Semantic Relation Extraction and Its Applications

Chenguang WANG

CS591 Data Mining Seminar
More Informative
Information to Knowledge

Focus on the syntactic structure of the information

Example

Mary has a cat

More Informative

Mary has a cat
Example is_a(Obama, President)

Semantics:
Recover information about relations and entities

More Informative
Example
What states border Illinois?

More Informative
Wisconsin
Indiana
Kentucky
Missouri
Iowa

Knowledge
Recover complete meaning representation

Information to Knowledge

11/13/2013
Goal: Machine Reading

- DARPA Machine Reading program started in 2009
- Goal: acquire structured knowledge from unstructured text
What is Relation Extraction?

Why Relation Extraction?
“Google was founded by Larry Page and Sergey Brin while they were Ph.D. students at Stanford University.”

“They incorporated Google as a privately held company on September 4, 1998.”

“In 2006 Google moved to headquarters in Mountain View, California, nicknamed the Googleplex.”
Why?

- Create new structured knowledge bases, useful for any application
- Augment current knowledge bases
  - Adding words to WordNet thesaurus, facts to FreeBase or DBPedia
- Support Web Search/Question Answering and other applications

Who is the author of Gone with the Wind?

(author-of ?x “Gone with the Wind”)
We Are Here

Problem

Approaches

Applications

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

Supervised machine learning

Semi-supervised and unsupervised
  Bootstrapping (using seeds)
  Distant supervision
  Unsupervised learning from the web
Agar is a substance prepared from a mixture of red algae, such as *Gelidium*, for laboratory or industrial use.”

What does *Gelidium* mean?

---

### Rules for Extracting IS-A Relation

#### Automatic Acquisition of Hyponyms (Hearst, 1992)

<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 bow lute, such as the Bambara ndang...</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>
What Relations Hold Between 2 Entities?

PERSON

Founder?
Employee?
Investor?
Member?

ORGANIZATION

11/13/2013
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

11/13/2013
Hand-written patterns

Supervised machine learning

Semi-supervised and unsupervised
  - Bootstrapping (using seeds)
  - Distant supervision
  - Unsupervised learning from the web
- 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
Supervised Machine Learning for Relations

Sentence: $w_1 w_2 \ldots e_1 \ldots w_i \ldots e_2 \ldots w_{n-1} w_n$

Textual Analysis (POS, Parse, etc.)

- Features
  - Define the feature set
  - Similarity metrics like cosine distance can be used

- Kernels (Structural Presentation)
  - Need to define the similarity metric (Kernel)
  - Kernel similarity is integral to classifiers like SVMs.

Feature Extraction

Or

Classifier

Kernel $K(x,y)$
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
Classify the relation between two entities in a sentence.

Example: "American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said."
Features for Relation Extraction

- **Word Features**
  - Headwords (Airlines, Wagner); Combination; Bag of words; Bigrams; Particular positions

- **Gazetteer and Trigger Word Features**
  - Trigger list for family: parent, wife, husband, grandparent, etc.

- **Named Entity Type and Mention Level Features**
  - Named-entity types (M1:ORG; M2:PERSON); Concatenation; Entity Level (NAME, NOMINAL, PRONOUN)

- **Parse Features**
  - Base syntactic chunk sequence; Constituent path; Dependency path
# Features for Relation Extraction

## Entity-based features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity(_1) type</td>
<td>ORG</td>
</tr>
<tr>
<td>Entity(_1) head</td>
<td>airlines</td>
</tr>
<tr>
<td>Entity(_2) type</td>
<td>PERS</td>
</tr>
<tr>
<td>Entity(_2) head</td>
<td>Wagner</td>
</tr>
<tr>
<td>Concatenated types</td>
<td>ORG_PERS</td>
</tr>
</tbody>
</table>

## Word-based features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between-entity bag of words</td>
<td>{ a, unit, of, AMR, Inc., immediately, matched, the, move, spokesman }</td>
</tr>
<tr>
<td>Word(s) before Entity(_1)</td>
<td>NONE</td>
</tr>
<tr>
<td>Word(s) after Entity(_2)</td>
<td>said</td>
</tr>
</tbody>
</table>

