Natural Language Processing
Stuff we didn’t cover
Coreference resolution

Coreference resolution: the task of clustering together of expressions that refer to the same concept/entity

Michelle LaVaughn Robinson Obama is an American lawyer and writer. She is the wife of the 44th president of the United States, Barack Obama, and the first African-American first lady of the United States.
Importance

• Crucial for general natural language understanding
  • *Who is Barack Obama’s spouse?*
  • Information extraction
  • Anywhere you need to understand more than one sentence there are coreference issues
Main sub-tasks

• Entity extraction
• Coreference resolution
• Entity linking
Entity extraction

- Find all mentions of entities in a text
  - *Michelle LaVaughn Robinson Obama*
  - *She*
  - *The wife of...*
Coreference resolution

• Clustering of the entities extracted

Michelle Obama
she
the wife of...

Barack Obama
The president...
Entity linking

Michelle Obama
the wife of...
General approaches

• Score every pair of entity mentions
• Find the best clustering
• In practice this is hard so usually a more greedy approach is used
Question answering

• A central task in NLP
• One can argue that any task can be reduced to QA with natural language
• A lot of interest in last couple of years
• Dominated by many large scale corpora and deep learning
SQuAD

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called “showers”.

What causes precipitation to fall?
**gravity**

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
**graupel**

Where do water droplets collide with ice crystals to form precipitation?
**within a cloud**
Context
The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the “Top Gear” host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon “to an unprovoked physical and verbal attack.” . . .

Query
Producer X will not press charges against Jeremy Clarkson, his lawyer says.

Answer
Oisin Tymon
# WikiReading

<table>
<thead>
<tr>
<th>Categorization</th>
<th>Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Document</strong></td>
<td><strong>Extraction</strong></td>
</tr>
<tr>
<td>Folkart Towers are twin skyscrapers in the Bayrakli district of the Turkish city of Izmir. Reaching a structural height of 200 m (656 ft) above ground level, they are the tallest ...</td>
<td>Breaking Bad is an American crime drama television series created and produced by Vince Gilligan. The show originally aired on the AMC network for five seasons, from January 20, 2008, to ...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Property</strong></th>
<th><strong>country</strong></th>
<th><strong>original language of work</strong></th>
<th><strong>located next to body of water</strong></th>
<th><strong>start time</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Answer</strong></td>
<td>Turkey</td>
<td>Spanish</td>
<td>Atlantic Ocean, Arctic Ocean, Pacific Ocean</td>
<td>20 January 2008</td>
</tr>
</tbody>
</table>
Q: what is rba

A: Results-Based Accountability is a disciplined way of thinking and taking action that communities can use to improve the lives of children, youth, families, adults and the community as a whole.

p1: Since 2007, the RBA's outstanding reputation has been affected by…
p2: The Reserve Bank of Australia (RBA) came into being on 14 January…
p3: Results-Based Accountability® (also known as RBA) is a disciplined…
…
**TriviaQA**

**Question:** The Dodecanese Campaign of WWII that was an attempt by the Allied forces to capture islands in the Aegean Sea was the inspiration for which acclaimed 1961 commando film?  
**Answer:** The Guns of Navarone  
**Excerpt:** The Dodecanese Campaign of World War II was an attempt by Allied forces to capture the Italian-held Dodecanese islands in the Aegean Sea following the surrender of Italy in September 1943, and use them as bases against the German-controlled Balkans. The failed campaign, and in particular the Battle of Leros, inspired the 1957 novel *The Guns of Navarone* and the successful 1961 movie of the same name.

**Question:** American Callan Pinckney’s eponymously named system became a best-selling (1980s-2000s) book/video franchise in what genre?  
**Answer:** Fitness  
**Excerpt:** Callan Pinckney was an American fitness professional. She achieved unprecedented success with her Callanetics exercises. Her 9 books all became international best-sellers and the video series that followed went on to sell over 6 million copies. Pinckney’s first video release ”Callanetics: 10 Years Younger In 10 Hours” outsold every other fitness video in the US.
James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle. After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

2) What did James pull off of the shelves in the grocery store?
A) pudding
B) fries
C) food
D) splinters

3) Where did James go after he went to the grocery store?
A) his deck
B) his freezer
C) a fast food restaurant
D) his room
“… Water is split, providing a source of electrons and protons (hydrogen ions, H\(^+\)) and giving off O\(_2\) as a by-product. Light absorbed by chlorophyll drives a transfer of the electrons and hydrogen ions from water to an acceptor called NADP\(^+\) …”

Q What can the splitting of water lead to?
   a Light absorption
   b Transfer of ions
## Natural Language Inference

<table>
<thead>
<tr>
<th>Text</th>
<th>Judgments</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>A man inspects the uniform of a figure in some East Asian country.</td>
<td>contradiction C C C C C</td>
<td>The man is sleeping</td>
</tr>
<tr>
<td>An older and younger man smiling</td>
<td>neutral N N E N N</td>
<td>Two men are smiling and laughing at the cats playing on the floor.</td>
</tr>
<tr>
<td>A black race car starts up in front of a crowd of people.</td>
<td>contradiction C C C C C</td>
<td>A man is driving down a lonely road.</td>
</tr>
<tr>
<td>A soccer game with multiple males playing.</td>
<td>entailment E E E E E</td>
<td>Some men are playing a sport.</td>
</tr>
<tr>
<td>A smiling costumed woman is holding an umbrella.</td>
<td>neutral N N E C N</td>
<td>A happy woman in a fairy costume holds an umbrella.</td>
</tr>
</tbody>
</table>
Dependency parsing

• In constituency parsing:
  • We saw both greedy and graph-based methods
  • Similar distinctions also occur in dependency parsing, but the algorithms are different
Character-based models
More...

- Multi-task learning
- ILP inference and LP relaxation
- Summarization
- Interactive learning
- ...
Broad takeaways
What did we cover?

• Word embeddings
• Language models
• Sequence tagging
• Syntactic parsing
• Semantic parsing
Main technical tools

- Structured prediction
- Deep learning
- General recipe:
  - Define a parameterized mapping from input to output
  - Define a loss function
  - Optimize
  - Find best output at test time
High-level observations

• Often linear models can be replaced with non-linear ones without change to the guarantees

• A growing trend is to move the burden from inference to learning. Let information flow between various variables in a neural network and make inference very simple

• It is still an active research area
Hopefully you

• Appreciate the complexity of building systems for natural language

• Understand the main tools used to build state-of-the-art systems nowadays

• Have solid background to read papers

• Have solid background to develop models for NLP
Thank you!