

NLP 201:

NLP Tasks II

Jeffrey Flanigan

December 2, 2021

University of California Santa Cruz
jmflanig@ucsc.edu

Final Exam

- Final Tuesday Dec 7, 12pm (24 hours)
- On Canvas
- Similar format to Midterm

Plan For Today

- Question Answering
- Semantic Parsing
- Semantic Role Labeling
- Information Extraction

us congress

[All](#) [News](#) [Images](#) [Videos](#) [Maps](#) [More](#)[Settings](#) [Tools](#)

About 793,000,000 results (0.57 seconds)

Ad · www.ballotpedia.org/116th/Congress

116th U.S. Congress | Full List | ballotpedia.org

Find out who holds leadership positions in Congress and see a list of members. Find out who is part of the 116th Congress including membership and leadership. Candidate Facts. Encyclopedic information. Highlights: Professional Staff Available, Founded In 2007.

About Ballotpedia

Digital Encyclopedia
Covering U.S. Elections

Support Our Work

Grow The Encyclopedia
Of American Politics

www.congress.gov ▾

Congress.gov | Library of Congress

U.S. Congress legislation, Congressional Record debates, Members of Congress, legislative process educational resources presented by the Library of ...

Show all members · Roll Call Votes · Browse · Contact Us

www.house.gov ▾

House.gov: Home

Speaker Nancy Pelosi addresses the Members of the 116th Congress. Watch Live House Proceedings. The Office of the Clerk gives you a front-row seat. U.S. ...

Find Your Representative · Representatives · The House Explained · Committees

en.wikipedia.org › wiki › United_States_Congress ▾

United States Congress - Wikipedia

The United States Congress is the bicameral legislature of the federal government of the United States, and consists of two chambers: the House of Representatives and the Senate. The Congress meets in the United States Capitol in Washington, D.C.

House of Representatives last election: [November 6, 2018](#) · Houses: [Senate](#); [House of Representatives](#)

House of Representatives next election: [November 3, 2020](#) · Speaker of the House: [Nancy Pelosi](#) (D); since ...

[More images](#)

United States Congress

congress.gov

The United States Congress is the bicameral legislature of the federal government of the United States, and consists of two chambers: the House of Representatives and the Senate. The Congress meets in the United States Capitol in Washington, D.C. [Wikipedia](#)

President of the Senate: [Mike Pence](#) (R); since [January 20, 2017](#)

President pro tempore of the Senate: [Chuck Grassley](#) (R); since [January 3, 2019](#)

Senate political groups: [Republican](#) (53); [Democratic](#) (45); Independent (2)

Senate last election: November 6, 2018

Senate next election: November 3, 2020

Number of members: Congress has 535 voting members: 435 representatives and 100 senators. [wikipedia.org](#)



who is the speaker of the house



All

News

Images

Books

Shopping

More

Settings

Tools

About 1,350,000,000 results (0.69 seconds)

United States House of Representatives / Speaker of the House

Nancy Pelosi



People also search for



Rudy
Giuliani
Trending



Donald
Trump



Dianne
Feinstein
Trending



Adam
Schiff



Devin
Nunes
Trending



Chuck
Schumer



Alexandria
Ocasio-Cortez

View 15+ more

Claim this knowledge panel

Feedback

Hey Siri who won the Alaska
democratic primary?

Tap to Edit ➔

**According to the Associated Press,
Joe Biden has won the Alaska
Democratic primary with 55.3% of
the vote.**



2020 ELECTIONS COVERAGE

Alaska Democratic Party Primary

2020 Presidential Primaries Results

CANDIDATE	VOTE%	DELEGATES
Joe Biden	55.3%	11
Bernie Sanders	44.7%	4
Elizabeth Warren	—	—
Tulsi Gabbard	—	—

100% of precincts reporting

Updated at 07:27PM PDT
Election results from The Associated Press



Watson



Question Answering: Tasks

Textual Question Answering (Reading Comprehension)

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, *Il milione* (or, *The Million*, known in English as the *Travels of Marco Polo*), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge through contact with Persian traders since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: through contact with Persian traders

Textual Question Answering

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

Conversational Question Answering

The Virginia governor's race, billed as the marquee battle of an otherwise anticlimactic 2013 election cycle, is shaping up to be a foregone conclusion. Democrat Terry McAuliffe, the longtime political fixer and moneymen, hasn't trailed in a poll since May. Barring a political miracle, Republican Ken Cuccinelli will be delivering a concession speech on Tuesday evening in Richmond. In recent ...

Q: What are the candidates **running** for?

A: Governor

Q: **Where?**

A: Virginia

Q: Who is the democratic candidate?

A: Terry McAuliffe

Q: Who is **his** opponent?

A: Ken Cuccinelli

Q: What party does **he** belong to?

A: Republican

Q: Which of **them** is winning?

Long-form Question Answering

Question | How do jellyfish function without a brain and a nervous system?

 Abstractive

Jellyfish don't have brains. Jellyfish have a nervous system. They know how to swim, they don't even know what to do with their bodies. They have a "nervous system", which is basically a bunch of nerve cells that connect to the body. The nervous system is made up of neurons, neurons, nerves, etc. Each neuron doesn't have a specific set of "neurons", it just has a different set of nerves. It has a set of sensory neurons, which are all connected to each other. It just knows that it does react to the stimulus, and it knows that the stimulus is there. They have a bunch of nervous systems that act like a filter to get information back.

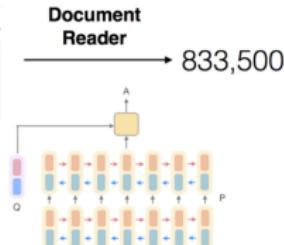
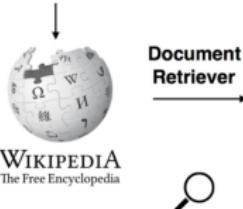
 Extractive

But none of these hypotheses make much sense for a jellyfish because they don't have brains at all. They just have a nerve net — a loose ring of neurons that runs around the rim of their pulsating bells. They have an unusual nervous system, 451 because jellyfish are not bilaterally symmetrical — that is, they don't have a left side and a right side. Jellyfish don't have brains, but their nervous systems detect smells, light and other stimuli, and they coordinate their physical responses.

Open-domain Question Answering

DrQA

Q: How many of Warsaw's inhabitants spoke Polish in 1933?



```
>>> process('What is the answer to life, the universe, and everything?')
```

Top Predictions:

Rank	Answer	Doc	Answer Score	Doc Score
1	42	Phrases from The Hitchhiker's Guide to the Galaxy	47242	141.26

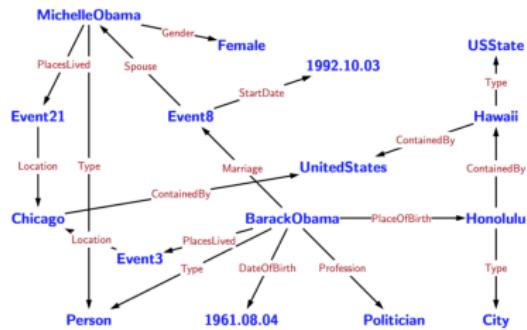
(Chen et al, 2017): Reading Wikipedia to Answer Open-Domain Questions

Knowledge Base Question Answering



100M entities (nodes)

1B assertions (edges)



Which states' capitals are also their largest cities by area?

semantic parsing

$\mu x. \text{Type.USState} \sqcap \text{Capital}.\text{argmax}(\text{Type.City} \sqcap \text{ContainedBy}.x, \text{Area})$

execute

Arizona, Hawaii, Idaho, Indiana, Iowa, Oklahoma, Utah

(Berant et al, 2013): Semantic Parsing on Freebase from Question-Answer Pairs

Table-based Question Answering

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

x = Greece held its last Summer Olympics in which year?

y = 2004

Visual Question Answering



What color are her eyes?

What is the mustache made of?



How many slices of pizza are there?

Is this a vegetarian pizza?

Question Answering Datasets

- **Reading Comprehension**

CNN/Daily Mail, CoQA, HotpotQA, QuAC, RACE, SQuAD, SWAG, Receipt QA, NarrativeQA, DROP, Story Cloze Test

- **Open-domain question answering**

DuReader, Quasar, SearchQA, ...

