CS388: Natural Language Processing

Lecture 20: Information Extraction, SRL, etc.

Greg Durrett
Final project presentation slots announced

Project 2 due today
This Lecture

- How do we represent information for information extraction?
- Semantic role labeling / abstract meaning representation
- Relation extraction
- Slot filling
- Open Information Extraction
Representing Information
Semantic Representations

- “World” is a set of entities and predicates

<table>
<thead>
<tr>
<th>person</th>
<th>president</th>
<th>stab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brutus</td>
<td>Obama</td>
<td>stab(Brutus, Caesar) =&gt; true</td>
</tr>
<tr>
<td>Caesar</td>
<td>Bush</td>
<td></td>
</tr>
<tr>
<td>Obama</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bush</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Statements are logical expressions that evaluate to true or false

*Brutus stabs Caesar*  
*Caesar was stabbed*  

Example credit: Asad Sayeed
Neo-Davidsonian Events

Brutus stabbed Caesar with a knife at the theater on the Ides of March

\[
\exists e \text{ stabs}(e, \text{Brutus}, \text{Caesar}) \land \text{with}(e, \text{knife}) \land \text{location}(e, \text{theater}) \\
\land \text{time}(e, \text{Ides of March})
\]

- Lets us describe events as having properties
- Unified representation of events and entities:

some clever driver in America

\[
\exists x \text{ driver}(x) \land \text{clever}(x) \land \text{location}(x, \text{America})
\]
Barack Obama signed the Affordable Care act on Tuesday. He gave a speech later that afternoon on how the act would help the American people. Several prominent Republicans were quick to denounce the new law.

\[ \exists e \text{ sign}(e, \text{Barack Obama}) \land \text{patient}(e, \text{ACA}) \land \text{time}(e, \text{Tuesday}) \]

- Need to impute missing information, resolve coreference, etc.
- Still unclear how to represent some things precisely or how that information could be leveraged (several prominent Republicans)
Bob and Alice were friends until he moved away to attend college

∃e1∃e2 friends(e1, Bob, Alice) ∧ moved(e2, Bob) ∧ end_of(e1, e2)

- How to represent temporal information?

Bob and Alice were friends until around the time he moved away to attend college

- Representing truly open-domain information is very complicated! We don’t have a formal representation that can capture everything
(At least) Three Solutions

- Crafted annotations to capture some subset of phenomena: predicate-argument structures (semantic role labeling), time (temporal relations), ...

- Slot filling: specific ontology, populate information in a predefined way
  
  (Earthquake: magnitude=8.0, epicenter=central Italy, ...)

- Entity-relation-entity triples: focus on entities and their relations (note that entities is pretty broad: can include events like *World War II*, etc.)
  
  (Lady Gaga, singerOf, Bad Romance)
Entity-relation-entity triples aren’t necessarily grounded in an ontology

Extract strings and let a downstream system figure it out

Barack Obama signed the Affordable Care act on Tuesday. He gave a speech later that afternoon on how the act would help the American people. Several prominent Republicans were quick to denounce the new law.

(Barack Obama, signed, the Affordable Care act)
(Several prominent Republicans, denounce, the new law)
IE: The Big Picture

- How do we represent information? What do we extract?
  - Semantic roles
  - Abstract meaning representation
  - Slot fillers
  - Entity-relation-entity triples (fixed ontology or open)
Semantic Role Labeling/
Abstract Meaning Representation
Performing event
- Subject: Lady Gaga
- Object: a concert
- Audience: students

Same event described but the representation looks different

Verb (predicate) associated with several arguments (roles): “Agent”, “Theme”, and “Beneficiary”
VerbNet

- Defines the semantics of verbs, arguments for every verb in English

<table>
<thead>
<tr>
<th>Frames</th>
<th>Example</th>
<th>Syntax</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP V NP</td>
<td>&quot;Sandy sang a song.&quot;</td>
<td>Agent V Theme</td>
<td><code>perform(during(E), Agent, Theme)</code></td>
</tr>
<tr>
<td>NP V</td>
<td>&quot;Sandy sang.&quot;</td>
<td>Agent V</td>
<td><code>perform(during(E), Agent, ?Theme)</code></td>
</tr>
<tr>
<td>NP V NP PP, Beneficiary</td>
<td>&quot;Sandy sang a song for me.&quot;</td>
<td>Agent V Theme {for} Beneficiary</td>
<td><code>perform(during(E), Agent, Theme) benefit(E, Beneficiary)</code></td>
</tr>
</tbody>
</table>
Semantic Roles

