Chapter 18: Information Extraction

1

Review and Pointers

- From logistic regression to neural nets (Chapter 7)
  - Each neural unit multiplies input values by a weight vector, adds a bias, and then applies a non-linear activation function (e.g., sigmoid, etc.)
  - Early layers learn representations that can be utilized by later layers
- More recent developments
  - Unassigned chapters 9, 10
    - Sequence processing with recurrent networks
    - Encoder-decoder models, attention, and contextual embeddings
  - From simple word embeddings to BERT (linked article)
    - GLUE: A MULTI-TASK BENCHMARK AND ANALYSIS PLATFORM FOR NATURAL LANGUAGE UNDERSTANDING

2

Machines Beat Humans on a Reading Test. But Do They Understand? (article cont.)

- "Before 2018, one of NLP's main pretraining tools was something like a dictionary. Known as word embeddings, this dictionary encoded associations between words as numbers in a way that deep neural networks could accept as input ….But a neural network pretrained with word embeddings is still blind to the meaning of words at the sentence level. "It would think that 'a man bit the dog' and 'a dog bit the man' are exactly the same thing,"…
- A better method would use pretraining to equip the network with richer rulebooks — not just for vocabulary, but for syntax and context as well — before training it to perform a specific NLP task…
- .. Each of these three ingredients — a deep pretrained language model, attention and bidirectionality — existed independently before BERT. But until Google released its recipe in late 2018, no one had combined them in such a powerful way."

3

Administrivia

- Exam and project
- Homework 3

4

Information Extraction

- A "dumbing down" of more lofty goal of true Natural Language Understanding (i.e., semantics)
  - more technologically manageable
    - often domain and application specific
  - useful practical applications

5

Information Extraction

- Information extraction (IE) systems
  - Find and understand limited relevant parts of texts
  - Gather information from many pieces of text
  - Produce a structured representation of relevant information:
    - *relations* (in the database sense)
    - a *knowledge base*
  - Goals:
    1. Organize information so that it is useful to people
    2. Put information in a semantically precise form that allows further inferences to be made by computer algorithms

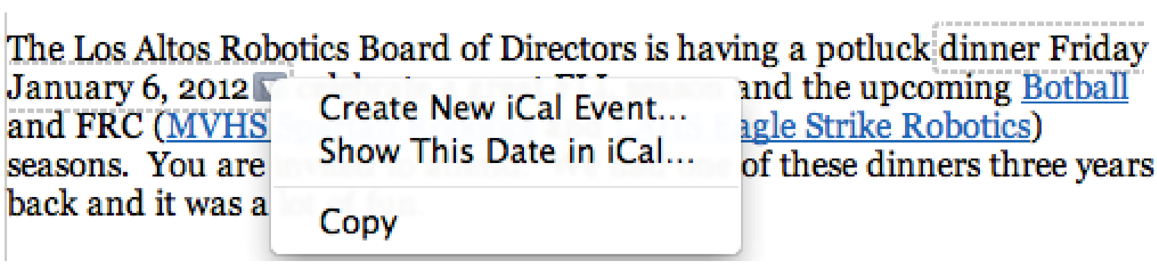
Slides based on Jurafsky and Manning

Information Extraction (IE)

- IE systems extract clear, factual information
  - Roughly: *Who did what to whom when?*
- E.g.,
  - Gathering earnings, profits, board members, headquarters, etc. from company reports
    - The headquarters of BHP Billiton Limited, and the global headquarters of the combined BHP Billiton Group, are located in Melbourne, Australia.
    - headquarters("BHP Biliton Limited", "Melbourne, Australia")
  - Learn drug-gene product interactions from medical research literature