- **Knowledge base question answering**

Check out more datasets: http://nlpprogress.com/english/question_answer.html

Reading Comprehension: Methods



SQuAD

- ▶ Single-document, single-sentence question-answering task where the answer is always a substring of the passage
- ▶ Predict start and end indices of the answer in the passage

One of the most famous people born in Warsaw was Maria Skłodowska-Curie, who achieved international recognition for her research on radioactivity and was the first female recipient of the Nobel Prize. Famous musicians include Władysław Szpilman and Frédéric Chopin. Though Chopin was born in the village of Żelazowa Wola, about 60 km (37 mi) from Warsaw, he moved to the city with his family when he was seven months old. Casimir Pulaski, a Polish general and hero of the American Revolutionary War, was born here in 1745.

What was Maria Curie the first female recipient of?
Ground Truth Answers: Nobel Prize Nobel Prize Nobel Prize

What year was Casimir Pulaski born in Warsaw?
Ground Truth Answers: 1745 1745 1745

Who was one of the most famous people born in Warsaw?
Ground Truth Answers: Maria Skłodowska-Curie Maria Skłodowska-Curie Maria Skłodowska-Curie

Feature Based Methods

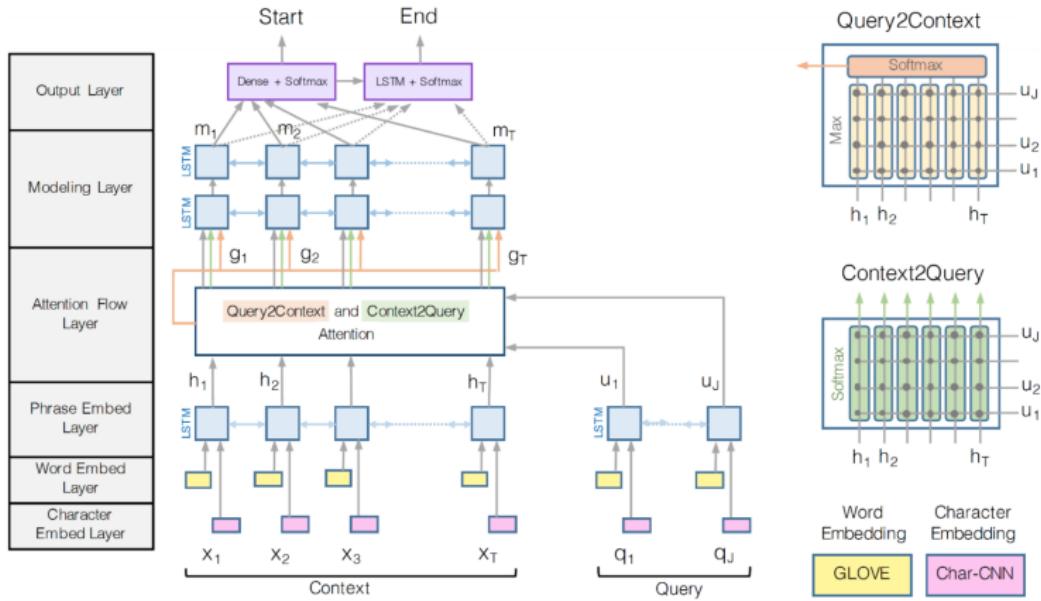
- Generate a list of candidate answers (a_1, a_2, \dots, a_M)
 - Considered only the constituents in parse trees
- Define a feature vector $\phi(p, q, a_i) \in R^d$:
 - Word/bigram features
 - Parse tree matches
 - Dependency labels, length, part-of-speech tags
- Apply a multi-class logistic regression model

(Rajpurkar et al, 2016): SQuAD: 100,000+ Questions for Machine Comprehension of Text

BiLSTM-based Models

- Encode the question using word/char embeddings; pass on an biLSTM encoder
- Encode the passage similarly
- Passage-to-question and question-to-passage attention
- Modeling layer: another BiLSTM layer
- Output layer: two classifiers for predicting start and end points
- The entire model can be trained in an end-to-end way

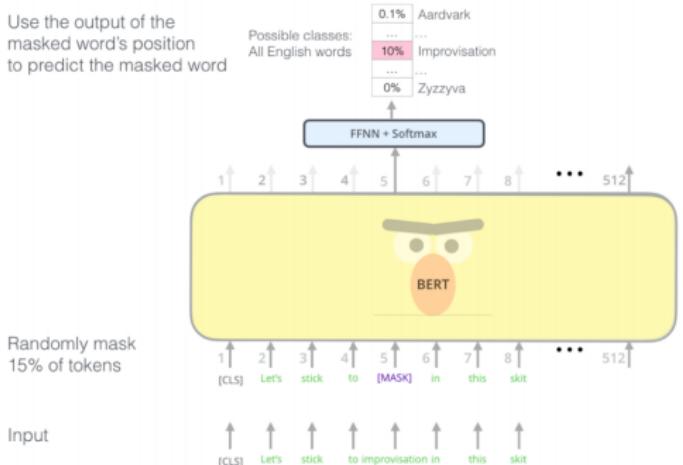
BiDAF



(Seo et al, 2017): Bidirectional Attention Flow for Machine Comprehension

BERT-based Models

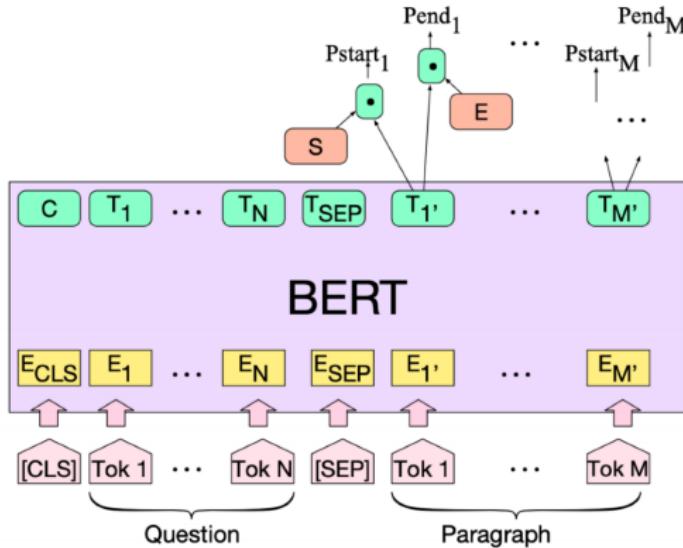
Use the output of the masked word's position to predict the masked word



Pre-training

$$Pstart_i = \frac{e^{S \cdot T_i}}{\sum_j e^{S \cdot T_j}}$$

$$Pend_i = \frac{e^{E \cdot T_i}}{\sum_j e^{E \cdot T_j}}$$



- Concatenate question and passage as one single sequence separated with a **[SEP]** token, then pass it to the BERT encoder
- Train two classifiers on top of passage tokens



SQuAD SOTA: Today

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.831	89.452
1 <small>Sep 18, 2019</small>	ALBERT (ensemble model) Google Research & TTIC https://arxiv.org/abs/1909.11942	89.731	92.215
2 <small>Jul 22, 2019</small>	XLNet + DAAF + Verifier (ensemble) PINGAN Omni-Sinotic	88.592	90.859
2 <small>Sep 16, 2019</small>	ALBERT (single model) Google Research & TTIC https://arxiv.org/abs/1909.11942	88.107	90.902
2 <small>Jul 26, 2019</small>	UPM (ensemble) Anonymous	88.231	90.713
3 <small>Aug 04, 2019</small>	XLNet + SG-Net Verifier (ensemble) Shanghai Jiao Tong University & CloudWalk https://arxiv.org/abs/1908.05147	88.174	90.702
4 <small>Aug 04, 2019</small>	XLNet + SG-Net Verifier++ (single model) Shanghai Jiao Tong University & CloudWalk https://arxiv.org/abs/1908.05147	87.238	90.071

► Performance is very saturated

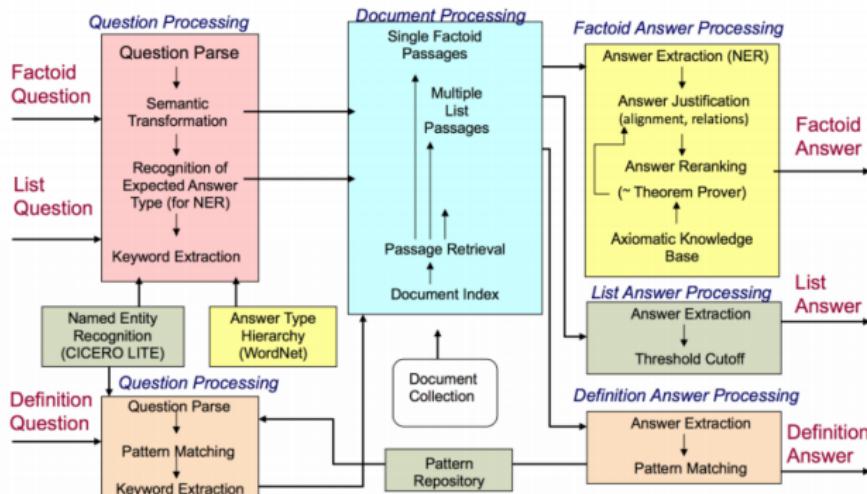
► Harder QA settings are needed!

