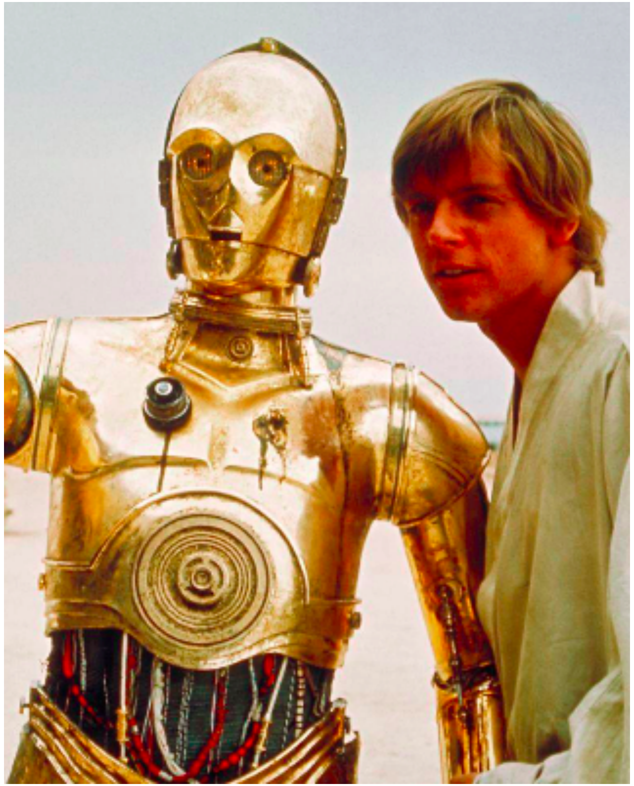
Understanding Natural Language Understanding

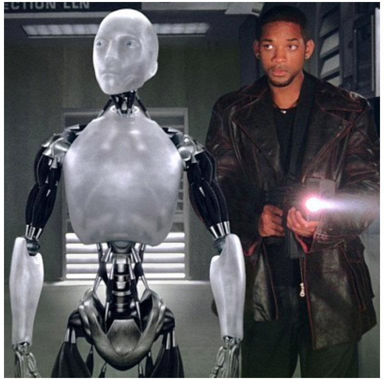
Bill MacCartney

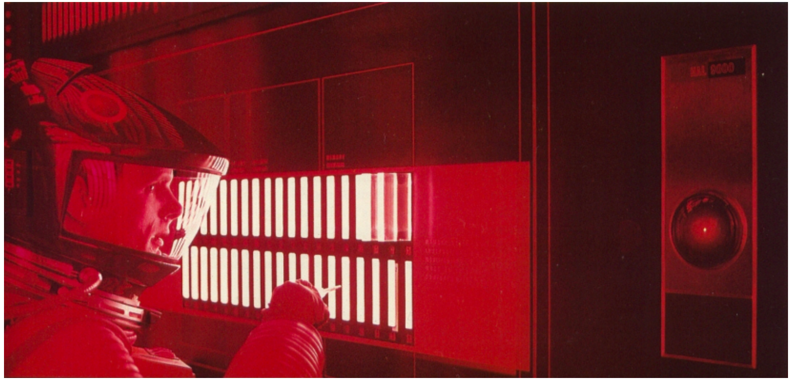
ACM SIGAI Bay Area Chapter Inaugural Meeting

16 July 2014

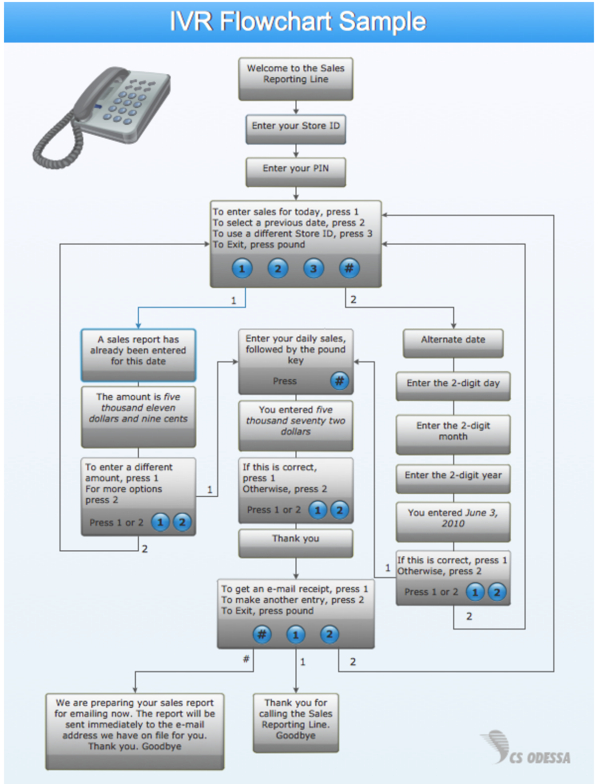
How we imagine talking to computers







How we actually talk to computers



Operator.

OPERATOR.

OPERATOR!!

AGENT!!!