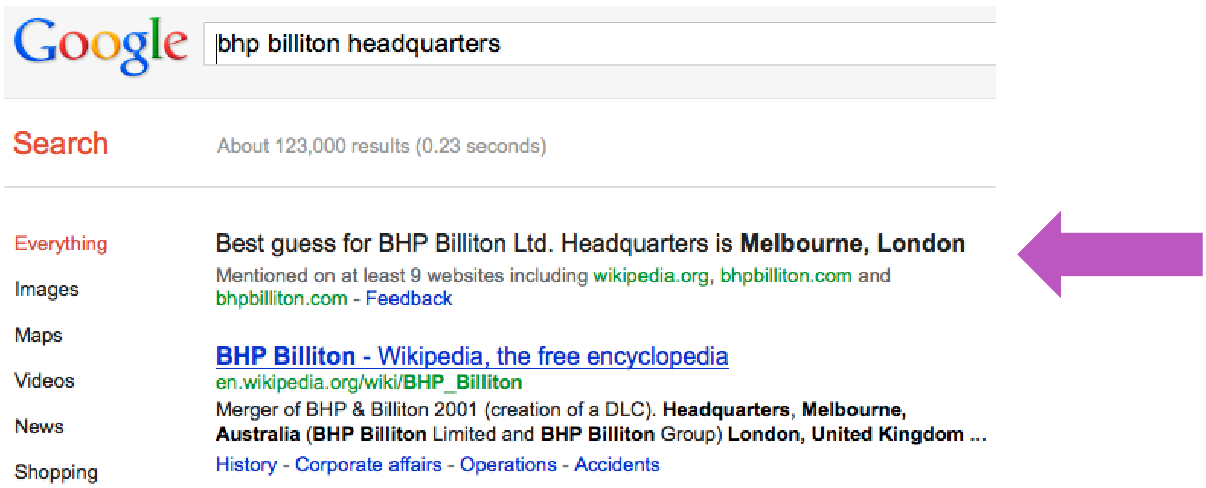
Low-level information extraction

- Is now available in applications like Apple or Google mail, and web indexing



- Often seems to be based on regular expressions and name lists

Low-level information extraction



Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:
  - The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:
  - The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:
  - The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Person
Date
Location
Organi-
zation

Named Entity Recognition (NER)

- The uses:
  - Named entities can be indexed, linked off, etc.
  - Sentiment can be attributed to companies or products
  - A lot of IE relations are associations between named entities
  - For question answering, answers are often named entities.
- Concretely:
  - Many web pages tag various entities, with links to bio or topic pages, etc.
    - Reuters' OpenCalais, Evri, AlchemyAPI, Yahoo's Term Extraction, …
  - Apple/Google/Microsoft/… smart recognizers for document content

As usual, the problem of ambiguity!

[$_{\text{PER}}$ Washington] was born into slavery on the farm of James Burroughs.
[$_{\text{ORG}}$ Washington] went up 2 games to 1 in the four-game series.
Blair arrived in [$_{\text{LOC}}$ Washington] for what may well be his last state visit.
In June, [$_{\text{GPE}}$ Washington] passed a primary seatbelt law.
The [$_{\text{VEH}}$ Washington] had proved to be a leaky ship, every passage I made...

Figure 17.3 Examples of type ambiguities in the use of the name *Washington*.

14

The Named Entity Recognition Task

Task: Predict entities in a text

Foreign	ORG
Ministry	ORG
spokesman	O
Shen	PER
Guofang	PER
told	O
Reuters	ORG
:	:

Standard evaluation is per entity, *not* per token

Precision/Recall/F1 for IE/NER

- Recall and precision are straightforward for tasks where there is only one grain size
- The measure behaves a bit funnily for IE/NER when there are *boundary errors* (which are *common*):
  - First Bank of Chicago announced earnings …
- This counts as both a false positive and a false negative
- Selecting *nothing* would have been better
- Some other metrics (e.g., MUC scorer) give partial credit (according to complex rules)

The ML sequence model approach to NER

Training

1. Collect a set of representative training documents
2. Label each token for its entity class or other (O)
3. Design feature extractors appropriate to the text and classes
4. Train a sequence classifier to predict the labels from the data

Testing

1. Receive a set of testing documents
2. Run sequence model inference to label each token
3. Appropriately output the recognized entities

Encoding classes for sequence labeling

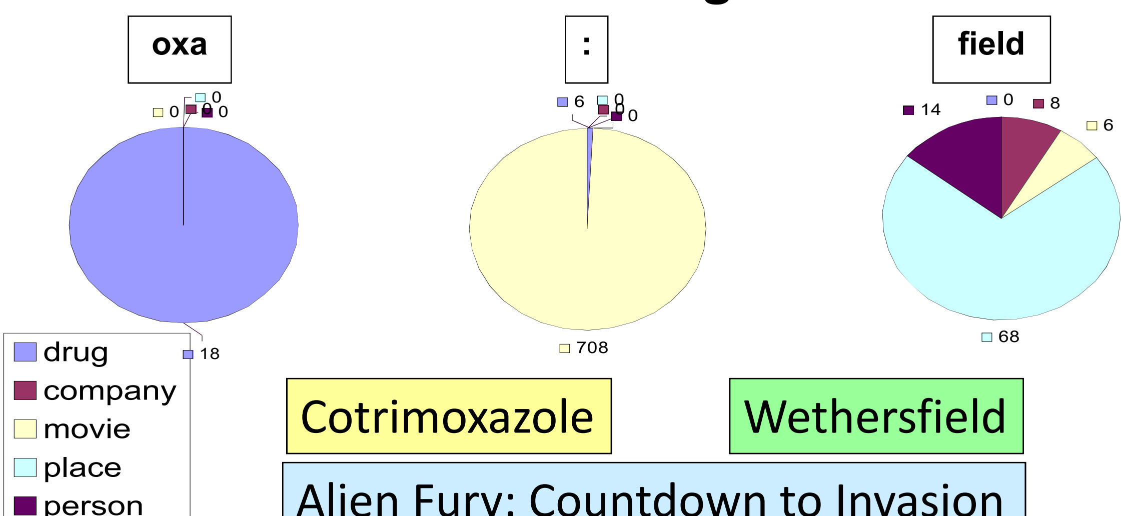
	IO encoding (Stanford)	IOB encoding
Fred	PER	B-PER
showed	O	O
Sue	PER	B-PER
Mengqiu	PER	B-PER
Huang	PER	I-PER
‘s	O	O
new	O	O
painting	O	O