Open-Domain QA: Methods

IBM's Watson and Jeopardy! Challenge



Traditional QA System



Architecture of LCC
(Harabagiu/Moldovan) QA
system, circa 2003

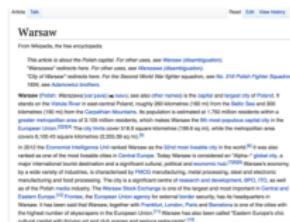
Open-domain QA

SQuAD, TREC, WebQuestions, WikiMovies

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

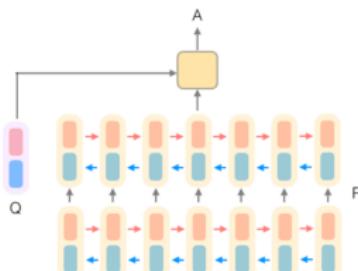


WIKIPEDIA
The Free Encyclopedia



Document Reader

833,500



Open-domain QA

Q: How many of Warsaw's inhabitants spoke Polish in 1933? _____



DrQA: Two-Stage Retriever and Reader

<https://github.com/facebookresearch/DrQA>

Hi!



Hello! Please ask a question.

What is question answering?



a computer science discipline within the fields of information retrieval and natural language processing

Who was the winning pitcher in the 1956 World Series?



Don Larsen

What is the answer to life, the universe, and everything?

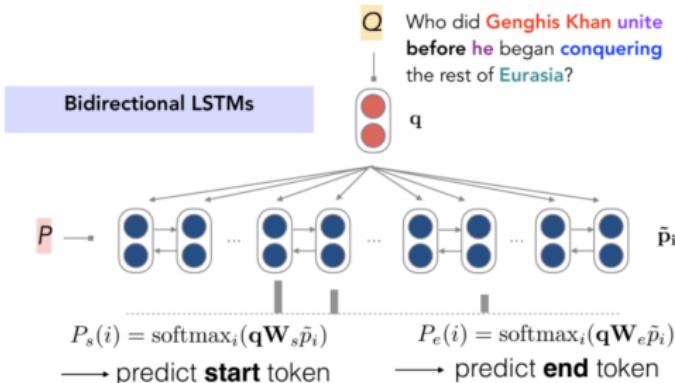


42

Document Retriever: Two Steps

1. TF-IDF bag-of-words vectors
2. Efficient bigram hashing (Weinberger et al., 2009)
 - Map the bigram to 2^{24} bins with an unsigned murmur3 hash
 - Preserving speed and memory efficiency
 - Murmur3: map a word or string to a 32-bit or 128 bit value
 - Online: <http://murmurhash.shorelabs.com/>

Document Reader

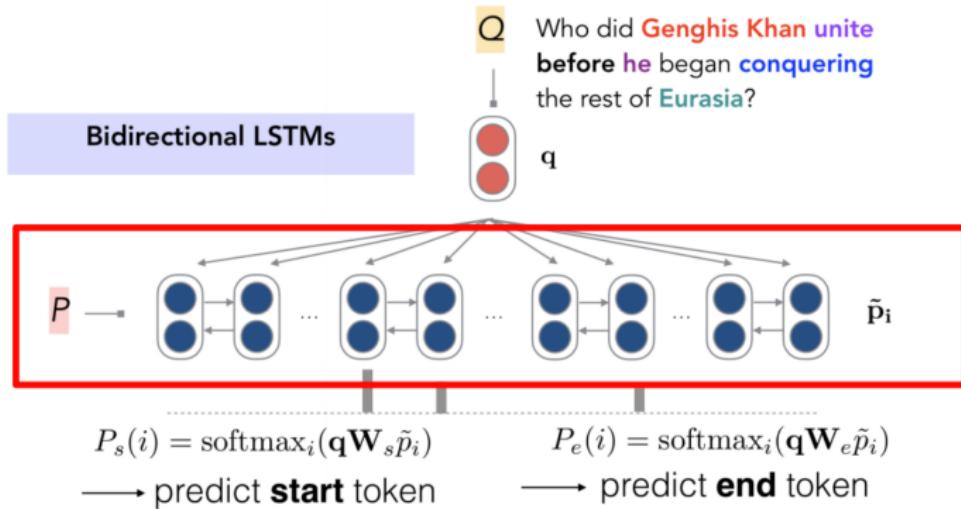


Three steps:

1. Paragraph encoding
2. Question encoding
3. Prediction

similar to AttentiveReader (Hermann et al, 2015;
Chen et al, 2016)

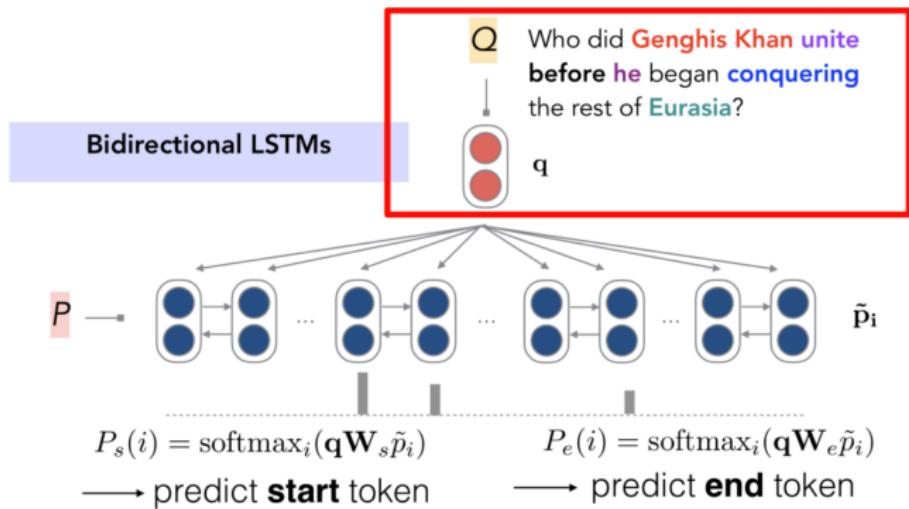
Document Reader: Paragraph Encoding



Document Reader: Paragraph Encoding

1. Represent tokens in a paragraph as a sequence of feature vectors
 - Word embedding
 - Exact match
 - Token features
 - Aligned question embedding
2. Pass features as the input to a RNN (multi-layer Bidirectional LSTM)

