CSE 454
Advanced Internet Systems
Slot-Filling Architectures
Dan Weld

Logistics
- Datasets
  - Stored on Amazon
- Computing
  - $100 EC2 credit per student
  - You’ll need to set up your own account
- Teams
  - 8-10 teams of 2-3 people
  - Bi-weekly meetings with Dan & Xiao
  - Additional mentor [optional]
- Classes
  - Steady state – one / week

High-Level Architecture

Entity Linking

Query = “James Parsons”

[Chen and Ji, EMNLP2011]

Entity Linking

Named Entity Recognition vs. Linking

• NEL returns entity in KB which has been mentioned
• NER identifies proper names in texts, and classifies them into a set of predefined categories of interest.
  - Especially: person, location and organisation
  - Sometimes date/time expressions, measures (percent, money, weight etc), email addresses etc.
  - Sometimes domain-specific entities:
    names of drugs
    medical conditions,
    names of ships, etc.
Basic Problems in NE

- Variation of NEs – e.g. John Smith, Mr Smith, John.
- Ambiguity of NE types
  - John Smith (company vs. person)
  - May (person vs. month)
  - Washington (person vs. location)
  - 1945 (date vs. time)
- Ambiguity with common words, e.g. “may”

Example

Top Kurnish Militant is Among Three Men in Paris

Battleships, movies, characters, paintings, songs...

Common NEL Features

<table>
<thead>
<tr>
<th>Feature Category</th>
<th>Feature Type</th>
<th>Feature Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Spelling error</td>
<td>Name with incorrect spelling.</td>
</tr>
<tr>
<td></td>
<td>Ambiguity</td>
<td>Name with multiple meanings.</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Name as a location.</td>
</tr>
</tbody>
</table>

Common NEL Features

<table>
<thead>
<tr>
<th>Feature Category</th>
<th>Feature Type</th>
<th>Feature Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document</td>
<td>Position</td>
<td>Position relative to the named entity.</td>
</tr>
<tr>
<td></td>
<td>Local</td>
<td>Name as a local concept.</td>
</tr>
</tbody>
</table>

Paris, France

Paris (English pronunciation: [pəˈʁɛ], [le ˈpæris]; French pronunciation: [paʁi]) is the capital and largest city of France. It is situated on the Seine, in northern Ile de France, at the heart of the Île-de-France region. The city of Paris, within its administrative limits (the 2015 census) has a population of about 2,230,000. Its metropolitan area is one of the largest population concentrations in the world, with more than 12 million inhabitants.

An important settlement for more than 2 millennia, Paris had become, by the 12th century, one of the foremost centres of learning and the arts and the largest city in the Western world until the turn of the 16th century. Paris is today one of the world’s leading business and cultural centres and influences in politics, education, entertainment, media, science, and the arts all contribute to its status as one of the world’s major global cities.

Paris and the Paris Region, with €172.4 billion in 2010, produce more than a quarter of the gross domestic product of France, and has one of the largest city GDPs in the world.

Paris, Mythology

Paris was a child of Penelope (see the list of Thespians). Just before his birth, his mother, the Thespians built a temple for him. At the age of 18, he was the first of the Trojan War ambassadors. In the war, he was one of the original Trojan War. In the end, he was killed by the gods. The city of Paris was named after him.
Common NEL Features

<table>
<thead>
<tr>
<th>Feature Category</th>
<th>Feature Type</th>
<th>Feature Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling Suzuki</td>
<td>Name mapping</td>
<td>Name mapping based on edit distance, order of capital letters, subsequents, word shape, among component words</td>
</tr>
<tr>
<td>Noun</td>
<td>Person</td>
<td>Organization and geographical entity observation markers</td>
</tr>
<tr>
<td>Lexical</td>
<td>Vocabularies</td>
<td>Results in KB items, KB relationships, query text, query set</td>
</tr>
<tr>
<td>Document slicing</td>
<td>Position</td>
<td>Output some snippets with KB texts</td>
</tr>
<tr>
<td>Context</td>
<td>Relations</td>
<td>Entities co-occurrence, or involved in some attribute relationships with the query</td>
</tr>
<tr>
<td>Content</td>
<td>Confidence</td>
<td>Confidence level of entity mentions in the source document and the KB text</td>
</tr>
<tr>
<td>Profile</td>
<td>Meta-data of query KB databases</td>
<td></td>
</tr>
<tr>
<td>Topic Sets</td>
<td>Topic labels and feature candidates for the query text and KB text</td>
<td></td>
</tr>
<tr>
<td>KB I/O Mining</td>
<td>Knowledge extracted from the KB through the KBI and KBM of the KB module</td>
<td></td>
</tr>
<tr>
<td>Popularity</td>
<td>Width</td>
<td>Top KB text ranked by small number and KB length</td>
</tr>
<tr>
<td>Frequency</td>
<td>Frequency in KB text</td>
<td></td>
</tr>
</tbody>
</table>