HAL-9000

In the 1967 Stanley Kubrick movie *2001: A Space Odyssey*, the spaceship's computer HAL can

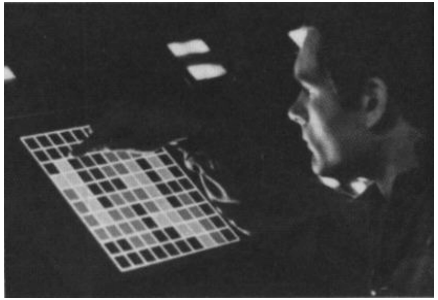
- display graphics;
- play chess; and
- conduct natural, open-domain conversations with humans.

How well did the filmmakers do at predicting what computers would be capable in 2001?

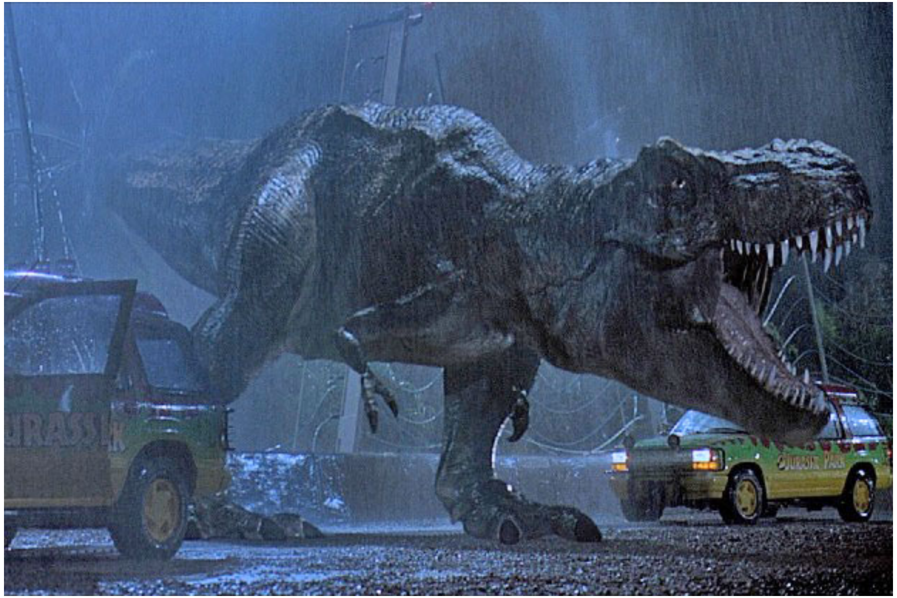
example from Andrew McCallum

HAL-9000: graphics

HAL



Jurassic Park (1993)



example from Andrew McCallum

HAL-9000: chess

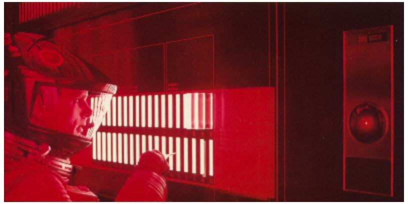
HAL

Deep Blue (1997)

example from Andrew McCallum

HAL-9000: natural language understanding

HAL



David Bowman: Open the pod bay doors, HAL.

HAL: I'm sorry, Dave, I'm afraid I can't do that.

David: What are you talking about, HAL?