Document Reader: Question Encoding



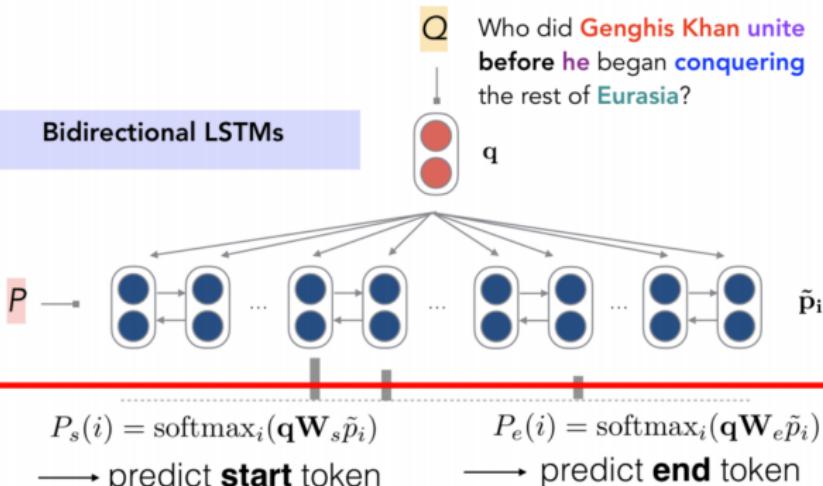
Document Reader: Question Encoding

1. Apply another RNN to top of word embeddings of q_i and get \mathbf{q}_i
2. Combining the resulting units into one single vector

$$\mathbf{q} = \sum_j b_j \mathbf{q}_j$$

Here $b_j = \frac{\exp(\mathbf{w} \cdot \mathbf{q}_j)}{\sum_{j'} \exp(\mathbf{w} \cdot \mathbf{q}_{j'})}$ and \mathbf{w} is a weight vector to learn

Document Reader: Prediction



Document Reader: Prediction

- Goal: predict the span of tokens that is most likely the correct answer
- Method: train two classifiers independently for predicting ends of span

$$\max_{i,j} P_{start}(i) \times P_{end}(j)$$

such that $i \leq j \leq i + 15$ and $P_{start}(i)/P_{end}(j)$ is probability of each token being start/end

$$P_{start}(i) \propto \exp(p_i W_s \mathbf{q})$$

$$P_{end}(i) \propto \exp(p_i W_e \mathbf{q})$$

Dataset and Example Training Data

Dataset	Example	Article / Paragraph
SQuAD	Q: How many provinces did the Ottoman empire contain in the 17th century? A: 32	Article: Ottoman Empire Paragraph: ... At the beginning of the 17th century the empire contained 32 provinces and numerous vassal states. Some of these were later absorbed into the Ottoman Empire, while others were granted various types of autonomy during the course of centuries.
CuratedTREC	Q: What U.S. state's motto is "Live free or Die"? A: New Hampshire	Article: Live Free or Die Paragraph: "Live Free or Die" is the official motto of the U.S. state of New Hampshire , adopted by the state in 1945. It is possibly the best-known of all state mottos, partly because it conveys an assertive independence historically found in American political philosophy and partly because of its contrast to the milder sentiments found in other state mottos.
WebQuestions	Q: What part of the atom did Chadwick discover? A: neutron	Article: Atom Paragraph: ... The atomic mass of these isotopes varied by integer amounts, called the whole number rule. The explanation for these different isotopes awaited the discovery of the neutron , an uncharged particle with a mass similar to the proton, by the physicist James Chadwick in 1932. ...
WikiMovies	Q: Who wrote the film Gigli? A: Martin Brest	Article: Gigli Paragraph: Gigli is a 2003 American romantic comedy film written and directed by Martin Brest and starring Ben Affleck, Jennifer Lopez, Justin Bartha, Al Pacino, Christopher Walken, and Lainie Kazan.

Result

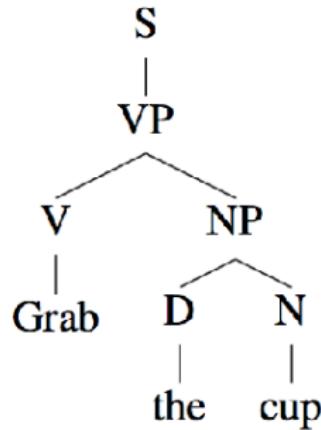
Dataset	YodaQA	DrQA		
		SQuAD	+Fine-tune (DS)	+Multitask (DS)
SQuAD (<i>All Wikipedia</i>)	n/a	27.1	28.4	29.8
CuratedTREC	31.3	19.7	25.7	25.4
WebQuestions	39.8	11.8	19.5	20.7
WikiMovies	n/a	24.5	34.3	36.5

Table 6: Full Wikipedia results. Top-1 exact-match accuracy (in %, using SQuAD eval script). +Fine-tune (DS): Document Reader models trained on SQuAD and fine-tuned on each DS training set independently. +Multitask (DS): Document Reader single model trained on SQuAD and all the distant supervision (DS) training sets jointly. YodaQA results are extracted from <https://github.com/brmson/yodaqa/wiki/Benchmarks> and use additional resources such as Freebase and DBpedia, see Section 2.

Quick Summary: DrQA

- DrQA was the first attempt to scale up reading comprehension to open-domain question answering, by **combining IR techniques and neural reading comprehension models.**
- Although we achieved good accuracy on SQuAD in 2017 (EM = 70.. vs state-of-the-art EM = 90 in 2020), the final QA accuracy still remains low: 20.7 - 36.5.
- Distant supervision + multi-task learning helps!

Semantic Parsing



- “Grab” = execute GrabbingFunction()
- “the cup” = object ID 9AF1948A81CD22

Following directions



Linguistic	Non-linguistic
Verbs like <i>grab, release</i>	Actions a robot can execute
Nouns like <i>cup, block</i>	Specific entities in the world
transitive VP (V NP)	An action executed upon an object

Question answering

Barack Obama was the 44th President of the United States. Obama was born on August 4 in Honolulu, Hawaii. In late August 1961, Obama's mother moved with him to the University of Washington in Seattle for a year...

Was Barack Obama born in the United States?

Non-linguistic

Hawaii is a place

Hawaii is a place within the United States

“born in X” entails “from x”

Meaning representation

Geo	which state has the most rivers running through it? $(\text{argmax } \$0 \ (\text{state}:t \ \$0) \ (\text{count } \$1 \ (\text{and} \ (\text{river}:t \ \$1) \ (\text{loc}:t \ \$1 \ \$0))))$
ATIS	all flights from dallas before 10am $(\text{lambda } \$0 \ e \ (\text{and} \ (\text{flight } \$0) \ (\text{from } \$0 \ \text{dallas:ci}) \ (< \ (\text{departure time } \$0 \ 1000:\text{ti})))$
SQL	What record company did conductor Mikhail Snitko record for after 1996? $\text{SELECT Record Company WHERE (Year of Recording > 1996) AND (Conductor = Mikhail Snitko)}$
Django	if length of bits is lesser than integer 3 or second element of bits is not equal to string 'as' $\text{if len(bits) < 3 or bits[1] != 'as':}$



SQL Generation

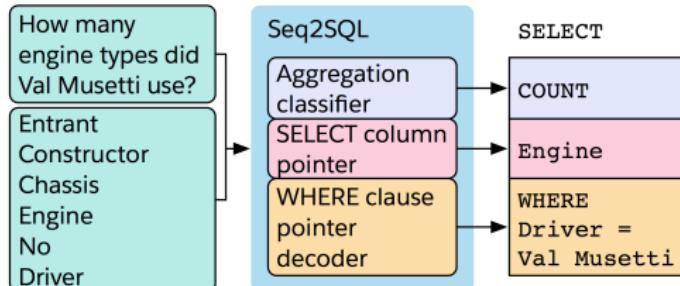
- ▶ Convert natural language description into a SQL query against some DB
- ▶ How to ensure that well-formed SQL is generated?
 - ▶ Three seq2seq models
- ▶ How to capture column names + constants?
 - ▶ Pointer mechanisms, to be discussed later

Question:

How many CFL teams are from York College?