## Syntactic features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constituent path</td>
<td>(NP \uparrow NP \uparrow S \uparrow S \downarrow NP)</td>
</tr>
<tr>
<td>Base syntactic chunk path</td>
<td>(NP \rightarrow NP \rightarrow PP \rightarrow NP \rightarrow VP \rightarrow NP \rightarrow NP)</td>
</tr>
<tr>
<td>Typed-dependency path</td>
<td>(\text{Airlines} \leftarrow \text{sub}_i \text{matched} \leftarrow \text{comp} \text{said} \rightarrow \text{sub}_i \text{Wagner})</td>
</tr>
</tbody>
</table>
Kernel K(x,y) defines similarity between objects x and y implicitly in a higher dimensional space.

(x,y) can be:
- Strings: similarity $\propto$ number of common substrings (or subsequences) between x and y
- Example: $\text{sim}(\text{cat, cant}) > \text{sim}(\text{cat, contact})$
- Excellent introduction to string kernels in Lodhi et. al. (2002)

Extend string kernels to word sequences and parse trees for relation extraction.
Kernels (Trees)

- Similarity computed by counting the number of common subtrees
- Complexity (polynomial)
- Culotta & Sorensen (2004)
  - Use dependency trees. Node attributes are
    - Word, POS, Generalized POS, Chunk tag, Entity type, Entity-level, Relation argument

1. Match attributes of parent nodes
2. If parent nodes match, add 1 to similarity score else return score of 0
3. Compare child-subsequences and continue recursively

11/13/2013
Now we 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
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


**Problem**

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

**Approaches**

**Applications**
<Mark Twain, Elmira> Seed tuple

Grep (bing/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

Iterate
Dipre: Extract <author,book> Pairs

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

- Start with 5 seeds:
  - 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 (group by middle, take longest common prefix/suffix)
  - ?x, by ?y, ?x, one of ?y's

- Now iterate, finding new seeds that match the pattern
Distant Supervision

Learning Syntactic patterns for automatic hypernym discovery. (Snow, Jurafsky, Ng. NIPS 2005)

Autonomously Semantifying Wikipeida. (Fei Wu and Daniel S. Weld. CIKM 2007)

Distant supervision for relation extraction without labeled data. (Mintz, Bills, Snow, Jurafsky. ACL 2009)

- 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
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
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
   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
   \[ P(born−in \mid f_1, f_2, f_3, \ldots, f_{70000}) \]
Unsupervised Learning for Relation Extraction

DIRT – Discovery of Inference Rules from Text (Lin, Pantel. 2003)

- Looks at MINIPAR dependency paths between noun pairs
  - N:subj:V ← find → V:obj:N → solution → N:to:N
  - i.e., X finds solution to Y

- Applies "extended distributional hypothesis"
  - If two paths tend to occur in similar contexts, the meanings of the paths tend to be similar.

- So, defines path similarity in terms of co-occurrence counts with various slot fillers
The top-20 most similar paths to “X solves Y”:

Y is solved by X
X resolves Y
X finds a solution to Y
X tries to solve Y
X deals with Y
Y is resolved by X
X addresses Y
X seeks a solution to Y
X do something about Y
X solution to Y

Y is resolved in X
Y is solved through X
X rectifies Y
X copes with Y
X overcomes Y
X eases Y
X tackles Y
X alleviates Y
X corrects Y
X is a solution to Y
Query Paraphrasing in Web Search

Question Answering

Knowledge Base Construction
Problem

Approaches

Applications

- Query Paraphrasing for Web Search
- Question Answering
- Knowledge Base Construction

- **Mismatch** between queries and documents is a key issue for the web search task
- Caused by expressing the same meaning in different natural language ways
  - E.g.
    - X is the author of Y
    - Y was written by X

**Who is the author of Gone with the Wind?**

Paraphrases (Relation Extraction Patterns)

**Gone with the Wind was written by whom?**
Solution Overview

Paraphrase Extraction
- Extract paraphrase pairs from various data sources

Paraphrase Model
- A search-oriented model generates candidates for each original query

Parameter Optimization
- Optimize the weights of the features used in paraphrasing model on development data

Ranking Model
- An enhanced ranking model by using augmented features computed on paraphrases of original queries