Features for sequence labeling

- Words
  - Current word (essentially like a learned dictionary)
  - Previous/next word (context)
- Other kinds of inferred linguistic classification
  - Part-of-speech tags
- Label context
  - Previous (and perhaps next) label

19

Features: Word substrings



Cotrimoxazole

Wethersfield

Alien Fury: Countdown to Invasion

Features: Word shapes

- Word Shapes
  - Map words to simplified representation that encodes attributes such as length, capitalization, numerals, Greek letters, internal punctuation, etc.

mRNA	xXXX
CPA1	XXXd

Sequence problems

- Many problems in NLP have data which is a sequence of characters, words, phrases, lines, or sentences …
- We can think of our task as one of labeling each item

VBG	NN	IN	DT	NN	IN	NN
Chasing	opportunity	in	an	age	of	upheaval

POS tagging

PERS	O	O	O	ORG	ORG
Murdoch	discusses	future	of	News	Corp.

Named entity recognition

B	B	I	I	B	I	B	I	B	B
而	相	对	于	这	些	品	牌	的	价

Word segmentation

Text segmentation

MEMM inference in systems

- For a Maximum Entropy Markov Model (MEMM), the classifier makes a single decision at a time, conditioned on evidence from observations and previous decisions
- Maximum entropy is an outdated name for logistic regression

Local Context

Decision Point

-3	-2	-1	0	+1
DT	NNP	VBD	???	???
The	Dow	fell	22.6	%

Features

W_0	22.6
W_{+1}	%
W_{-1}	fell
T_{-1}	VBD
T_{-1}-T_{-2}	NNP-VBD
hasDigit?	true
…	…

(Ratnaparkhi 1996; Toutanova et al. 2003, etc.)

MEMMs

- Turn logistic regression onto a discriminative sequence model
  - HMM was generative
    - Easier to add arbitrary features into discriminative models
  - Logistic regression was not a sequence model
  - Optional details in Section 8.5
- **Run logistic regression on successive words, using the class assigned to the prior word as a feature in the classification of the next word**

HMM

$$\begin{aligned}\hat{T} &= \operatorname*{argmax}_T P(T|W) \\ &= \operatorname*{argmax}_T P(W|T)P(T) \\ &= \operatorname*{argmax}_T \prod_i P(word_i|tag_i) \prod_i P(tag_i|tag_{i-1})\end{aligned}$$

MEMM

$$\begin{aligned}\hat{T} &= \operatorname*{argmax}_T P(T|W) \\ &= \operatorname*{argmax}_T \prod_i P(t_i|w_i, t_{i-1})\end{aligned}$$

25

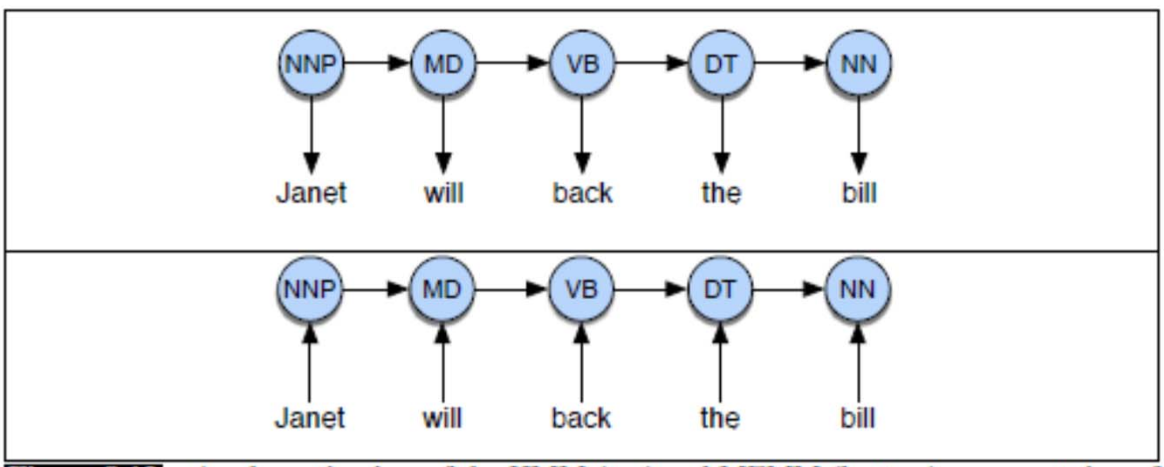


Figure 8.12 A schematic view of the HMM (top) and MEMM (bottom) representation of the probability computation for the correct sequence of tags for the *back* sentence. The HMM computes the likelihood of the observation given the hidden state, while the MEMM computes the posterior of each state, conditioned on the previous state and current observation.