SQL:

```
SELECT COUNT CFL Team FROM  
CFLDraft WHERE College = "York"
```

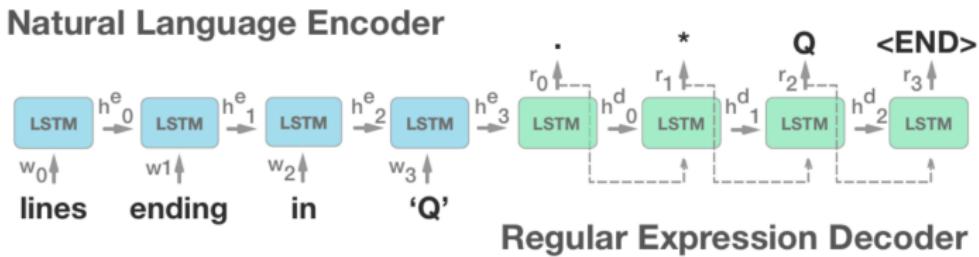


Zhong et al. (2017)



Regex Prediction

- Predict regex from text



- Problem: requires a lot of data: 10,000 examples needed to get ~60% accuracy on pretty simple regexes
- Does not scale when regex specifications are more abstract (*I want to recognize a decimal number less than 20*)

Locascio et al. (2016)



Semantic Parsing as Translation

"what states border Texas"



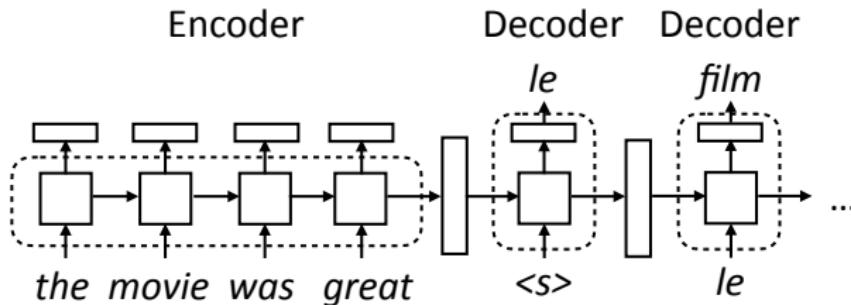
lambda x (state(x) and border(x , e89)))

- ▶ Write down a linearized form of the semantic parse, train seq2seq models to directly translate into this representation
- ▶ What are some benefits of this approach compared to grammar-based?
- ▶ What might be some concerns about this approach? How do we mitigate them?

Jia and Liang (2016)



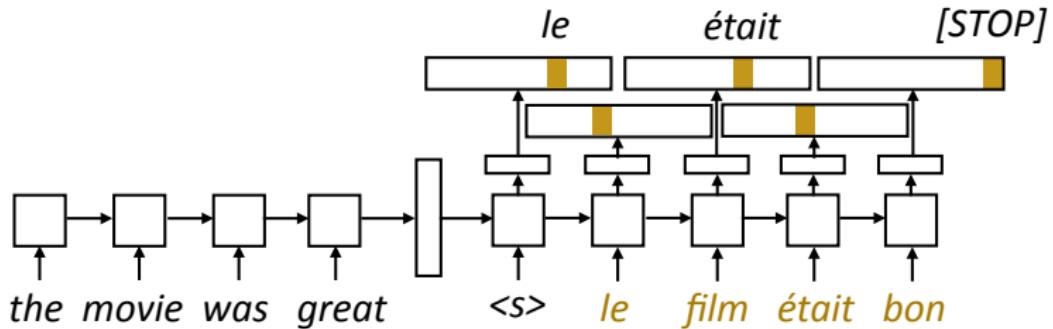
Implementing seq2seq Models



- ▶ Encoder: consumes sequence of tokens, produces a vector. Analogous to encoders for classification/tagging tasks
- ▶ Decoder: separate module, single cell. Takes two inputs: hidden state (vector h or tuple (h, c)) and previous token. Outputs token + new state



Training

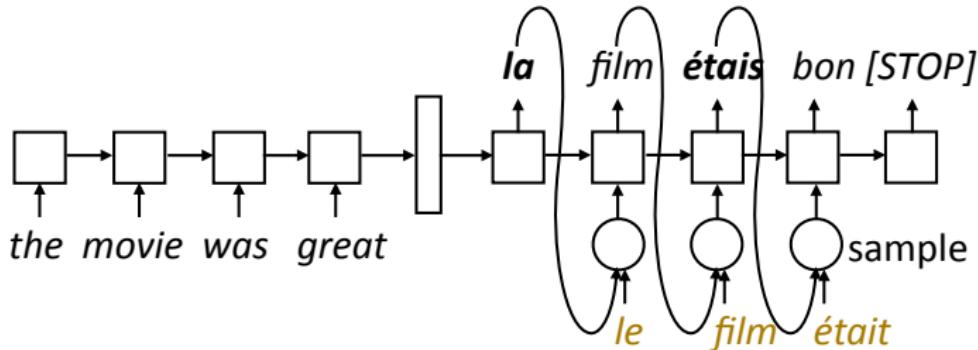


- ▶ Objective: maximize $\sum_{(\mathbf{x}, \mathbf{y})} \sum_{i=1}^n \log P(y_i^* | \mathbf{x}, y_1^*, \dots, y_{i-1}^*)$
- ▶ One loss term for each target-sentence word, feed the correct word regardless of model's prediction (called “teacher forcing”)



Training: Scheduled Sampling

- ▶ Model needs to do the right thing even with its own predictions



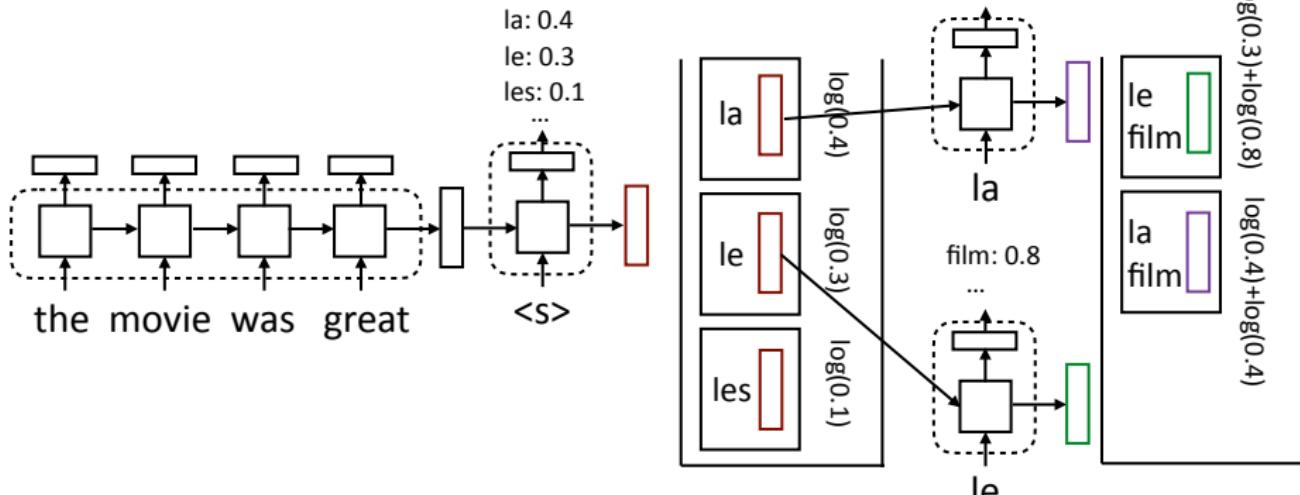
- ▶ Scheduled sampling: with probability p , take the gold as input, else take the model's prediction
- ▶ Starting with $p = 1$ (teacher forcing) and decaying it works best
- ▶ “Right” thing: train with reinforcement learning

Bengio et al. (2015)



Beam Search

- Maintain decoder state, token history in beam



- Keep both *film* states! Hidden state vectors are different

Semantic Parsing

- Seq2seq for semantic parsing is a good baseline
- We will see many other methods later in the course
- Many different possible meaning representations: first-order logic, Abstract Meaning Representation (AMR), SQL

Shallow Semantic Parsing: Semantic Role Labeling (SRL)

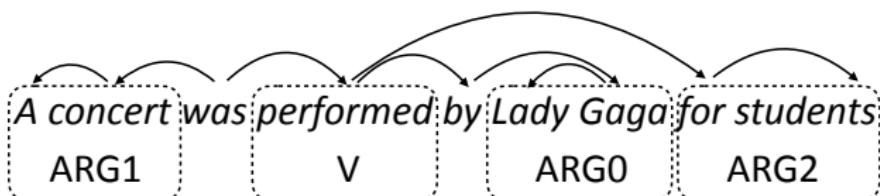
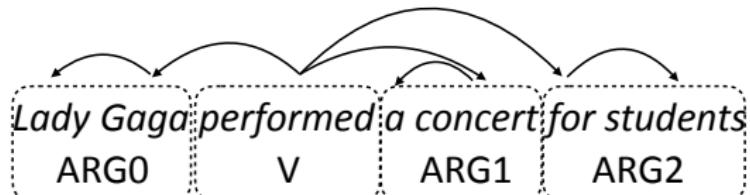
Semantic Role Labeling