Raw Data

Paraphrase Extraction

Original Query

Paraphrase Model

Original Query
+ N-best Candidates

Ranking Model

DEV Data

Model Optimization

\[
\sum_{i} \lambda_i \cdot h_i(\cdot)
\]
Parameter Optimization

Candidate is sent to the ranker, and returned by an NDCG score

Updated feature weights

After optimization, candidates with higher NDCGs are preferred and ranked on the top of the N-best list
### Impacts of Enhanced Ranking Model

<table>
<thead>
<tr>
<th>Ranking model baseline (Liu et al., 2007)</th>
<th>Dev Set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDCG@1</td>
<td>NDCG@5</td>
</tr>
<tr>
<td></td>
<td>BL-Rank</td>
<td>BL-Rank+Para</td>
</tr>
<tr>
<td>BL-Rank</td>
<td>25.31%</td>
<td>27.28%</td>
</tr>
<tr>
<td>BL-Rank+Para</td>
<td>28.59% (+3.28%)</td>
<td>28.42% (+1.14%)</td>
</tr>
<tr>
<td></td>
<td>33.76%</td>
<td>34.79%</td>
</tr>
<tr>
<td></td>
<td>34.25% (+0.49%)</td>
<td>35.68% (+0.89%)</td>
</tr>
</tbody>
</table>

**Enhanced ranking model**

**BL-Rank:** 
*Query-documents* matching features (unigram/bigram/trigram BM25 and original/normalized Perfect-Match)

**BL-Rank+Para:** 
*Query+Top 1 Paraphrase-documents* matching features (unigram/bigram/trigram BM25 and original/normalized Perfect-Match)

11/13/2013
Problem

Query Paraphrasing in Web Search

Approaches

Question Answering

Applications

Knowledge Base Construction
A QA system:
- Input: a natural language question
- Output: a concise answer to the question

A QA system is not a search engine

Example questions:
- What was the monetary value of the Nobel Peace Prize in 1989?
- What does the Peugeot company manufacture?
- How did Socrates die?
Examples

Q: What does the [BMW company] is-a produce?
A: “[BMW cars] make-produce are sold ..”

Q: Where have nuclear incidents occurred?
A: “The [ (Three Mile Island) (nuclear incident)] loc caused a DOE policy crisis..”

Q: What causes malaria?
A: “..to protect themselves and others from being bitten by [malaria mosquitoes] cause..”
Problem

Approaches

Applications

- Query Paraphrasing in Web Search
- Question Answering
- Knowledge Base Construction
Knowledge Base Construction

The Problem:
- Building knowledge bases (i.e., ontologies) by hand is a laborious, time-intensive process that requires domain expertise (and possibly, knowledge representation expertise).

The Solution:
- Automatic extraction of important concepts and semantic relations from text documents.
- Advantages:
  - less domain expertise required up front
  - less time required to acquire knowledge for a new domain
  - provides greater flexibility than traditional approaches
Knowledge Base Construction

- KAT- Basic Idea (Moldovan & Girju 2001)

- Example:

  “When the US economy enters a boom, mortgage interest rates rise.”

- new concept mortgage interest rate

- state of US economy and the value of mortgage interest rate are in a DIRECT RELATIONSHIP
Gate: general architecture for NLP
  www.gate.shef.ac.uk

UIMA
  Largely inspired by Gate

Open Calais
  Web Service to automatically create rich semantic metadata for submitted content (NER, Events, etc.)
  http://www.opencalais.com/

NLTK: http://nltk.org/#

R: http://cran.cnr.berkeley.edu/
Future Directions

- A better understanding of the core modules of a semantic relation extraction: A tendency towards context/hybrid approaches

- More focus on both open-domain and domain specific applications

- More research on relation extraction in other languages than English and more cross-linguistic analyses of semantic relations

- Design reasonable approach to extract relations from various big data web pages, social network, query log, etc.

  with adapted machine learning techniques
Take-aways

- **Semantic relation extraction**
  - A challenging task
  - A hot topic in both NLP and Text mining
  - Currently three main (different) approaches
    - Hand-written patterns; Supervised machine learning; Semi-supervised and unsupervised
    - None is perfect; each is useful in applications
Thank You!

Q&A
Some slides are adapted from Dan Jurafsky, Bill MacCartney, Roxana Girju


11/13/2013
References