Features in a MEMM

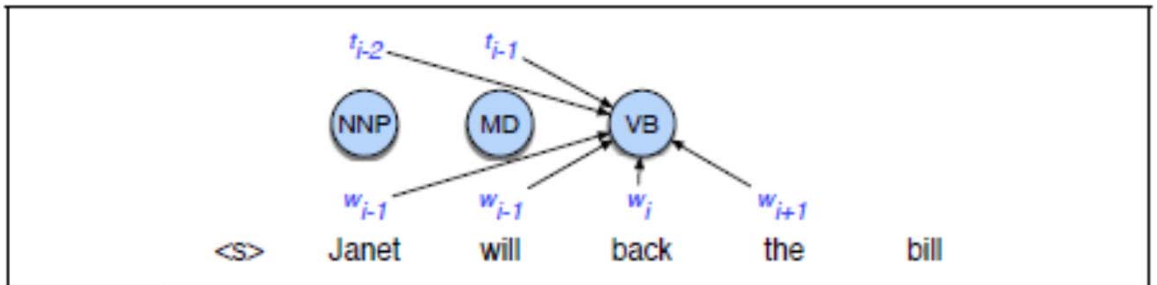


Figure 8.13 An MEMM for part-of-speech tagging showing the ability to condition on more features.

Example: POS Tagging

- Scoring individual labeling decisions is no more complex than standard classification decisions
  - We have some assumed labels to use for prior positions
  - We use features of those and the observed data (which can include current, previous, and next words) to predict the current label

Local Context

Decision Point

-3	-2	-1	0	+1
DT	NNP	VBD	???	???
The	Dow	fell	22.6	%

Features

W_0	22.6
W_{+1}	%
W_{-1}	fell
T_{-1}	VBD
T_{-1}-T_{-2}	NNP-VBD
hasDigit?	true
…	…

(Ratnaparkhi 1996; Toutanova et al. 2003, etc.)

Example: POS Tagging

- POS tagging Features can include:
  - Current, previous, next words in isolation or together.
  - Previous one, two, three tags.
  - Word-internal features: word types, suffixes, dashes, etc.

Local Context

Decision Point

-3	-2	-1	0	+1
DT	NNP	VBD	???	???
The	Dow	fell	22.6	%

Features

W_0	22.6
W_{+1}	%
W_{-1}	fell
T_{-1}	VBD
T_{-1}-T_{-2}	NNP-VBD
hasDigit?	true
…	…

(Ratnaparkhi 1996; Toutanova et al. 2003, etc.)

Greedy Inference

- Greedy inference:
  - We just start at the left, and use our classifier at each position to assign a label
  - The classifier can depend on previous labeling decisions as well as observed data
- Advantages:
  - Fast, no extra memory requirements
  - Very easy to implement
  - With rich features including observations to the right, it may perform quite well
- Disadvantage:
  - Greedy. We make commit errors we cannot recover from

Beam Inference

- Beam inference:
  - At each position keep the top k complete sequences.
  - Extend each sequence in each local way.
  - The extensions compete for the k slots at the next position.
- Advantages:
  - Fast; beam sizes of 3–5 are almost as good as exact inference in many cases.
  - Easy to implement (no dynamic programming required).
- Disadvantage:
  - Inexact: the globally best sequence can fall off the beam.

CRFs [Lafferty, Pereira, and McCallum 2001]

- Another sequence model: Conditional Random Fields (CRFs)
- A whole-sequence conditional model rather than a chaining of local models

Recently also Neural Methods

33

Extracting *relations* from text

- Company report: "International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…"
- Extracted Complex Relation:

  Company-Founding

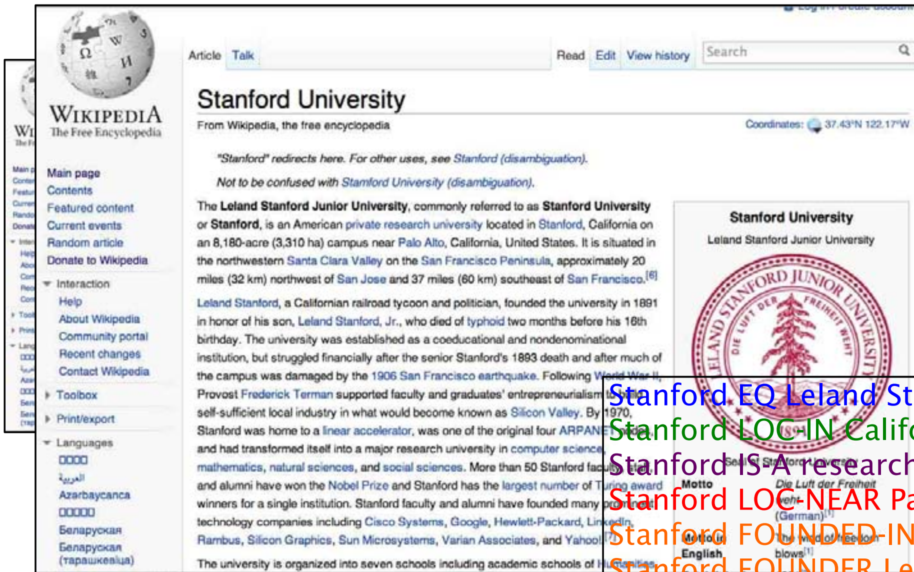
Company	IBM
Location	New York
Date	June 16, 1911
Original-Name	Computing-Tabulating-Recording Co.