2011 Results NEL Monolingual

Joint Inference

2011 Results NEL Monolingual

High-Level Architecture

Information Extraction = Relation Extraction = Slot Filling

As a task: Filling slots in a database from sub-segments of text.

Teams

- Named Entity Linking (1)
- Time (1)
- Distant Supervision (1)
- InstaRead (1)
- Relation-Specific (3-5)
- Lexicon Bootstrapping (0-1)
What is “Information Extraction”? 
As a task: Filling slots in a database from sub-segments of text.

Landscape of IE Tasks (2/4): Pattern Scope
Web site specific
- Formatting
- Amazon Book Pages

Genre specific
- Layout
- Resumes

Wide, non-specific
- Language
- University Names

Landscape of IE Tasks (3/4): Pattern Complexity
E.g. word patterns:
- Closed set
  - U.S. states
    - He was born in Alabama.

- Regular set
  - U.S. phone numbers
    - The CALD main office can be reached at (413) 545-1323

- Complex pattern
  - U.S. postal addresses
    - University of Arkansas
      - Phone: (870) 972-2600

- Ambiguous patterns, needing context and many sources of evidence
  - Person names
    - ...was among the six houses sold by Hope Feldman that year.

Landscape of IE Models
- Lexicons
- Classify Pre-segmented Candidates
- Sliding Window
- Boundary Models
- Finite State Machines
- Context Free Grammars

Any of these models can be used to capture words, formatting or both...
Supremacy of Machine Learning

- Machine learning is preferred approach to
  - Speech recognition, Natural language processing
  - Web search – result ranking
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - Computational biology
  - Sensor networks
  - ...

- This trend is accelerating
  - Improved machine learning algorithms
  - Improved data capture, networking, faster computers
  - Software too complex to write by hand
  - New sensors / IO devices
  - Demand for self-customization to user, environment

Space of ML Problems

<table>
<thead>
<tr>
<th>Type of Supervision</th>
<th>Labeled Examples</th>
<th>Reward</th>
<th>Nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete Function</td>
<td>- Classification</td>
<td></td>
<td>Clustering</td>
</tr>
<tr>
<td>Continuous Function</td>
<td>- Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Policy</td>
<td>- Apprenticeship Learning</td>
<td>Reinforcement Learning</td>
<td></td>
</tr>
</tbody>
</table>

Classification

from data to discrete classes

Weather prediction

The classification pipeline
Machine Learning

Supervised Learning
- Parametric
- Non-parametric

Unsupervised Learning
- Reinforcement Learning

Reinforcement Learning
- Y Continuous
- Y Discrete

Parametric Reinforcement Learning
- Decision Trees
- Greedy search; pruning

Non-parametric Reinforcement Learning
- Probability of class | features
- Learn P(Y), P(X|Y); apply Bayes
- Learn P(Y|X) w/ gradient descent

Non-probabilistic Linear Classifier
- Learn w/ gradient descent

Classifier

Hypothesis: Function for labeling examples

Terminology

- Examples
- Features
- Labels

Examples, Labels & Features for RE

Citigroup has taken over EMI, the British ...
Citigroup's acquisition of EMI comes just ahead of ... 
Google's Adwords system has long included ways to connect to Youtube.

Terminology

- Examples
- Features
- Labels
- Training Sample
- Validation Sample
- Test Sample
- Loss Function
- Hypothesis Space
A Learning Problem

Hypothesis Spaces

Knowledge-Based Weak Supervision

Precision & Recall

Precision/Recall Tradeoff

- Can get high precision (but low recall)
  - How?
- Can get high recall (but low precision)
  - How?
- Recall is a non-decreasing function of the number of docs retrieved
  - Precision usually decreases (in a good system)
Precision-Recall Curves
- May return any # of results ordered by similarity
- By varying numbers of docs (levels of recall)
  - Produce a precision-recall curve