- Rather than parse to a logic/executable semantic representation, let's try something easier: just label the semantics of events (**shallow semantic parsing**)



Semantic Role Labeling

- ▶ 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"

Semantic Role Labeling

- ARG0: Agent
- ARG1: Patient
- ARG2+: Semantics vary
- Each event has it's own semantic roles

Semantic Role Labeling

- Easier task for humans to annotate
- Can annotate large amounts of data to train SRL systems
- Useful for downstream tasks (more abstract than syntax)



Semantic Role Labeling

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

Gold

ARG1

V

ARG2

ARG3

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

quicken:

Arg0-PAG: *causer 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)



Semantic Role Labeling

- ▶ 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

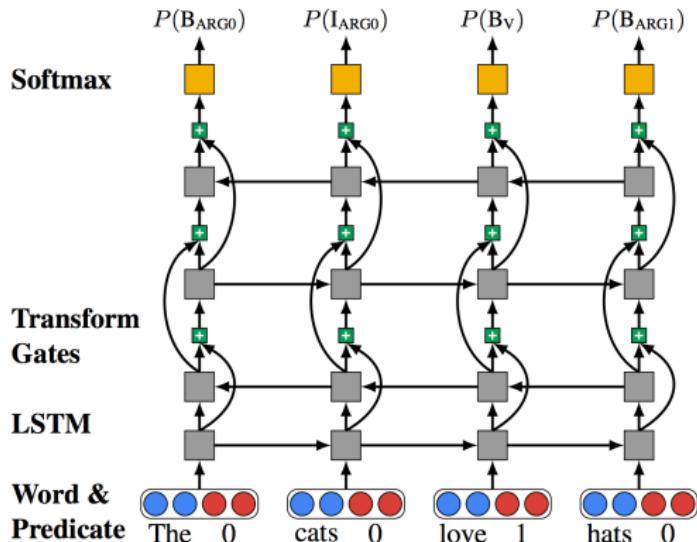


Figure from He et al. (2017)

SRL Datasets

- FrameNet
- PropBank
- OntoNotes

Information extraction (IE)

Investigating(SEC, Tesla)

 INVESTOPEDIA Topics ▾ Reference ▾ Advisors ▾ Markets ▾ Simulator ▾ Academy NEW

LEARN TO INVEST

Don't know where to start? Sign up
for the Investing Basics newsletter.



SEC Probing Tesla CEO Musk's Tweets: Reports

By Deborah DSouza | Updated August 9, 2018 — 4:47 AM EDT

On Tuesday, Tesla Inc. ([TSLA](#)) CEO Elon Musk made the dramatic announcement that he was considering taking Tesla private for \$420 a share on Twitter. In an email sent to Tesla employees [posted on the company's official blog](#), Musk explained that he is mulling taking the firm private to protect it from short sellers and wild swings in stock prices. However, the email didn't provide any details regarding financing. (See also: [What if Tesla Goes Private?](#))

Withdraws-from(Sanders, US Presidential Race)

Bernie Sanders Drops Out of 2020 Democratic Race for President



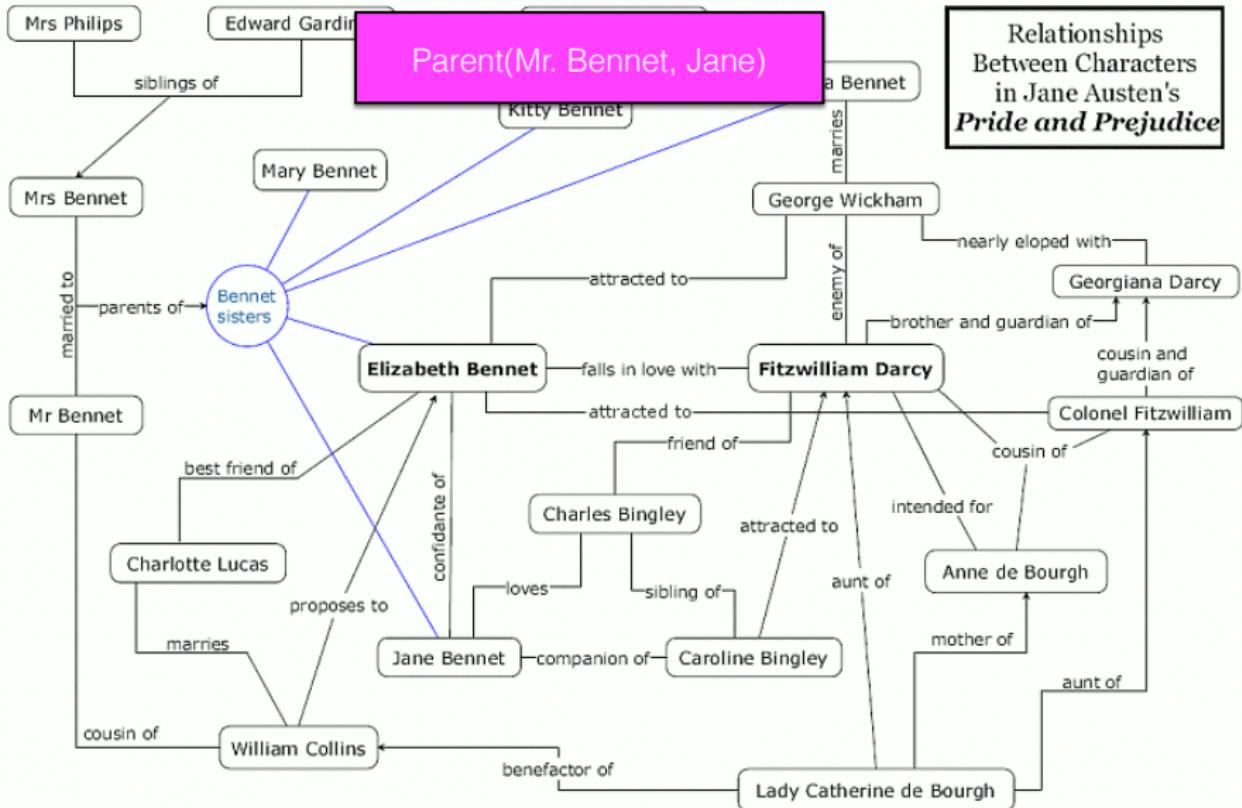
By Sydney Ember

April 8, 2020



Senator [Bernie Sanders](#) of Vermont ended his presidential candidacy on Wednesday, concluding a quest that elevated him as a standard-bearer of American liberalism and clearing the way for a general election between the presumptive Democratic nominee, Joseph R. Biden Jr., and President Trump at a time of national crisis.

Relationships
Between Characters
in Jane Austen's
Pride and Prejudice



Information extraction

- Named entity recognition
- Entity linking
- Relation extraction

Named entity recognition

[tim cook]_{PER} is the ceo of [apple]_{ORG}

- Identifying spans of text that correspond to typed entities

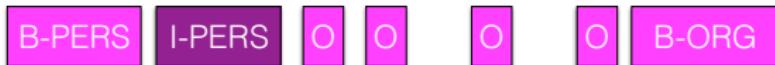
Named entity recognition

Type	Tag	Sample Categories	Example sentences
People	PER	people, characters	Turing is a giant of computer science.
Organization	ORG	companies, sports teams	The IPCC warned about the cyclone.
Location	LOC	regions, mountains, seas	The Mt. Sanitas loop is in Sunshine Canyon.
Geo-Political Entity	GPE	countries, states, provinces	Palo Alto is raising the fees for parking.
Facility	FAC	bridges, buildings, airports	Consider the Golden Gate Bridge.
Vehicles	VEH	planes, trains, automobiles	It was a classic Ford Falcon.

Figure 17.1 A list of generic named entity types with the kinds of entities they refer to.

