Goal: “Machine Reading”

Information Extraction (IE)

- Extract entities
  - People, organizations, locations, times, dates, prices, ...
  - Or sometimes: genes, proteins, diseases, medicines, ...
- Extract the relations between entities
  - Located in, employed by, part of, married to, ...
- Figure out the larger events that are taking place
  - attack, transport, die, ...

Information Extraction (IE)

- IE systems extract clear, factual information
  - Roughly: Who did what to whom when? (and maybe where too)
- 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 Billiton Limited", "Melbourne, Australia")
- Learn drug-gene interactions from medical research literature

Machine-readable summaries

- Textual abstract: summary for human
- Structured knowledge extraction: summary for machine
More applications of IE

- Building & extending knowledge bases and ontologies
- Scholarly literature databases: Google Scholar, CiteSeer
- People directories: Rapleaf, Spoke, Naymz
- Shopping engines & product search
- Bioinformatics: clinical outcomes, gene interactions, ...
- Patent analysis
- Stock analysis: deals, acquisitions, earnings, hiring & firings
- SEC filings
- Intelligence analysis for business & government

Information Extraction

- Named Entity Recognition
  - Relation 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.

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.
- Apple/Google/Microsoft/... smart recognizers for document content
- Dialogue systems, like Alexa, Google Home, etc.
Evaluation of Named Entity Recognition

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 like IR and text categorization, where there is only one grain size (documents)
- The measure behaves a bit funny for IE/NER when there are boundary errors (which are common):
  - First Bank of Chicago announced earnings ...

Sequence Models for Named Entity Recognition

Precision/Recall/F1 for IE/NER

- Recall and precision are straightforward for tasks like IR and text categorization, where there is only one grain size (documents)
- The measure behaves a bit funny 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
  - Partial credit?
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

<table>
<thead>
<tr>
<th>IO encoding</th>
<th>IOB encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fred</td>
<td>PER</td>
</tr>
<tr>
<td>showed</td>
<td>O</td>
</tr>
<tr>
<td>Sue</td>
<td>PER</td>
</tr>
<tr>
<td>Meng Shiu</td>
<td>PER</td>
</tr>
<tr>
<td>Huang</td>
<td>PER</td>
</tr>
<tr>
<td>'s</td>
<td>O</td>
</tr>
<tr>
<td>new</td>
<td>O</td>
</tr>
<tr>
<td>painting</td>
<td>O</td>
</tr>
</tbody>
</table>

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

Features: Word substrings

- Drug
- Company
- Movie
- Place
- Person
- 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.

<table>
<thead>
<tr>
<th>Varicella-zoster</th>
<th>Xx-xxx</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA</td>
<td>xXXX</td>
</tr>
<tr>
<td>CPA1</td>
<td>XXXd</td>
</tr>
</tbody>
</table>
Maximum Entropy Sequence Models

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: labeling each item in the sequence

<table>
<thead>
<tr>
<th>VBG</th>
<th>NN</th>
<th>IN</th>
<th>DT</th>
<th>NN</th>
<th>IN</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chasing</td>
<td>opportunity</td>
<td>in</td>
<td>an</td>
<td>age</td>
<td>of</td>
<td>upheaval</td>
</tr>
</tbody>
</table>

POS tagging

<table>
<thead>
<tr>
<th>PERSON</th>
<th>O</th>
<th>O</th>
<th>ORG</th>
<th>ORG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murdoch</td>
<td>discusses</td>
<td>future</td>
<td>of</td>
<td>News Corp.</td>
</tr>
</tbody>
</table>

Named entity recognition

Maximum Entropy

• Make a probabilistic model from the linear combination \( \sum \theta f(c, d) \)

\[
P(c | d, \lambda) = \frac{\exp \sum \theta f(c, d)}{\sum \exp \sum \theta f(c', d)}
\]

Makes votes positive
Normalizes votes

\( f(c, d) \) = \[c = \text{LOCATION} \land w = \text{“in”} \land \text{isCapitalized}(w) \] => weight 1.8
\( f(c, d) \) = \[c = \text{LOCATION} \land \text{hasAccentedLatinChar}(w) \] => weight -0.6
\( f(c, d) \) = \[c = \text{DRUG} \land \text{endsWith}(w, c) \] => weight 0.3

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.