- “Postprocessing” layer on top of dependency parsing that exposes useful information, canonicalizes across grammatical constructions
- Related to theta roles in linguistics
- Agent (~ subject), patient/theme (~ object), goal (~ indirect object)
  - ARG0  ARG1  ARG2+ (semantics vary)
Semantic Role Labeling

- Identify predicate, disambiguate it, identify that predicate’s arguments
- Verb roles from Propbank (Palmer et al., 2005)

Gold: 

![Diagram showing semantic role labeling]

Housing starts are expected to quicken a bit from August’s pace

quicken:

Arg0-PAG: cause of speed-up
Arg1-PPT: thing becoming faster (vnrole: 45.4-patient)
Arg2-EXT: EXT
Arg3-DIR: old speed
Arg4-PRD: new speed

Figure from He et al. (2017)
- Identify predicates (*love*) using a classifier (not shown)
- Identify ARG0, ARG1, etc. as a tagging task with a BiLSTM conditioned on *love*
- Other systems incorporate syntax, joint predicate-argument finding
Question and several answer candidates

Q: Who discovered prions?

AC1: In 1997, Stanley B. Prusiner, a scientist in the United States, discovered prions...

AC2: Prions were researched by...

Score by matching expected answer phrase (EAP) against answer candidate (AC)

Shen and Lapata (2007)
Abstract Meaning Representation

- Graph-structured annotation
- Superset of SRL: full sentence analyses, contains coreference and multi-word expressions as well
- F1 scores in the 60s: hard!
- So comprehensive that it’s hard to predict, but still doesn’t handle tense or some other things…

The boy wants to go
First predict mapping from concepts to graph nodes (many-to-many)

Then use an edge scoring module similar to dependency parsers to predict edges

Predicting a coherent graph is hard, lots of constraints on it and no dynamic program

Flanigan et al. (2016), Lyu et al. (2018)
Summarization with AMR

- Merge AMRs across multiple sentences
- Summarization = subgraph extraction
- No real systems actually work this way (more when we talk about summarization)

Liu et al. (2015)
Slot Filling
Most conservative, narrow form of IE

Indian Express — A massive earthquake of magnitude 7.3 struck Iraq on Sunday, 103 kms (64 miles) southeast of the city of As-Sulaymaniyah, the US Geological Survey said, reports Reuters. US Geological Survey initially said the quake was of a magnitude 7.2, before revising it to 7.3.

Speaker: Alan Clark

“Gender Roles in the Holy Roman Empire”

Old work: HMMs, later CRFs trained per role

Freitag and McCallum (2000)
Slot Filling: MUC

Key aspect: need to combine information across multiple mentions of an entity using coreference

Haghighi and Klein (2010)
Slot Filling: Forums

- Extract product occurrences in cybercrime forums, but not everything that looks like a product is a product.

  TITLE: [ buy ] Backconnect bot
  BODY: Looking for a solid backconnect bot. If you know of anyone who codes them please let me know

  (a) File 0-initiator4856

  TITLE: Exploit cleaning ?
  BODY: Have some Exploits i need fud.

  (b) File 0-initiator10815

Not a product in this context

Portnoff et al. (2017), Durrett et al. (2017)
Relation Extraction
Relation Extraction

- Extract entity-relation-entity triples from a fixed inventory

- Located_In

- Nationality

During the war in **Iraq**, **American journalists** were sometimes caught in the line of fire

- Use NER-like system to identify entity spans, classify relations between entity pairs with a classifier

- Systems can be feature-based or neural, look at surface words, syntactic features (dependency paths), semantic roles

- Problem: limited data for scaling to big ontologies

ACE (2003-2005)
Hearst Patterns

- Syntactic patterns especially for finding hypernym-hyponym pairs ("is a" relations)

  \[ Y \text{ is a } X \quad \text{Berlin is a city} \]

  \[ X \text{ such as } [\text{list}] \quad \text{cities such as Berlin, Paris, and London.} \]

  \[ \text{other } X \text{ including } Y \quad \text{other cities including Berlin} \]