- But we will focus on the simpler task of extracting relation **triples**

  Founding-year(IBM,1911)

  Founding-location(IBM,New York)

Extracting Relation Triples from Text



d Junior University,
to as Stanford
rd, is an American
iversity located in
… near Palo Alto,
Stanford…founded
91

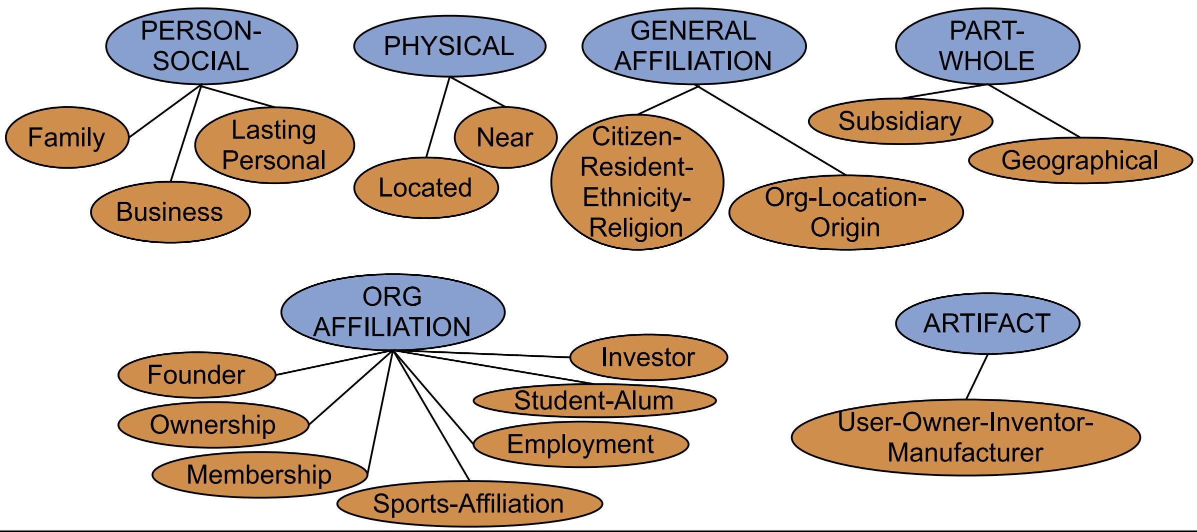
Stanford EQ Leland Stanford Junior University
Stanford LOC-IN California
Stanford IS-A research university
Stanford LOC-NEAR Palo Alto
Stanford FOUNDED-IN 1891
Stanford FOUNDER Leland Stanford

Why Relation Extraction?

- Create new structured knowledge bases, useful for any app
- Augment current knowledge bases
  - Adding words to WordNet thesaurus
- But which relations should we extract?

Automated Content Extraction (ACE)

17 relations from 2008 "Relation Extraction Task"



Automated Content Extraction (ACE)

- Part-Whole-Subsidiary ORG-ORG

  `XYZ, the parent company of ABC`

- Person-Social-Family PER-PER

  `John's wife Yoko`

- Org-AFF-Founder PER-ORG

  `Steve Jobs, co-founder of Apple…`

-

38

UMLS: Unified Medical Language System

- 134 entity types, 54 relations

Injury	disrupts	Physiological Function
Bodily Location	location-of	Biologic Function
Anatomical Structure	part-of	Organism
Pharmacologic Substance	causes	Pathological Function
Pharmacologic Substance	treats	Pathologic Function

Extracting UMLS relations from a sentence

```
Doppler echocardiography can be used to
diagnose left anterior descending artery
stenosis in patients with type 2 diabetes
```

↓

Echocardiography, Doppler DIAGNOSES Acquired stenosis

40

Databases of Wikipedia Relations

Wikipedia Infobox

```
{{Infobox university
|image_name= Stanford University seal.svg
|image_size= 210px
|caption = Seal of Stanford University
|name =Stanford University
|native_name =Leland Stanford Junior Uni
|motto = {{lang|de|"Die Luft der Freiheit v
name="casper">{{cite speech|title=Die Lu
Casper|first=Gerhard|last=Casper|author
05|url=http://www.stanford.edu/dept/pr tml}}</ref>
|mottoeng = The wind of freedom blows<
|established = 1891<ref>{{cite web |
url=http://www.stanford.edu/home/stan ty History |
publisher = Stanford University | accessd
|type = [[private university|Private]]
|calendar= Quarter
|president = [[John L. Hennessy]]
|provost = [[John Etchemendy]]
|city = [[Stanford, California|Stanford]]
|state = California
|country = U.S.
```

Type	Private
Endowment	US$ 16.5 billion (2011)[3]
President	John L. Hennessy
Provost	John Etchemendy
Academic staff	1,910[4]
Students	15,319
Undergraduates	6,878[5]
Postgraduates	8,441[5]
Location	Stanford, California, U.S.
Campus	Suburban, 8,180 acres (3,310 ha)[6]
Colors	Cardinal red and white

Relations extracted from Infobox

Stanford state California

Stanford motto “Die Luft der Freiheit weht”