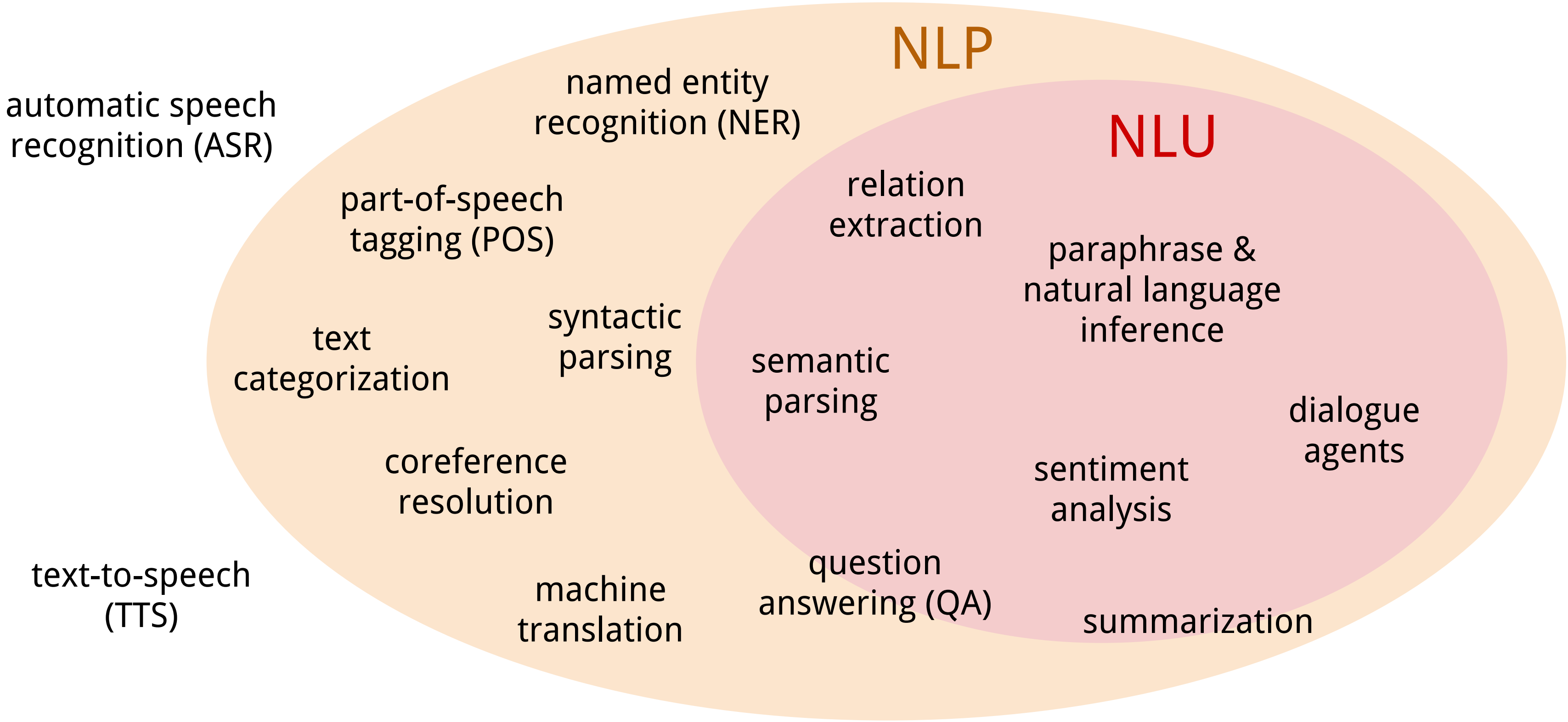
HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

Siri (2011)

Colbert: … I don't want to search for anything! I want to write the show!
Siri: Searching the Web for "search for anything. I want to write the shuffle."
Colbert: … For the love of God, the cameras are on, give me something?
Siri: What kind of place are you looking for? Camera stores or churches?

example from Andrew McCallum

Terminology: NLU vs. NLP vs. ASR



Outline

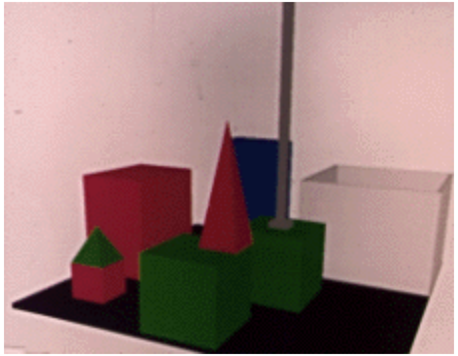
- Introduction
- NLU yesterday and today
- Relation extraction
- Semantic parsing
- Vector-space semantics
- Sentiment analysis

A brief history of NLU

- 1960s: Pattern-matching with small rule-sets
- 1970-80s: Linguistically rich, logic-driven, grounded systems; restricted applications
- 1990s: the “statistical revolution” in NLP leads to a decrease in NLU work
- 2010s: NLU returns to center stage, mixing techniques from previous decades

SHRDLU (Winograd 1972)

H. What does the box contain?
C. THE BLUE PYRAMID AND THE BLUE BLOCK.
H. What is the pyramid supported by?
C. THE BOX.
H. How many blocks are not in the box?
C. FOUR OF THEM.
H. Is at least one of them narrower than the one which I told you to pick up?
C. YES, THE RED CUBE.
H. Is it supported?
C. YES, BY THE TABLE.



http://youtube.com/watch?v=8SvD-lNg0TA
http://hci.stanford.edu/winograd/shrdlu/

CHAT-80 (Pereira 1980)

Is there more than one country in each continent?	No.
What are the countries from which a river flows into the Black_Sea?	[romania].
What is the total area of countries south of the Equator and not in Australasia?	10239 ksqmiles.
Which country bordering the Mediterranean borders a country that is bordered by a country whose population exceeds the population of India?	turkey.
What countries border Denmark?	west_germany.
What countries is Denmark adjacent to?	I don’t understand!

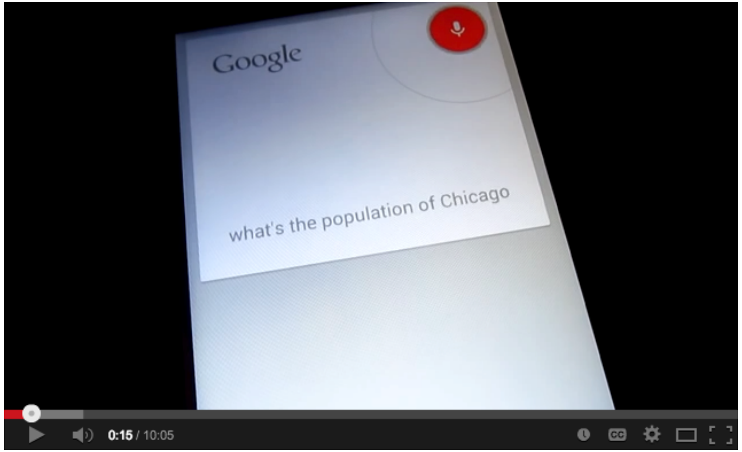
NLU today

An explosion of interest, in both academia and industry:

- voice-driven assistants (Siri, Google Now, Microsoft Cortana)
- natural-language search (Google, Facebook Graph Search)
- question answering (Google, IBM's Watson, Wolfram Alpha)
- web-scale relation extraction (Google, many startups)
- sentiment analysis for automated trading (many hedge funds)
- legal discovery (Cataphora, H5)
- business intelligence (Palantir, Quid)
- social media analytics (a zillion startups)
- content summarization (Summly, other startups)

Conversational search at Google

what's the population of Chicago
who's the mayor
how old is he
who is he married to



https://www.youtube.com/watch?v=yiQX-_Y0gms

OK Google, where am I
how is traffic in San Diego
show me things to do there
when did the San Diego Zoo open
is it open
how far is it
call them

when is Thanksgiving
I meant the Canadian one

IBM’s Watson

NLU in automated trading

- Most financial trading is now done by automated systems (especially in “high-frequency trading”, or HFT)
- Many trading strategies rely in part on automated analysis of unstructured data feeds: newswires, analyst reports, etc.
- You can make vast profits if you can discover and act on market-moving news a few milliseconds faster than rivals
- Essentially, they’re using NLU to predict the markets

The 2008 United Airlines "bankruptcy"

- Newspaper accidentally republished old bankruptcy story
- Automated trading reacted within seconds
- $1B in market value evaporated within 12 minutes

Read more at http://nyti.ms/1dBzJSK

The 2013 @AP Twitter hack

@AP Twitter feed hacked.

Within seconds,
Dow plunged 140 points.

Recovered in 6 minutes.

S&P 500 temporarily lost
$136B in market cap!

Oops.

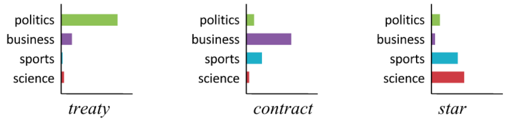
The 2013 @AP Twitter hack

The rapid fire trading also highlights the role of computers and algorithmic trading on Wall Street. "That goes to show you how algorithms read headlines and create these automatic orders — you don't even have time to react as a human being," said Kenny Polcari of O'Neill Securities, on Power Lunch. "I'd imagine the SEC's going to look into how this happens. It's not about banning computers, but it's about protection and securing our markets."

http://www.cnbc.com/id/100646197

Semantic representations

One way of organizing NLU topics: by output representation

<table>
<tr><td>relation extraction</td><td rowspan="2">discrete</td><td>relation instances /
database triples</td><td>(Larry Page, founder, Google)
(Google, located in, Mountain View)</td></tr>
<tr><td>semantic parsing</td><td>logical forms /
other rich structures</td><td>argmax(λx.state(x), λx.size(x))</td></tr>
<tr><td>sentiment analysis</td><td rowspan="2">continuous</td><td>scalars</td><td></td></tr>
<tr><td>vector space models</td><td>vectors /
topic distributions</td><td></td></tr>
</table>

Outline

- Introduction
- NLU yesterday and today
- Relation extraction
- Semantic parsing
- Vector-space semantics
- Sentiment analysis

Relation extraction example

CHICAGO (AP) — Citing high fuel prices, United Airlines said Friday it has increased fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL, said the increase took effect Thursday night and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Atlanta and Denver to San Francisco, Los Angeles and New York.

Subject	Relation	Object
American Airlines	subsidiary	AMR
Tim Wagner	employee	American Airlines
United Airlines	subsidiary	UAL

example from Jim Martin

Extending the Knowledge Graph

Google’s Knowledge Graph: >500M entities, >40B relationships

Curation is an ongoing challenge — things change!

Relies heavily on relation extraction from the web

/film/film/starring	
Bad Words	Jason Bateman
Divergent	Shailene Woodley
Non-Stop	Liam Neeson

/organization/organization/parent	
WhatsApp	Facebook
Nest Labs	Google
Nokia	Microsoft

/music/artist/track	
Macklemore	White Privilege
Phantogram	Mouthful of Diamonds
Lorde	Royals

/people/person/date_of_death	
Nelson Mandela	2013-12-05
Paul Walker	2013-11-30
Lou Reed	2013-10-27

Approaches to relation extraction

Many possible approaches to the problem:

- Hand-built patterns (80s and 90s)
- Bootstrapping methods (late 90s)
- Supervised methods (late 90s to late 00s)
- Unsupervised methods (mid 00s to present)

The one I'll focus on is *distant supervision*:

- Use relation instances you already know about to learn how the relation is expressed in text
- Then apply that to discover new relation instances

Distant supervision: training data

Knowledge Graph

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Web documents

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, …
Bill Gates attended Harvard from…
Google was founded by Larry Page …