A combined measure: F
- Combined measure assessing tradeoff is F measure (weighted harmonic mean):
  \[ F = \frac{1}{\alpha \frac{1}{P} + (1-\alpha) \frac{1}{R}} \]
- People usually use balanced F1 measure
  - i.e., with \( \alpha = \frac{1}{2} \)
- Harmonic mean is conservative average
  - See CJ van Rijsbergen, Information Retrieval

2011 Slot-filling Results

High-Level Architecture

Rules are Crucial
For both hand-engineered and learned systems
- Example
  \[ r(x,y) = \begin{cases} \text{minutes} & \text{if} \ y = \text{time} \\ \text{thing} & \text{if} \ y = \text{thing} \end{cases} \]
- Rules as patterns
- Rules as features
- Rules (or space of rules) typically supplied

Desiderata and Challenges
1. User writes rules in simple, expressive rule language
2. User instantly sees rule extractions on large amounts of text
Writing Rules in Logic

- FOL\(^1\) is simple, expressive, extensible, widely used
- Introduce predicates for NER, dependencies, ...

\[ P(x) \rightarrow \text{NER}(x, a) \land \text{POS}(x, b) \land \text{LKC}() \]

- Rules are deterministic, execute in defined order

\(^1\) We use the subset referred to as “safe domain relational calculus”

Rule Composition

- Define new predicates for similar substructure

[Diagram showing predicate definitions]

- More compact set of rules, better generalization

Enabling Instant Execution

- Translation to SQL allows dynamic optimization

[Diagram showing translation to SQL]

- Each predicate maps to a fragment of SQL
- Intensional and extensional predicates
- Indices, caching, SSDs important

Desiderata and Challenges

1. User writes rules in simple, expressive rule language
2. User instantly sees rule extractions on large amounts of text
3. User gets automatic rule suggestions based on distribution of data

Bootstrap Feature

[Diagram showing feature implementation]

- Provide seed query
- See rule suggestions (lexicalized dependency paths, follows coreference)
- Sort by extractions or mutual information
- Investigate matching sentences
Keyword Feature

Find sentences by keywords

Select words to bring up rule suggestions

See related words and frequencies

See sentence structure to develop rule

Experimental Setup

Datasets

- 22M news sentences
- 1M news sentences
- 22M news sentences
- 5K selected sentences (some gold annotations)

Procedure

Develop

- 4 relational extractors in 55min each

Compare

- To a weakly supervised system
  - Bootstrap, Keyword, Morphology, and Decomposition features
  - To gold annotations for error analysis

Comparison to Weakly Supervised Extraction

<table>
<thead>
<tr>
<th>Rules</th>
<th>attendedSchool</th>
<th>founded</th>
<th>killed</th>
<th>married</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>1.00</td>
<td>.91</td>
<td>.90</td>
<td>.90</td>
</tr>
<tr>
<td>Restrictions</td>
<td>1,411</td>
<td>997</td>
<td>189</td>
<td>4,604</td>
</tr>
<tr>
<td>Weakly supervised Precision</td>
<td>0</td>
<td>.71</td>
<td>N/A</td>
<td>.50</td>
</tr>
<tr>
<td>Weakly supervised Restrictions</td>
<td>5</td>
<td>14</td>
<td>N/A</td>
<td>2</td>
</tr>
</tbody>
</table>

• Precision consistently at least 90%
• Works well, even when weak supervision fails

Comparison of Development Phases

- Bootstrap initially effective, but recall limited
- Decomposition effective for some relations

Error Analysis

- On CoNLL04 ‘killed’ wrt gold annotations: Re .34, Pr .98

<table>
<thead>
<tr>
<th>False Negatives due to Preprocessing (missing predictions)</th>
<th>False Positives due to Preprocessing (wrong predictions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NER</td>
<td>NER</td>
</tr>
<tr>
<td>Dependencies</td>
<td>Dependencies</td>
</tr>
<tr>
<td>Co-references</td>
<td>Co-references</td>
</tr>
<tr>
<td>Rules (missing predictions)</td>
<td>Rules (wrong predictions)</td>
</tr>
<tr>
<td>Lexical items</td>
<td>Lexical items</td>
</tr>
<tr>
<td>Syntactic variation</td>
<td>Syntactic variation</td>
</tr>
<tr>
<td>Reasoning chain</td>
<td>Reasoning chain</td>
</tr>
</tbody>
</table>

• 36% of all errors are due to incorrect pre-processing