Relation databases that draw from Wikipedia

- Resource Description Framework (RDF) triples

  subject predicate object

  `Golden Gate Park location San Francisco`

  dbpedia:Golden_Gate_Park dbpedia-owl:location dbpedia:San_Francisco
- DBPedia: 1 billion RDF triples, 385 from English Wikipedia
- Frequent Freebase relations:

  people/person/nationality, location/location/contains
  people/person/profession, people/person/place-of-birth
  biology/organism_higher_classification film/film/genre

42

Ontological relations

Examples from the WordNet Thesaurus

- IS-A (hypernym): subsumption between classes
  - `Giraffe` IS-A `ruminant` IS-A `ungulate` IS-A `mammal` IS-A `vertebrate` IS-A `animal…`
- Instance-of: relation between individual and class
  - `San Francisco` instance-of `city`

Review

- Admin Questions?
  - Hw3
  - Project (groups, schedule, details)?
- Introduction to Information Extraction
- Named Entity Recognition
  - What?
  - How?
- Relation Extractors

How to build relation extractors

1. Hand-written patterns
2. Supervised machine learning
3. Semi-supervised and unsupervised
   - Bootstrapping (using seeds)
   - Distant supervision
   - Unsupervised learning from the web

Rules for extracting IS-A relation

Early intuition from **Hearst (1992)**

- "Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use"

- What does *Gelidium* mean?
- How do you know?`

Rules for extracting IS-A relation

Early intuition from **Hearst (1992)**

- "Agar is a substance prepared from a mixture of **red algae, such as Gelidium,** for laboratory or industrial use"

- What does *Gelidium* mean?
- How do you know?`

Hearst's Patterns for extracting IS-A relations

Hearst pattern	Example occurrences
X and other Y	...temples, treasuries, and other important civic buildings.
X or other Y	Bruises, wounds, broken bones or other injuries...
Y such as X	The bow lute, such as the Bambara ndang...
Such Y as X	...such authors as Herrick, Goldsmith, and Shakespeare.
Y including X	...common-law countries, including Canada and England...
Y , especially X	European countries, especially France, England, and Spain...

Extracting Richer Relations Using Rules

- Intuition: relations often hold between specific entities
  - located-in (ORGANIZATION, LOCATION)
  - founded (PERSON, ORGANIZATION)
  - cures (DRUG, DISEASE)
- Start with Named Entity tags to help extract relation!
- (Maps well to logical representations)

Named Entities aren't quite enough. Which relations hold between 2 entities?

Drug

Cure?

Prevent?

Cause?

Disease

What relations hold between 2 entities?

PERSON

Founder?

Investor?

Member?

Employee?

President?

ORGANIZATION

Extracting Richer Relations Using Rules and Named Entities

Who holds what office in what organization?

PERSON , POSITION `of` ORG

- George Marshall, Secretary of State of the United States

PERSON `(named|appointed|chose|`*etc.*`)` PERSON Prep? POSITION

- Truman appointed Marshall Secretary of State

PERSON [be]? (`named|appointed|`*etc.*) Prep? ORG POSITION

- George Marshall was named US Secretary of State

Hand-built patterns for relations

- Plus:
  - Human patterns tend to be high-precision
  - Can be tailored to specific domains
- Minus
  - Human patterns are often low-recall
  - A lot of work to think of all possible patterns!
  - Don't want to have to do this for every relation!
  - We'd like better accuracy

Supervised machine learning for relations

- Choose a set of relations we'd like to extract
- Choose a set of relevant named entities
- Find and label data
  - Choose a representative corpus
  - Label the named entities in the corpus
  - Hand-label the relations between these entities
  - Break into training, development, and test
- Train a classifier on the training set

How to do classification in supervised relation extraction

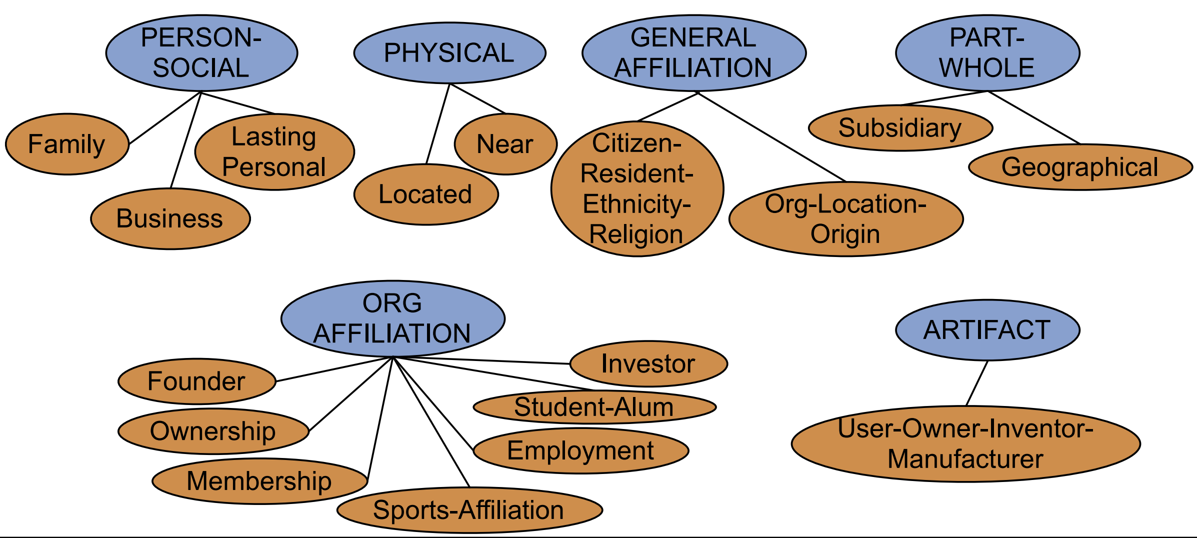
1. Find all pairs of named entities (usually in same sentence)
2. Decide if 2 entities are related
3. If yes, classify the relation