Information Extraction

• Named Entity Recognition
  • Relation Extraction
Relation Extraction

Relation extraction example

CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Relation</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Airlines</td>
<td>subsidiary</td>
<td>AMR</td>
</tr>
<tr>
<td>Tim Wagner</td>
<td>employee</td>
<td>American Airlines</td>
</tr>
<tr>
<td>United Airlines</td>
<td>subsidiary</td>
<td>UAL</td>
</tr>
</tbody>
</table>

Why Relation Extraction?

• Create new structured knowledge bases, useful for any application
• Augment current knowledge bases
  • Adding words to WordNet thesaurus, facts to FreeBase or DBPedia
• Support question answering
  • The granddaughter of which actor stared in the movie "E.T."
• But which relations should we extract?

Automated Content Extraction (ACE)

17 relations from 2008 “Relation Extraction Task”

Automated Content Extraction (ACE)

• Physical-Located PER-GPE
  He was in Tennessee
• 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...

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
How to build relation extractors

1. Hand-written patterns (also can be used as features)
2. Supervised machine learning
3. Semi-supervised and unsupervised
   • Bootstrapping (using seeds)
   • Distant supervision
   • Unsupervised learning from the web

Hand-written Patterns

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, 1992): Automatic Acquisition of Hyponyms

“Y such as X”
“such Y as X”
“X or other Y”
“X and other Y”
“Y including X”
“Y, especially X”

Hearst’s Patterns for extracting IS-A relations

(Hearst, 1992): Automatic Acquisition of Hyponyms

<table>
<thead>
<tr>
<th>Hearst pattern</th>
<th>Example occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>X and other Y</td>
<td>...temples, treasuries, and other important civic buildings.</td>
</tr>
<tr>
<td>X or other Y</td>
<td>Bruises, wounds, broken bones or other injuries...</td>
</tr>
<tr>
<td>Y such as X</td>
<td>The bowlute, such as the Bambarandang...</td>
</tr>
<tr>
<td>Such Y as X</td>
<td>such authors as Herrick, Goldsmith, and Shakespeare.</td>
</tr>
<tr>
<td>Y including X</td>
<td>...common-law countries, including Canada and England...</td>
</tr>
<tr>
<td>Y, especially X</td>
<td>European countries, especially France, England, and Spain...</td>
</tr>
</tbody>
</table>
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!

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

What relations hold between 2 entities?

- Founder?
- Investor?
- Member?
- Employee?
- President?

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.
Features

- Lightweight features — require little pre-processing
  - Bags of words & bigrams between, before, and after the entities
  - Stemmed versions of the same
  - The types of the entities
  - The distance (number of words) between the entities

- Medium-weight features — require base phrase chunking
  - Base-phrase chunk paths
  - Bags of chunk heads

- Heavyweight features — require full syntactic parsing
  - Dependency-free paths
  - Constituent-free paths
  - Tree distance between the entities
  - Presence of particular constructions in a constituent structure

Word Features for Relation Extraction

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

Entity Mention 1

Entity Mention 2

- Headwords of M1 and M2, and combination
  - American Airlines
  - Tim 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
  - M1: +2 spokesman
  - M2: +2 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

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, PP, NP, ADVP, VP, NP, NP

Phrase label paths

PTPH = [NP-Airlines, S-matched, NP-Wagner]
American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.

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

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

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

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

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

Semi-supervised and Unsupervised Relation Extraction

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 to 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"
    - "Elmira is Mark Twain’s final resting place"
  - Use those patterns to grep for new tuples
  - Iterate

Dipre: Extract <author,book> pairs

- Start with 5 seeds:
  - Jane Austen: Pride and Prejudice
  - 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
  - The Comedy of Errors, by William Shakespeare, is one of William Shakespeare’s earliest attempts
  - The Comedy of Errors, one of William Shakespeare’s most

- Extract patterns (group by middle, take longest common prefix/suffix)
  - \(?x\), by \(?y\)
  - \(?x\), one of \(?y\)

- Now iterate, finding new seeds that match the pattern

Distant Supervision

- Combine bootstrapping with supervised learning
  - Instead of small number of seeds,
    - Use a large database to get huge # of seed examples
  - Create lots of features from all these examples
  - Combine in a supervised classifier

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