- Totally unsupervised way of harvesting world knowledge for tasks like parsing and coreference (Bansal and Klein, 2011-2012)

Hearst (1992)
Distant Supervision

- Lots of relations in our knowledge base already (e.g., 23,000 film-director relations); use these to bootstrap more training data
- If two entities in a relation appear in the same sentence, assume the sentence expresses the relation

[Steven Spielberg]’s film [Saving Private Ryan] is loosely based on the brothers’ story

Allison co-produced the Academy Award-winning [Saving Private Ryan], directed by [Steven Spielberg]

Mintz et al. (2009)
Distant Supervision

- Learn decently accurate classifiers for ~100 Freebase relations
- Could be used to crawl the web and expand our knowledge base

<table>
<thead>
<tr>
<th>Relation name</th>
<th>100 instances</th>
<th>1000 instances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Syn</td>
<td>Lex</td>
</tr>
<tr>
<td>/film/director/film</td>
<td>0.49</td>
<td>0.43</td>
</tr>
<tr>
<td>/film/writer/film</td>
<td>0.70</td>
<td>0.60</td>
</tr>
<tr>
<td>/geography/river/basin_countries</td>
<td>0.65</td>
<td>0.64</td>
</tr>
<tr>
<td>/location/country/administrative.divisions</td>
<td>0.68</td>
<td>0.59</td>
</tr>
<tr>
<td>/location/location/contains</td>
<td>0.81</td>
<td>0.89</td>
</tr>
<tr>
<td>/location/us_county/county_seat</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>/music/artist/origin</td>
<td>0.64</td>
<td>0.66</td>
</tr>
<tr>
<td>/people/deceased_person/place_of_death</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td>/people/person/nationality</td>
<td>0.61</td>
<td>0.70</td>
</tr>
<tr>
<td>/people/person/place_of_birth</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>0.67</strong></td>
<td><strong>0.66</strong></td>
</tr>
</tbody>
</table>

Mintz et al. (2009)
**FewRel**

- Treats relation classification as a few-shot classification problem

- 100 classes x 700 instances, goal is to generalize to each class with just a few instances

- BERT can handle this fairly well (Soares et al., 2019)

- “FewRel 2.0”: new dataset with “none of the above” type, which makes things much harder

<table>
<thead>
<tr>
<th>Supporting Set</th>
<th>Test Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) capital_of</td>
<td>Euler was elected a foreign member of the Royal Swedish Academy of Sciences.</td>
</tr>
<tr>
<td>(B) member_of</td>
<td></td>
</tr>
<tr>
<td>(C) birth_name</td>
<td></td>
</tr>
</tbody>
</table>

Han et al. (2018), Gao et al. (2019)
Entity Tracking / Procedural Text
Information extraction for “procedural text”: text describing some kind of process

For a recipe: what ingredients are involved at each timestep?

Involves global constraints and being able to model complex entity interactions

Kiddon et al. (2016), Bosselut et al. (2018)
Slide credit: Aditya Gupta
Entity Tracking

- Process paragraphs: predict when objects are created, moved, or destroyed in a scientific process.

- Structured prediction problem, tied to the particular information conveyed in these paragraphs.

- Use a neural CRF to make a coherent prediction for each entity.

ProPara Dataset

<table>
<thead>
<tr>
<th>Seq. of Steps</th>
<th>water</th>
<th>mixture</th>
<th>sugar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roots absorb water from soil.</td>
<td>M</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>The water flows to the leaf.</td>
<td>M</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Light from the sun and CO₂ enter the leaf.</td>
<td>E</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Light, water, and CO₂ combine into mixture.</td>
<td>D</td>
<td>C</td>
<td>O</td>
</tr>
<tr>
<td>Mixture forms sugar.</td>
<td>O</td>
<td>D</td>
<td>C</td>
</tr>
</tbody>
</table>

Implicit Events requiring Global Knowledge

<table>
<thead>
<tr>
<th>Structural Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>C → M → D</td>
</tr>
</tbody>
</table>

Dalvi et al. (2018), Gupta and Durrett (2019)

Slide credit: Aditya Gupta
Open IE
Open Information Extraction

- “Open”ness — want to be able to extract all kinds of information from open-domain text

- Acquire commonsense knowledge just from “reading” about it, but need to process lots of text (“machine reading”)

- Typically no fixed relation inventory
TextRunner

- Extract positive examples of \((e, r, e)\) triples via parsing and heuristics
- Train a Naive Bayes classifier to filter triples from raw text: uses features on POS tags, lexical features, stopwords, etc.