- Why the extra step?
  - Faster classification training by eliminating most pairs
  - Can use distinct feature-sets appropriate for each task.

55

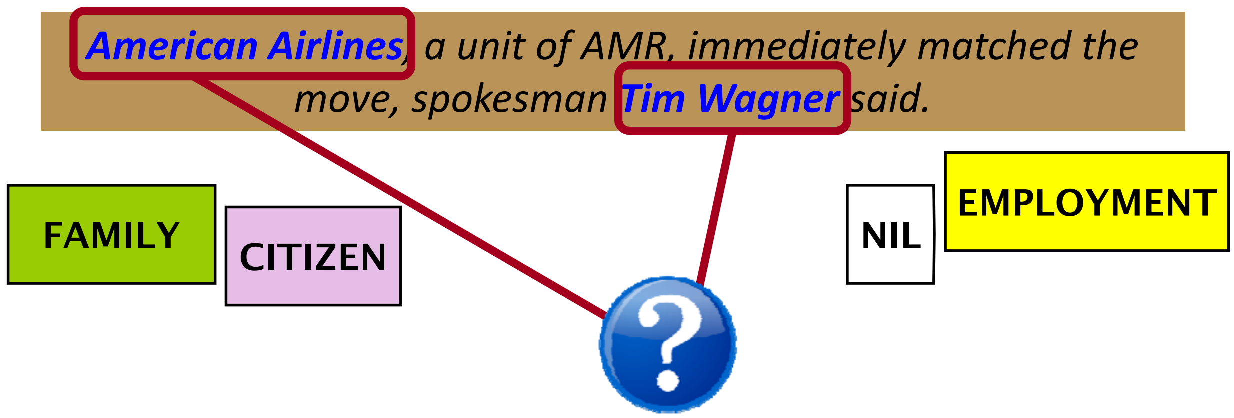
Automated Content Extraction (ACE)

17 sub-relations of 6 relations from 2008 "Relation Extraction Task"



Relation Extraction

Classify the relation between two entities in a sentence



Word Features for Relation Extraction

American Airlines, *a unit of AMR, immediately matched the move, spokesman* ***Tim Wagner*** *said*

Mention 1 Mention 2

- Headwords of M1 and M2, and combination
  - Airlines Wagner Airlines-Wagner
- Bag of words and bigrams in M1 and M2
  - {American, Airlines, Tim, Wagner, American Airlines, Tim Wagner}
- Words or bigrams in particular positions left and right of M1/M2
  - *M2: -1 spokesman*
  - *M2: +1 said*
- Bag of words or bigrams between the two entities
  - {a, AMR, of, immediately, matched, move, spokesman, the, unit}

Named Entity Type and Mention Level Features for Relation Extraction

American Airlines, *a unit of AMR, immediately matched the move, spokesman* ***Tim Wagner*** *said*
Mention 1 Mention 2

- Named-entity types
  - M1: ORG
  - M2: PERSON
- Concatenation of the two named-entity types
  - ORG-PERSON
- Entity Level of M1 and M2 (NAME, NOMINAL, PRONOUN)
  - M1: NAME [it or he would be PRONOUN]
  - M2: NAME [the company would be NOMINAL]

Parse Features for Relation Extraction

American Airlines, *a unit of AMR, immediately matched the move, spokesman* ***Tim Wagner*** *said*
Mention 1 Mention 2

- Base syntactic chunk sequence from one to the other

  NP NP PP VP NP NP
- Constituent path through the tree from one to the other

  NP ↑ NP ↑ S ↑ S ↓ NP
- Dependency path

  Airlines matched Wagner said

Gazeteer and trigger word features for relation extraction

- Trigger list for family: kinship terms
  - parent, wife, husband, grandparent, etc. [from WordNet]
- Gazeteer:
  - Lists of useful geo or geopolitical words
    - Country name list
    - Other sub-entities