Mintz et al. 2009

Training data

Distant supervision: training data

Knowledge Graph

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y

Web documents

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, …
Bill Gates attended Harvard from…
Google was founded by Larry Page …

Requires entity annotations in web documents
See http://lemurproject.org/clueweb12/FACC1/

Mintz et al. 2009

Distant supervision: training data

Knowledge Graph

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Web documents

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, …
Bill Gates attended Harvard from…
Google was founded by Larry Page …

Training data

(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y
Feature: X, founder of Y

Mintz et al. 2009

Distant supervision: training data

Knowledge Graph

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Web documents

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, …
Bill Gates attended Harvard from…
Google was founded by Larry Page …

Training data

(Bill Gates, Microsoft)

Label:	Founder
Feature:	X founded Y
Feature:	X, founder of Y

(Larry Page, Google)

Label:	Founder
Feature:	Y was founded by X

Mintz et al. 2009

Distant supervision: training data

Knowledge Graph

Founder: (Bill Gates, Microsoft)
Founder: (Larry Page, Google)
CollegeAttended: (Bill Gates, Harvard)

Web documents

Bill Gates founded Microsoft in 1975.
Bill Gates, founder of Microsoft, …
Bill Gates attended Harvard from…
Google was founded by Larry Page …

Training data

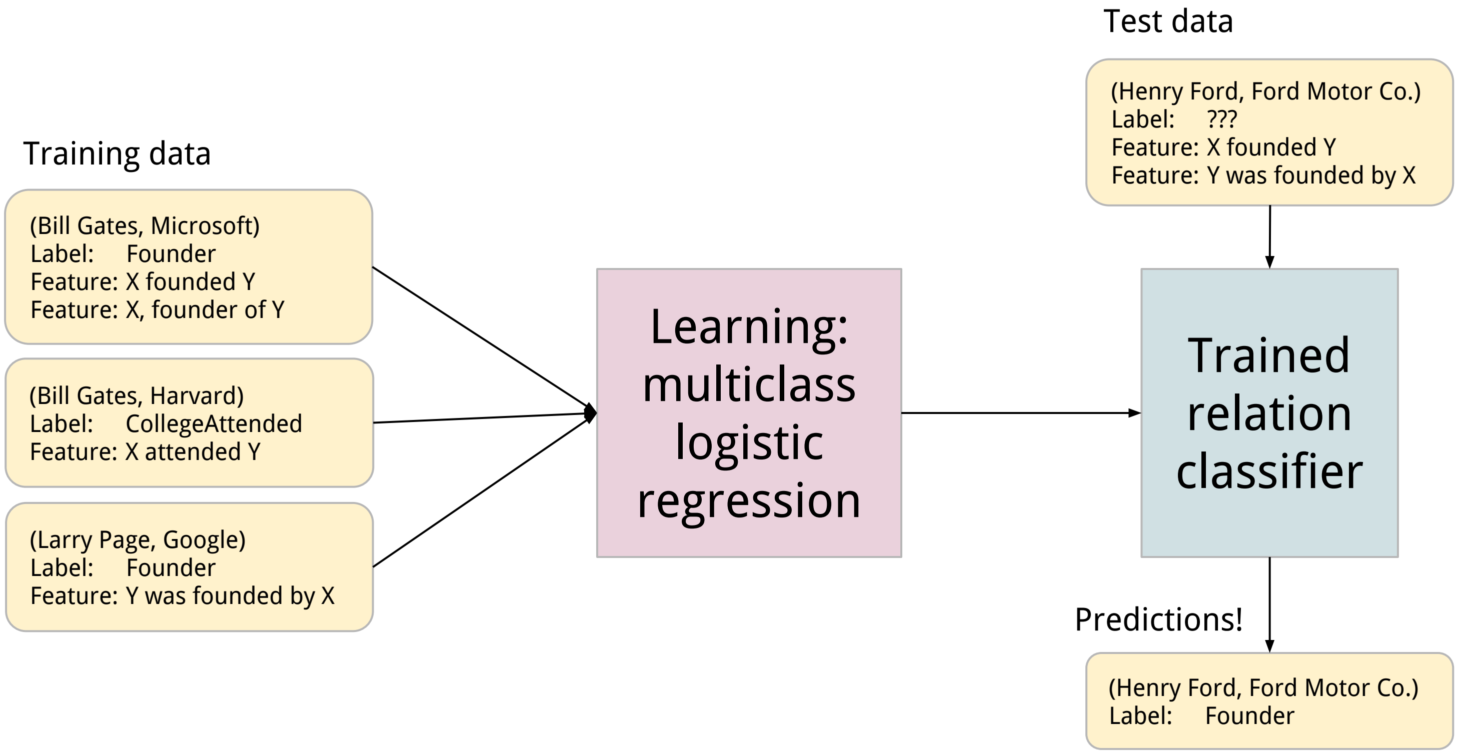
(Bill Gates, Microsoft)
Label: Founder
Feature: X founded Y
Feature: X, founder of Y