ACE NER categories (+weapon)

BIO notation



tim cook is the ceo of apple

- **B**eginning of entity
- **I**nside entity
- **O**utside entity

[tim cook]_{PER} is the ceo of [apple]_{ORG}

Named entity recognition

B-PERS B-PERS

After he saw Harry Tom went to the store

Named entity recognition

- Most **named** entity recognition datasets have flat structure (i.e., non-hierarchical labels).
 - ✓ [The University of California]**ORG**
 - ✗ [The University of [California]**GPE**]**ORG**
- Mostly fine for **named** entities, but more problematic for general entities:

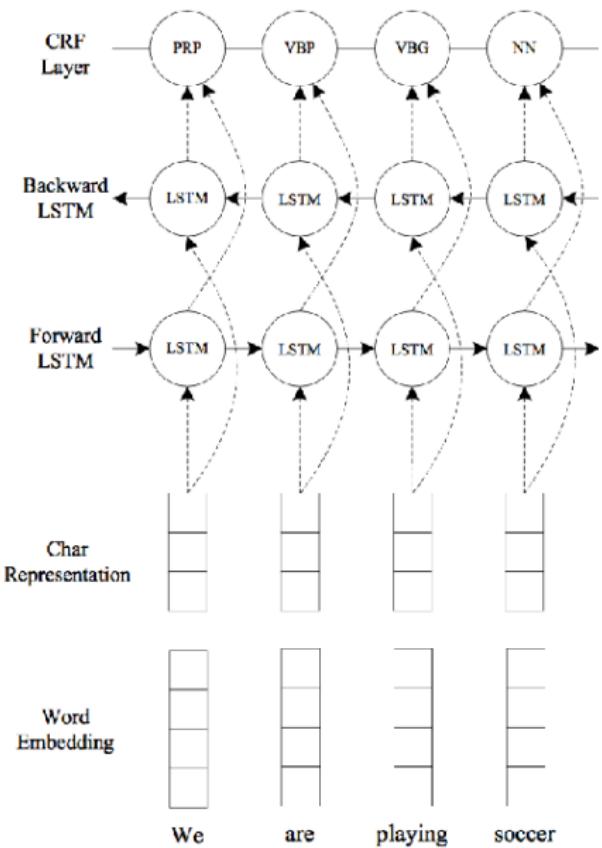
[[John]**PER**'s mother]**PER** said ...

Sequence labeling

$$x = \{x_1, \dots, x_n\}$$

$$y = \{y_1, \dots, y_n\}$$

- For a set of inputs x with n sequential time steps, one corresponding label y_i for each x_i
- Model correlations in the labels y .



Ma and Hovy (2016), "End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF"

Gazetteers

- List of place names; more generally, list of names of some typed category
- GeoNames (GEO), US SSN (PER), Getty Thesaurus of Geographic Placenames, Getty Thesaurus of Art and Architecture

Bun Cranncha
Dromore West
Dromore
Youghal Harbour
Youghal Bay
Youghal
Eochaill
Yellow River
Yellow Furze
Woodville
Wood View
Woodtown House
Woodstown
Woodstock House
Woodsgift House
Woodrooff House
Woodpark
Woodmount
Wood Lodge
Woodlawn Station
Woodlawn
Woodlands Station
Woodhouse
Wood Hill
Woodfort
Woodford River
Woodford
Woodfield House
Woodenbridge Junction Station
Woodenbridge
Woodbrook House
Woodbrook
Woodbine Hill
Wingfield House
Windy Harbour
Wind... G

Evaluation

- We evaluate NER with precision/recall/F1 over typed chunks.

Evaluation

	1	2	3	4	5	6	7
	tim	cook	is	the	CEO	of	Apple
<i>gold</i>	B-PER	I-PER	O	O	O	O	B-ORG
<i>system</i>	B-PER	O	O	O	B-PER	O	B-ORG

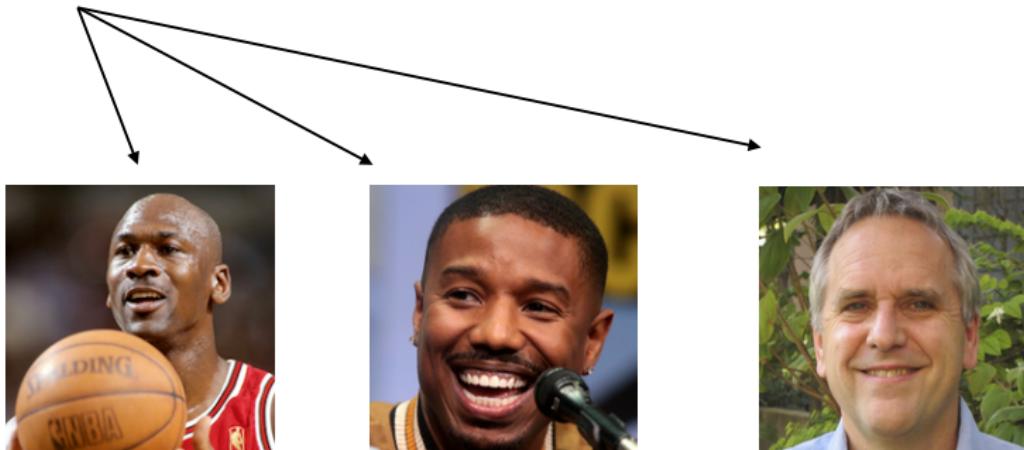
<start, end, type>

Precision	1/3
Recall	1/2

gold
<1,2,PER>
system
<1,1,PER>
<5,5,PER>
<7,7,ORG>

Entity linking

Michael	Jordan	can	dunk	from	the	free	throw	line
B-PER	I-PER							



Entity linking

- Task: Given a database of candidate referents, identify the correct referent for a mention in context.

Text	True wikipedia page
Hornets owner Michael Jordan thinks having one or two “superteams” is a detriment to the NBA because the other 28 teams “are going to be garbage.”	wiki/Michael_Jordan
In 2001, Michael Jordan and others resigned from the Editorial Board of <i>Machine Learning</i> .	wiki/Michael_I._Jordan
The stars are aligning for leading man Michael Jordan , who just signed on for a new film, according to Variety.	wiki/Michael_B._Jordan
Michael Jordan played in 1,072 regular-season games in his 15-season career	wiki/Michael_Jordan

Learning to rank

- Entity linking is often cast as a learning to rank problem: given a mention x , some set of candidate entities $\mathcal{Y}(x)$ for that mention, and context c , select the **highest scoring** entity from that set.

$$\hat{y} = \arg \max_{y \in \mathcal{Y}(x)} \Psi(y, x, c)$$

Eisenstein 2018

Some scoring function
over the mention x ,
candidate y , and context c

Learning to rank

Some scoring function
over the mention x,
candidate y, and context c

$$\Psi(y, x, c)$$

feature = $f(x, y, c)$
string similarity between x and y
popularity of y
NER type(x) = type(y)
cosine similarity between c and Wikipedia page for y

$$\Psi(y, x, c) = f(x, y, c)^T \beta$$

Neural learning to rank

Parameters measuring the compatibility of
the candidate and mention

Parameters measuring the compatibility of
the candidate and context

$$\Psi(y, x, c) = v_y^T \Theta^{(x,y)} x + v_y^T \Theta^{(y,c)} c$$

Embedding
for candidate

Embedding
for mention

Embedding
for context



Relation extraction

The Big Sleep is a 1946 film noir directed by [Howard Hawks](#),^{[2][3]} the first film version of [Raymond Chandler](#)'s 1939 [novel of the same name](#). The film stars [Humphrey Bogart](#) as private detective [Philip Marlowe](#) and [Lauren Bacall](#) as Vivian Rutledge in a story about the "process of a criminal investigation, not its results."^[4] [William Faulkner](#), [Leigh Brackett](#) and [Jules Furthman](#) co-wrote the screenplay.