***American Airlines**, a unit of AMR, immediately matched the move, spokesman **Tim Wagner** said.*

Entity-based features	
Entity$_1$ type	ORG
Entity$_1$ head	*airlines*
Entity$_2$ type	PERS
Entity$_2$ head	*Wagner*
Concatenated types	ORGPERS
Word-based features	
Between-entity bag of words	{ *a, unit, of, AMR, Inc., immediately, matched, the, move, spokesman* }
Word(s) before Entity$_1$	NONE
Word(s) after Entity$_2$	*said*
Syntactic features	
Constituent path	$NP \uparrow NP \uparrow S \uparrow S \downarrow NP$
Base syntactic chunk path	$NP \rightarrow NP \rightarrow PP \rightarrow NP \rightarrow VP \rightarrow NP \rightarrow NP$
Typed-dependency path	$Airlines \leftarrow_{subj} matched \leftarrow_{comp} said \rightarrow_{subj} Wagner$

Classifiers for supervised methods

- Now you can use any classifier you like
  - MaxEnt
  - Naïve Bayes
  - SVM
  - ...
- Train it on the training set, tune on the dev set, test on the test set

Evaluation of Supervised Relation Extraction

- Compute P/R/F_1 for each relation (both labeled and unlabeled versions, like with parsing)

$$\mathrm{P} = \frac{\text{\# of correctly extracted relations}}{\text{Total \# of extracted relations}}$$

$$F_1 = \frac{2PR}{P+R}$$

$$R = \frac{\text{\# of correctly extracted relations}}{\text{Total \# of gold relations}}$$

64

Summary: Supervised Relation Extraction

+ Can get high accuracies with enough hand-labeled training data, if test similar enough to training

- Labeling a large training set is expensive

- Supervised models are brittle, don't generalize well to different genres

Seed-based or bootstrapping approaches to relation extraction

- No training set? Maybe you have:
  - A few seed tuples or
  - A few high-precision patterns
- Can you use those seeds to do something useful?
  - Bootstrapping: use the seeds to directly learn to populate a relation

Relation Bootstrapping (Hearst 1992)

- Gather a set of seed pairs that have relation R
- Iterate:
  1. Find sentences with these pairs
  2. Look at the context between or around the pair and generalize the context to create patterns
  3. Use the patterns for grep for more pairs

Bootstrapping

- <Mark Twain, Elmira> Seed tuple
  - Grep (google) for the environments of the seed tuple

  "Mark Twain is buried in Elmira, NY."

  X is buried in Y

  "The grave of Mark Twain is in Elmira"

  The grave of X is in Y

  "Elmira is Mark Twain's final resting place"

  Y is X's final resting place.
- Use those patterns to grep for new tuples
  - Often use confidence values to avoid drift
- Iterate

Dipre: Extract <author,book> pairs

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

- Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

- Find Instances:

  The Comedy of Errors, by William Shakespeare, was
  The Comedy of Errors, by William Shakespeare, is
  The Comedy of Errors, one of William Shakespeare's earliest attempts
  The Comedy of Errors, one of William Shakespeare's most

- Extract patterns

  `?x , by ?y ,` `?x , one of ?y 's`

- Now iterate, finding new seeds that match the pattern

Distant Supervision

- Combine bootstrapping with supervised learning
  - Instead of 5 seeds,
    - Use a large database to get huge # of seed examples
  - Create lots of features from all these examples
  - Combine in a supervised classifier
  - Need to also create negative examples via random sampling

Distant supervision paradigm

- Like supervised classification:
  - Uses a classifier with lots of features
  - Supervised by detailed hand-created knowledge
  - Doesn't require iteratively expanding patterns
- Like unsupervised classification:
  - Uses very large amounts of unlabeled data
  - Not sensitive to genre issues in training corpus

Distantly supervised learning of relation extraction patterns

1. For each relation — Born-In
2. For each tuple in big database — <Edwin Hubble, Marshfield> <Albert Einstein, Ulm>
3. Find sentences in large corpus with both entities — Hubble was born in Marshfield; Einstein, born (1879), Ulm; Hubble's birthplace in Marshfield
4. Extract frequent features (parse, words, etc) — PER was born in LOC; PER, born (XXXX), LOC; PER's birthplace in LOC
5. Train supervised classifier using thousands of patterns — $P(\text{born-in} \mid f_1, f_2, f_3, \ldots, f_{70000})$

Unsupervised relation extraction

- Open Information Extraction:
  - extract relations from the web with no training data, no list of relations

73

Evaluation of Semi-supervised and Unsupervised Relation Extraction

- Since it extracts totally new relations from the web
  - There is no gold set of correct instances of relations!
    - Can't compute precision (don't know which ones are correct)
    - Can't compute recall (don't know which ones were missed)
- Instead, we can approximate precision (only)
  - Draw a random sample of relations from output, check precision manually

$$\hat{P} = \frac{\text{\# of correctly extracted relations in the sample}}{\text{Total \# of extracted relations in the sample}}$$

- Can also compute precision at different levels of recall.
  - Precision for top 1000 new relations, top 10,000 new relations, top 100,000
  - In each case taking a random sample of that set
- But no way to evaluate recall

74

Temporal IE for Temporal Reasoning

- Another area, optional

75

Chapter Summary

- IE involves techniques for extracting limited forms of semantic content, e.g.,
  - Named entities
  - Relations among entities
- Methods included various types of machine learning as well as hand crafted methods

76