(Larry Page, Google)
Label: Founder
Feature: Y was founded by X

(Bill Gates, Harvard)
Label: CollegeAttended
Feature: X attended Y

Mintz et al. 2009

Distant supervision: learning a model



Mintz et al. 2009

Distant supervision: new relation instances

Relation name	New instance
/location/location/contains	Paris, Montmartre
/location/location/contains	Ontario, Fort Erie
/music/artist/origin	Mighty Wagon, Cincinnati
/people/deceased_person/place_of_death	Fyodor Kamensky, Clearwater
/people/person/nationality	Marianne Yvonne Heemskerk, Netherlands
/people/person/place_of_birth	Wavell Wayne Hinds, Kingston
/book/author/works_written	Upton Sinclair, Lanny Budd
/business/company/founders	WWE, Vince McMahon
/people/person/profession	Thomas Mellon, judge

Mintz et al. 2009

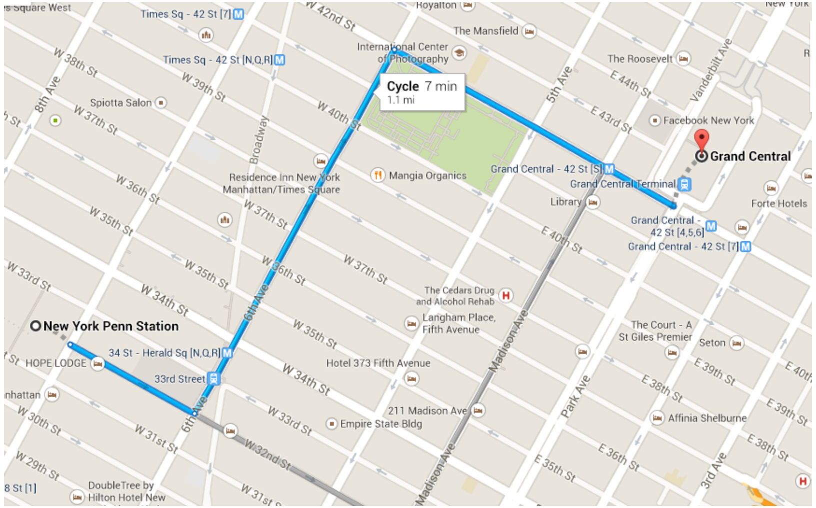
Outline

- Introduction
- NLU yesterday and today
- Relation extraction
- Semantic parsing
- Vector-space semantics
- Sentiment analysis

The semantic parsing task

“navigate me to Grand Central by bike”

```
(GetDirections
   (Destination /m/01rz3c)
   (Mode BIKE))
```



Natural-language interfaces to databases

To facilitate data exploration and analysis, you might want to parse natural language into database queries:

which country had the highest carbon emissions last year

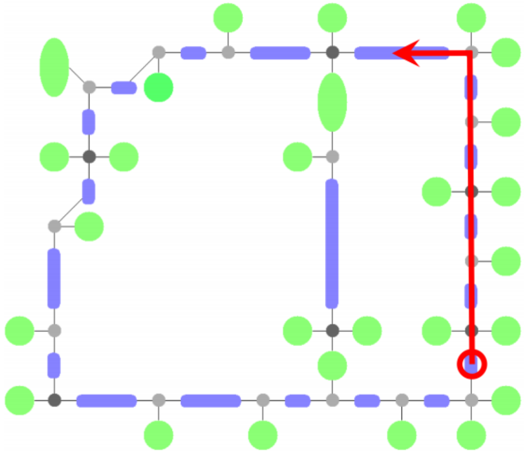
```
SELECT    country.name
FROM      country, co2_emissions
WHERE     country.id = co2_emissions.country_id
AND       co2_emissions.year = 2013
ORDER BY  co2_emissions.volume  DESC
LIMIT     1;
```

Robot control

For a robot control application, you might want a custom-designed procedural language:

Go to the third junction and take a left.

```
(do-sequentially
  (do-n-times 3
    (do-sequentially
      (move-to forward-loc)
      (do-until
        (junction current-loc)
        (move-to forward-loc))))
  (turn-left))
```



Matuszek et al. 2012

Semantic query parsing at Google

A growing proportion of queries require semantic interpretation.
Conventional keyword-based retrieval does not suffice!

directions to SF by train

```
(TravelQuery
  (Destination /m/0d6lp)
  (Mode TRANSIT))
```

angelina jolie net worth

```
(FactoidQuery
  (Entity /m/0f4vbz)
  (Attribute /person/net_worth))
```

weather friday austin tx

```
(WeatherQuery
  (Location /m/0vzm)
  (Date 2013-12-13))
```

text my wife on my way

```
(SendMessage
  (Recipient 0x31cbf492)
  (MessageType SMS)
  (Subject "on my way"))
```

play sunny by boney m

```
(PlayMedia
  (MediaType MUSIC)
  (SongTitle "sunny")
  (MusicArtist /m/017mh))
```

is REI open on sunday

```
(LocalQuery
  (QueryType OPENING_HOURS)
  (Location /m/02nx4d)
  (Date 2013-12-15))
```

Challenge: linguistic variation

I need to get to the train station

bike gct

grand central navigation

train station directions

Which train should I take to get to Grand Central Terminal?

