{\textit{California}}_{\textit{LOC}}$."

Named Entity Classes

- Person
  - Person names
- Organization
  - Companies, Government, Organizations, Committees, ..
- Location
  - Cities, Countries, Rivers, ..
- Date and time expression
- Measure
  - Percent, Money, Weight, ...
- Religious
- Book title
- Movie title
- Drug name

NER Task

- Assigning a label to each token of the text

Steven	PER
Paul	PER
Jobs	PER
,	O
co-founder	O
of	O
Apple	ORG
Inc	ORG
,	O
was	O
born	O
in	O
California	LOC
.	O

IO

Steven	B-PER
Paul	I-PER
Jobs	I-PER
,	O
co-founder	O
of	O
Apple	B-ORG
Inc	I-ORG
,	O
was	O
born	O
in	O
California	B-LOC
.	O

IOB

NER Ambiguity

- IO vs. IOB Encoding

John	PER
Shows	O
Mary	PER
Hermann	PER
Hesse	PER
’s	O
book	O
.	O

John	B-PER
Shows	O
Mary	B-PER
Hermann	B-PER
Hesse	I-PER
’s	O
book	O
.	O

- Although IOB is more accurate, most of the systems use IO for the following reasons
  - IO is much faster than IOB
  - The above case happens very rarely. Even in such cases achieving correct results with IOB is difficult and unlikely

NER Ambiguity

- Ambiguity between named entities and common words
  - May

- Ambiguity between named entity types
  - Washington (Location or Person)

Outline

Making Features from Data

- A feature f links some observed aspects of data d with a class c that we want to predict
- A feature specifies
  - A matching function of properties of the input data
  - A particular class
- The function returned value is 0 or 1

$$f_i(c, d) \equiv [\Phi(d) \wedge c = c_j]$$

$\Rightarrow$ Each feature picks out a data subset that matches the conditions and suggest a label for it.

Making Features from Data

$$f_i(c, d) \equiv [\Phi(d) \wedge c = c_j]$$

"$\underbrace{\text{Steven Paul Jobs}}_{PER}$, co-founder of $\underbrace{\text{Apple Inc}}_{ORG}$, was born in $\underbrace{\text{California}}_{LOC}$."

- Example

$$f_1(c, d) \equiv [w_{-1} = in \wedge isCapitalized(w) \quad \wedge \quad c = LOC]$$

$$f_2(c, d) \equiv [w_{-2} = founder \wedge w_{-1} = of \wedge isCapitalized(w) \quad \wedge \quad c = ORG]$$

Feature Weighting

- Each feature is assigned a positive or negative weight:
  - A positive weight means that the defined matching function is likely to be effective
  - A negative weight means that the defined matching function is likely to be ineffective

Feature Weighting

$$f_i(c,d) \equiv [\Phi(d) \wedge c = c_j]$$

"$\underbrace{\text{Steven Paul Jobs}}_{PER}$, co-founder of $\underbrace{\text{Apple Inc}}_{ORG}$, was born in $\underbrace{\text{California}}_{LOC}$."

- Example

$$1.6 \quad f_1(c,d) \equiv [w_{-1} = in \wedge isCapitalized(w) \quad \wedge \quad c = LOC]$$

$$0.7 \quad f_2(c,d) \equiv [w_{-2} = founder \wedge w_{-1} = of \wedge isCapitalized(w) \quad \wedge \quad c = ORG]$$

$$-1.1 \quad f_3(c,d) \equiv [w_{-1} = by \wedge isCapitalized(w) \quad \wedge \quad c = LOC]$$

Feature-based Linear Classification

1. For each input data item, find the features that matches the data
2. Vote for the class associated with that matching function in the feature set based on the feature weights
3. Calculate the overall vote for each class

$$vote(c) = \Sigma \lambda_i f_i(c, d)$$

4. Choose the class with the maximum vote

Maximum Entropy

$$\hat{c} = \text{argmax}_{c_j} P(c_j|d,\lambda)$$

Makes votes positive

$$P(c_j|d,\lambda) = \frac{exp \sum_i \lambda_i \cdot f_i(c,d)}{\sum_{c_j} exp \sum_i \lambda_i \cdot f_i(c_j,d)}$$

Normalizes votes

Building a MaxEnt Model

- Defining features $f_i(c, d)$
  - Features are often defined by try-and-error on development set
  - They are added during the model development to target errors

- Choosing weighting parameters λ_i
  - Parameters are chosen on the way that maximize the conditional log-likelihood of the training data

$$CLogLik(D) = \Sigma_{i=1}^{n} logP(c_i|d_i)$$

  - It is done by using one of the available numerical optimization packages

Outline

1. Named Entity Recognition
2. MaxEnt Classification
3. Sequential Modeling
4. Evaluation

Task

- Similar to a normal classification task
  - Feature Selection
  - Algorithm

Sequence Modeling

- Many of the NLP techniques should deal with data represented as sequence of items
  - Characters, Words, Phrases, Lines, ...

警察枪杀了那个逃

B I BI B B B B I

I[PRP] *saw*[VBP] *the*[DT] *man*[NN] *on*[IN] *the*[DT] *roof*[NN].