<i>subject</i>	<i>predicate</i>	<i>object</i>
The Big Sleep	directed_by	Howard Hawks
The Big Sleep	stars	Humphrey Bogart
The Big Sleep	stars	Lauren Bacall
The Big Sleep	screenplay_by	William Faulkner
The Big Sleep	screenplay_by	Leigh Brackett
The Big Sleep	screenplay_by	Jules Furthman

Relation extraction

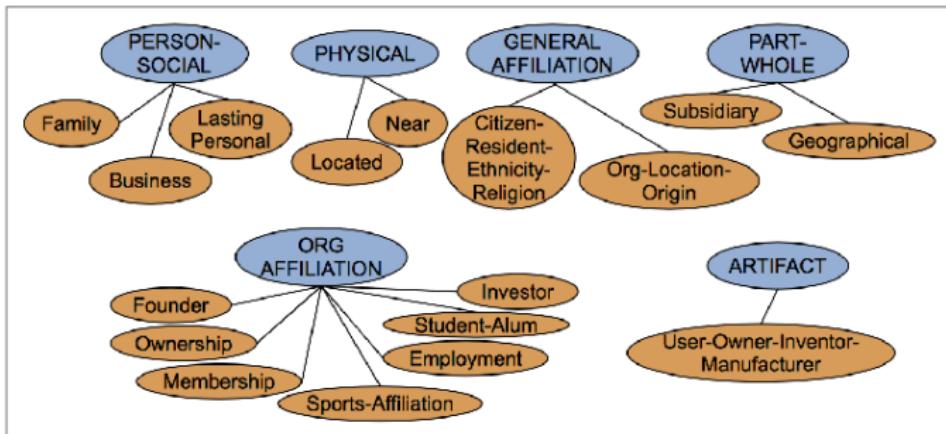


Figure 17.9 The 17 relations used in the ACE relation extraction task.

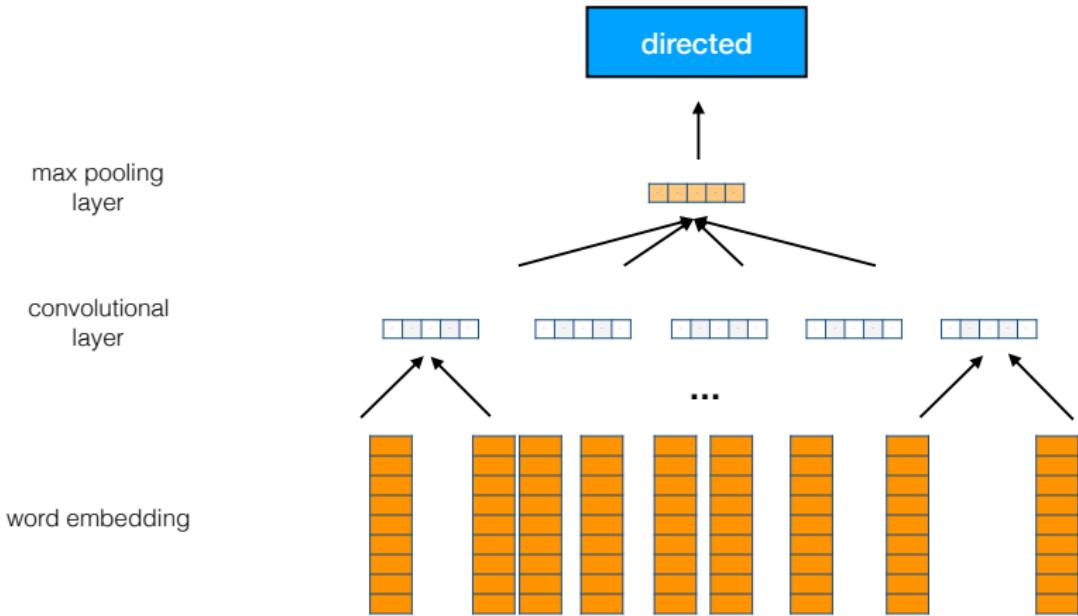
ACE relations, SLP3

Regular expressions

- Regular expressions are precise ways of extracting high-precision relations
- “NP₁ is a film directed by NP₂” → *directed_by*(NP₁, NP₂)
- “NP₁ was the director of NP₂” → *directed_by*(NP₂, NP₁)

Hearst patterns

<i>pattern</i>	<i>sentence</i>
NP {, NP}* {,} (and or) other NP _H	temples, treasuries, and other important civic buildings
NP _H such as {NP,*} {(or and)} NP	red algae such as Gelidium
such NP _H as {NP,*} {(or and)} NP	such authors as Herrick, Goldsmith, and Shakespeare
NP _H {,} including {NP,*} {(or and)} NP	common-law countries , including Canada and England
NP _H {,} especially {NP,*} {(or and)} NP	European countries , especially France, England, and Spain



[The Big Sleep]_{m1} is a 1946 film noir directed by [Howard Hawks]_{m2}

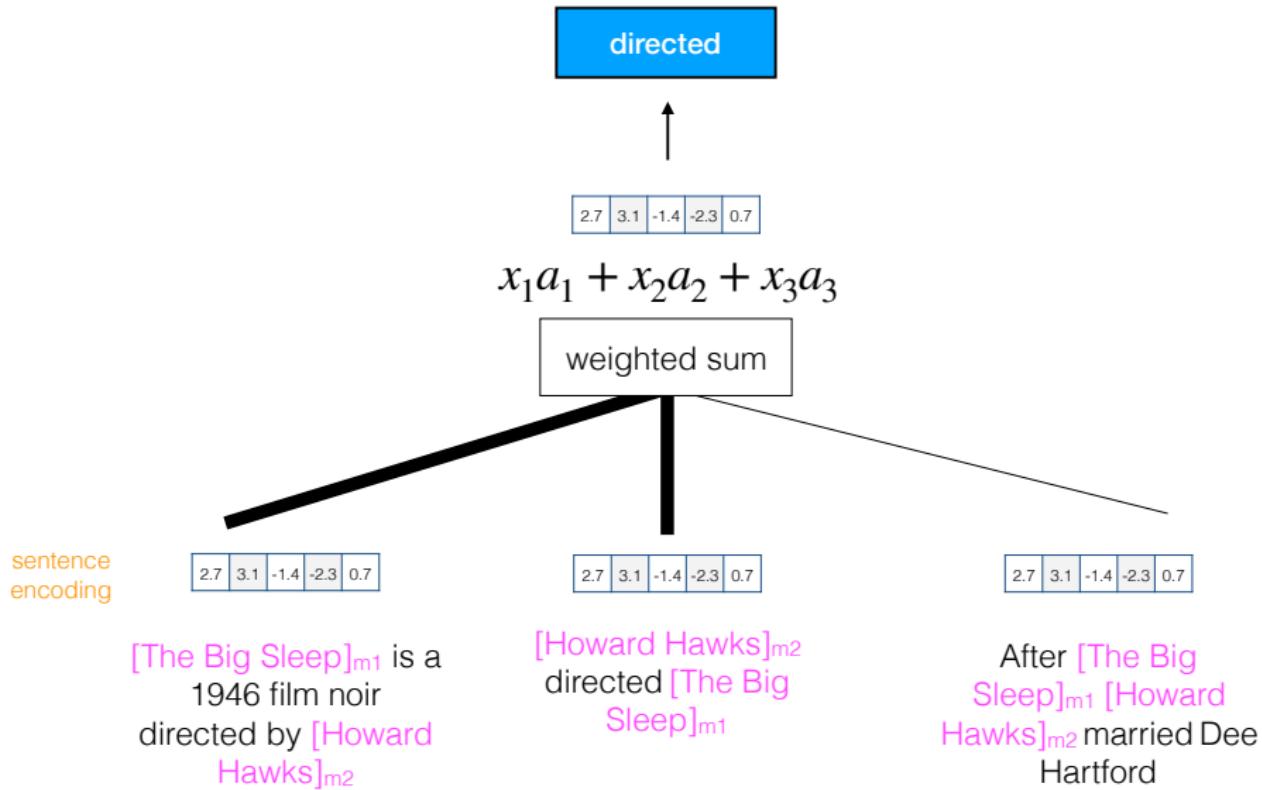
Distant supervision

- It's uncommon to have labeled data in the form of <sentence, relation> pairs

<i>sentence</i>	<i>relations</i>
[The Big Sleep] _{m1} is a 1946 film noir directed by [Howard Hawks] _{m2} , the first film version of Raymond Chandler's 1939 novel of the same name.	directed_by(The Big Sleep, Howard Hawks)

Distant supervision

- More common to have knowledge base data about entities and their relations that's separate from text.
- We know the text likely expresses the relations somewhere, but not **exactly where**.



Information Extraction

- Named entity recognition
- Entity linking
- Relation extraction
- Template filling
- Event detection
- Event coreference
- Extra-propositional information (veridicality, hedging)