How do I get to Grand Central?

how to get to grand central by subway

tell me how to go to the grand central

take me to grand central

best route gct

grand central navigation

directions to grand central

what's the best way to walk to grand central from here

walk to grand central

Challenge: internationalization

взять меня в Grand Central

I need to get to the train station

bike gct

grand central navigation

私は、鉄道駅に取得する必要があります

bici GCT

train station directions

Which train should I take to get to Grand Central Terminal?

directions à Grand Central

How do I get to Grand Central?

how to get to grand central by subway

Bahnhof Richtungen

tell me how to go to the grand central

take me to grand central

navegar grand central

التي القطار لجراند سنترال

best route gct

grand central navigation

directions to grand central

步行到中央火车站

what's the best way to walk to grand central from here

walk to grand central

Challenge: ambiguity

“italian reservation in palo alto”

`$Cuisine reservation in $Location`

“indian reservation in montana”

“mission bike directions”

`$Location $TransportationMode directions`

“mission bicycle directions”

Typical approaches

- Recent academic work explores a variety of related approaches
  - Cf. Zettlemoyer & Collins 2005, Kwiatkowski et al. 2013
  - Cf. Liang et al. 2011, Liang et al. 2013, Berant et al. 2013
- Context-free grammars with semantic attachments
- Log-linear scoring models learned from training data
- Leveraging annotators for numbers, locations, times, entities, ...
- Grammar induction to learn CFG rules from data

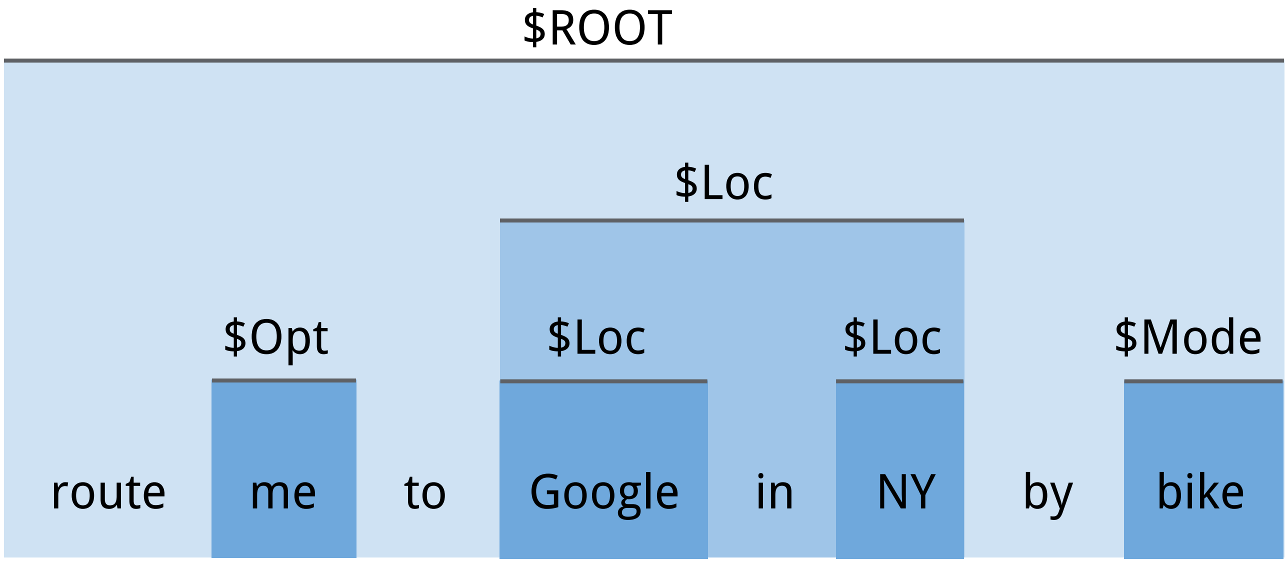
Context-free grammar

The *syntactic* part of the grammar is a fairly conventional CFG:

```
$Loc → Google
$Loc → NY
$Loc → $Loc in $Loc
$Opt → me
$Mode → bike
$Mode → car
$ROOT → route ($Opt)? to $Loc by $Mode
```

Usually *not* deterministic: many possible derivations per input.