Barack Obama, 44th president of the United States, was born on August 4, 1961 in Honolulu

=> Barack Obama, was born in, Honolulu

- 80x faster than running a parser (which was slow in 2007...)
- Use multiple instances of extractions to assign probability to a relation

Banko et al. (2007)
Exploiting Redundancy

- 9M web pages / 133M sentences
- 2.2 tuples extracted per sentence, filter based on probabilities
- Concrete: definitely true
  Abstract: possibly true but underspecified
- Hard to evaluate: can assess precision of extracted facts, but how do we know recall?

Banko et al. (2007)
More constraints: open relations have to begin with verb, end with preposition, be contiguous (e.g., *was born on*)

- Extract more meaningful relations, particularly with light verbs

<table>
<thead>
<tr>
<th>verb</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>is</em></td>
<td>is an album by, is the author of, is a city in</td>
</tr>
<tr>
<td><em>has</em></td>
<td>has a population of, has a Ph.D. in, has a cameo in</td>
</tr>
<tr>
<td><em>made</em></td>
<td>made a deal with, made a promise to</td>
</tr>
<tr>
<td><em>took</em></td>
<td>took place in, took control over, took advantage of</td>
</tr>
<tr>
<td><em>gave</em></td>
<td>gave birth to, gave a talk at, gave new meaning to</td>
</tr>
<tr>
<td><em>got</em></td>
<td>got tickets to, got a deal on, got funding from</td>
</tr>
</tbody>
</table>
For each verb, identify the longest sequence of words following the verb that satisfy a POS regex (V .* P) and which satisfy heuristic lexical constraints on specificity.

Find the nearest arguments on either side of the relation.

Annotators labeled relations in 500 documents to assess recall.

Fader et al. (2011)
(a) **CCG parse** builds an underspecified semantic representation of the sentence.

\[
\begin{array}{cccc}
\text{Former} & \text{municipalities} & \text{in} & \text{Brandenburgh} \\
N/N & N & N\backslash N/N & NP \\
\lambda f \lambda x. f(x) \land \text{former}(x) & \lambda x. \text{municipalities}(x) & \lambda f \lambda x \lambda y. f(y) \land \text{in}(y, x) & \text{Brandenburg} \\
\lambda x. \text{former}(x) \land \text{municipalities}(x) & N \backslash N & \lambda f \lambda y. f(y) \land \text{in}(y, \text{Brandenburg}) & N \\
l_0 = \lambda x. \text{former}(x) \land \text{municipalities}(x) \land \text{in}(x, \text{Brandenburg})
\end{array}
\]

(b) **Constant matches** replace underspecified constants with Freebase concepts

\[
\begin{align*}
l_0 &= \lambda x. \text{former}(x) \land \text{municipalities}(x) \land \text{in}(x, \text{Brandenburg}) \\
l_1 &= \lambda x. \text{former}(x) \land \text{municipalities}(x) \land \text{in}(x, \text{Brandenburg}) \\
l_2 &= \lambda x. \text{former}(x) \land \text{municipalities}(x) \land \text{location}.\text{containedby}(x, \text{Brandenburg}) \\
l_3 &= \lambda x. \text{former}(x) \land \text{OpenRel}(x, \text{Municipality}) \land \text{location}.\text{containedby}(x, \text{Brandenburg}) \\
l_4 &= \lambda x. \text{OpenType}(x) \land \text{OpenRel}(x, \text{Municipality}) \land \text{location}.\text{containedby}(x, \text{Brandenburg})
\end{align*}
\]
Takeaways

- SRL/AMR: handle a bunch of phenomena, but more or less like syntax++ in terms of what they represent
- Relation extraction: can collect data with distant supervision, use this to expand knowledge bases
- Slot filling: tied to a specific ontology, but gives fine-grained information
- Open IE: extracts lots of things, but hard to know how good or useful they are
  - Can combine with standard question answering
  - Add new facts to knowledge bases
- Many, many applications and techniques