Steven	*Paul*	*Jobs*,	*co-founder*	*of*	*Apple*	*Inc*,	*was*	*born*	*in*	*California*.
PER	PER	PER	O	O	ORG	ORG	O	O	O	LOC

Sequence Modeling

- Making a decision based on the
  - Current Observation
  - Surrounding observations
  - Previous decisions

POS Tagging

- Features

Word	the: the $\rightarrow$ DT
Prefixes	unbelievable: un- $\rightarrow$ JJ
Suffixes	slowly: -ly $\rightarrow$ RB
Lowercased word	Importantly: importantly $\rightarrow$ RB
Capitalization	Stefan: [CAP] $\rightarrow$ NNP
Word shapes	35-year: d-x $\rightarrow$ JJ

NER

- Features

Word	Germany: Germany
POS tag	Washington: NNP
Capitalization	Stefan: [CAP]
Punctuation	St.: [PUNC]
Lowercased word	Book: book
Suffixes	Spanish: -ish
Word shapes	1920-2008: dddd-dddd

- List lookup

List lookup

- Extensive list of names are available via various resources
- The name lists include lists of
  - Entities
    - Organisation, government, airline, educational, ..
    - Location, continent, country, state, city, ...
    - Person first name, last name, ...
  - Entity cues
    - Typical words in organization; e.g., “Limited" or “Incorporated"
    - Person title; e.g., “Mister", “Lord"
- The terms “gazetteer", “lexicon" and “dictionary" are often used interchangeably with the term “list"
  - Gazetteer originally referred to a large list of place names but it became a more general terminology in the NER task

Context Words

- NER
  - Sherwood Forest
  - Portobello Street
  - Mr Smith
  - Apple Inc
  - John earns 3000 €
  - John joined IBM

Learning Model

Training

$T_1 \rightarrow C_1$
$T_2 \rightarrow C_2$
$\ldots$
$T_n \rightarrow C_n$

$\longrightarrow$

F_1
F_2
$\ldots$
F_n

$\longrightarrow$ $\boxed{\text{Model(F,C)}}$

Testing

$T_{n+1} \rightarrow ?$ $\longrightarrow$ F_{n+1} $\longrightarrow$ C_{n+1}

Maximum Entropy Markov Model (MEMM)

- Also known as Conditional Markov Model (CMM)
- The classifier decision is conditioned on the evidence from observations and previous decisions

Conditional Random Field (CRF)

- Another alternative for sequence modeling

$$P(c_1^n|d_1^n,\lambda) = \frac{exp \sum_i \lambda_i \cdot f_i(c,d)}{\sum_{c\prime} exp \sum_i \lambda_i \cdot f_i(c\prime,d)}$$

- A whole-sequence of labels (classes) is conditioned to the whole-sequence of data items rather than a chaining of local models
  - The space of $c\prime$'s is now the space of sequences
- Training is slower than MEMM, but
  - CRFs avoid some of the competition biases in MEMM
  - In practice usually work much the same as MEMM

Challenge

- Dealing with low frequency words

Word class	Example	Intuition
twoDigitNum	90	Two digit year
fourDigitNum	1990	Four digit year
containsDigitAndAlpha	A8956-67	Product code
containsDigitAndDash	09-96	Date
containsDigitAndSlash	11/9/89	Date
containsDigitAndComma	23,000.00	Monetary amount
containsDigitAndPeriod	1.00	Monetary amount, percentage
othernum	456789	Other number
allCaps	BBN	Organization
capPeriod	M.	Person name initial
firstWord	first word of sentence	no useful capitalization information
initCap	Sally	Capitalized word
lowercase	can	Uncapitalized word
other	,	Punctuation marks, all other words

Outline

Precision/Recall Evaluation

- Evaluation is done per entity and not per token

Steven	PER
Paul	PER
Jobs	PER
,	O
co-founder	O
of	O
Apple	ORG
Inc	ORG
,	O
was	O
born	O
in	O
California	LOC
.	O

Steven	PER
Paul	PER
Jobs	PER
,	O
co-founder	O
of	O
Apple	O
Inc	O
,	O
was	O
born	O
in	O
California	LOC
.	O

$$P = \frac{2}{2} = 100\%$$

$$R = \frac{2}{3} = 66\%$$

Precision/Recall Evaluation

- Problem with boundary Errors

Steven	PER
Paul	PER
Jobs	PER
,	O
co-founder	O
of	O
Apple	ORG
Inc	ORG
,	O
was	O
born	O
in	O
California	LOC
.	O

Steven	PER
Paul	PER
Jobs	O
,	O
co-founder	O
of	O
Apple	LOC
Inc	LOC
,	O
was	O
born	O
in	O
California	LOC
.	O

$$P = \frac{1}{3} = 33\%$$

$$R = \frac{1}{3} = 33\%$$

- The boundary error is counted as both fp and fn
- Selecting nothing is even better!!!
- Same problem for wrong entity types

Behind Exact Matching

- Exact matching only accept the items whose both entity boundary and type are correct
- Alternative option is accepting items regardless their boundary or types or both
  - Exact match: detected entity has correct type and boundary
  - Type match: detected entity has correct type but wrong boundary
  - Boundary match: detected entity has correct boundary but wrong type

Further Reading

- Speech and Language Processing
  - Chapter 6: MaxEnt & HMM
  - Chapter 22.1: NER

- Named Entities