Example parse

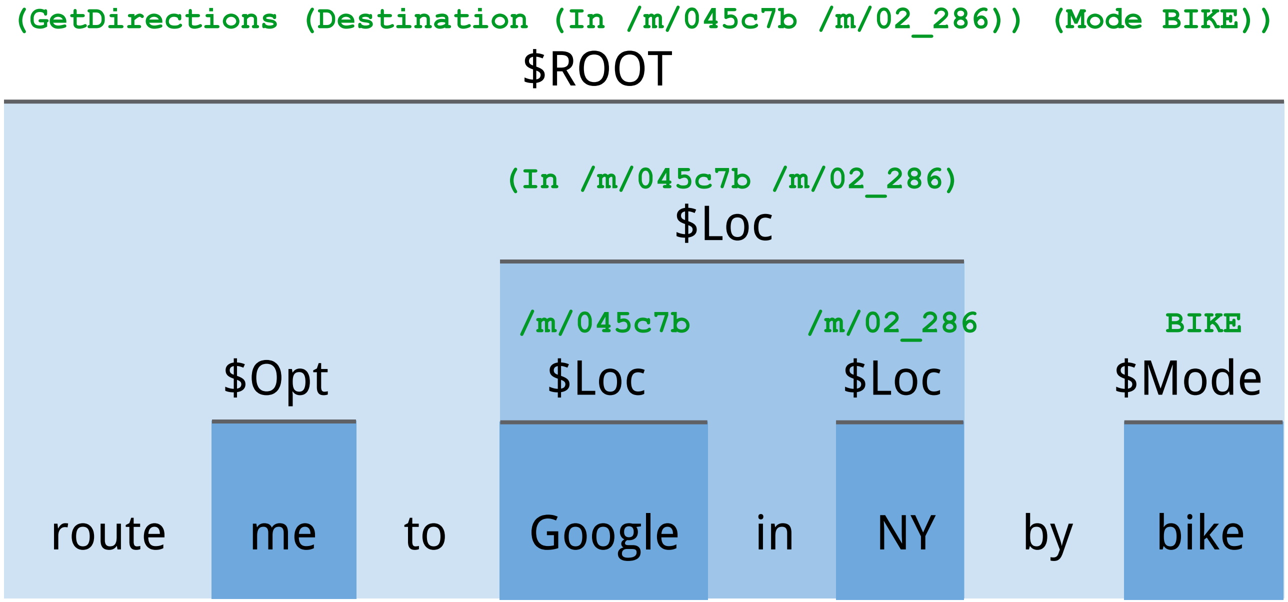


Semantic attachments to grammar rules

Little programs which compute semantic interpretation bottom-up

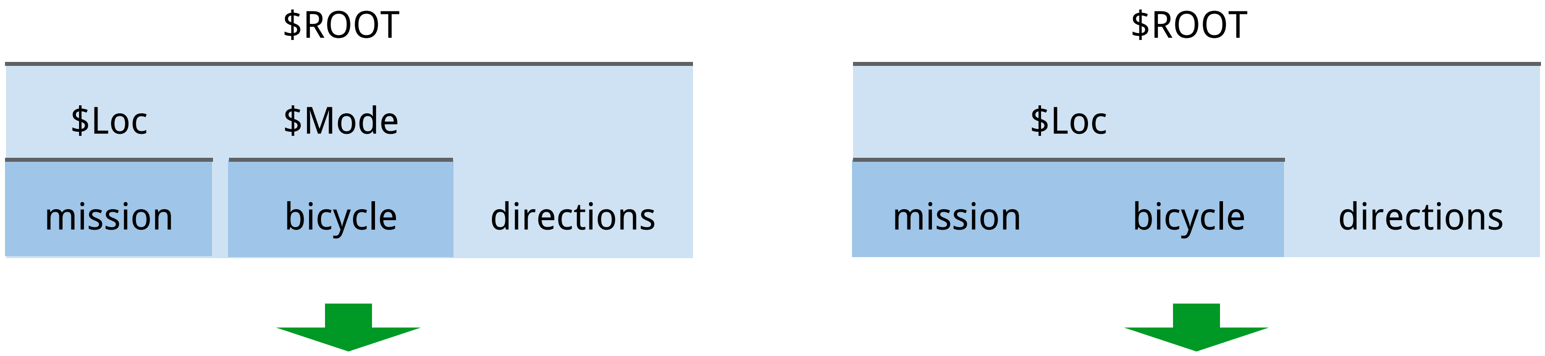
```
$Loc → Google [/m/045c7b]
$Loc → NY [/m/02_286]
$Loc → $Loc in $Loc [(In $1 $2)]
$Opt → me []
$Mode → bike [BIKE]
$Mode → car [CAR]
$ROOT → route ($Opt)? to $Loc by $Mode
  [(GetDirections (Destination $2) (Mode $3))]
```

Example parse, now with semantics!



Semantic ambiguity

When grammar supports multiple interpretations, how to choose?




```
(GetDirections
  (Destination /m/02ll7r)
  (Mode BIKE))
```

```
(GetDirections
  (Destination /g/1tfmcgjy))
```

Scoring model

- A log-linear model to score alternative derivations (parses)
- Features from input *x*, semantic yield *y*, and derivation *z*
  - E.g., co-occurrence of *"to"* in input and `Destination` in semantics
  - E.g., occurrence of specific CFG rules or categories in derivation
  - E.g., confidence score from an annotator

$$score(x, z) = features(x, z)^\top \theta$$

$$p(z|x, \theta) = \frac{e^{score(x,z)}}{\sum_{z' \in Z(x)} e^{score(x,z')}}$$

Learning

- Estimate parameters using EM-style training (Liang et al., 2011)
- Assume we have training data
- Sum out latent derivations in n-best list

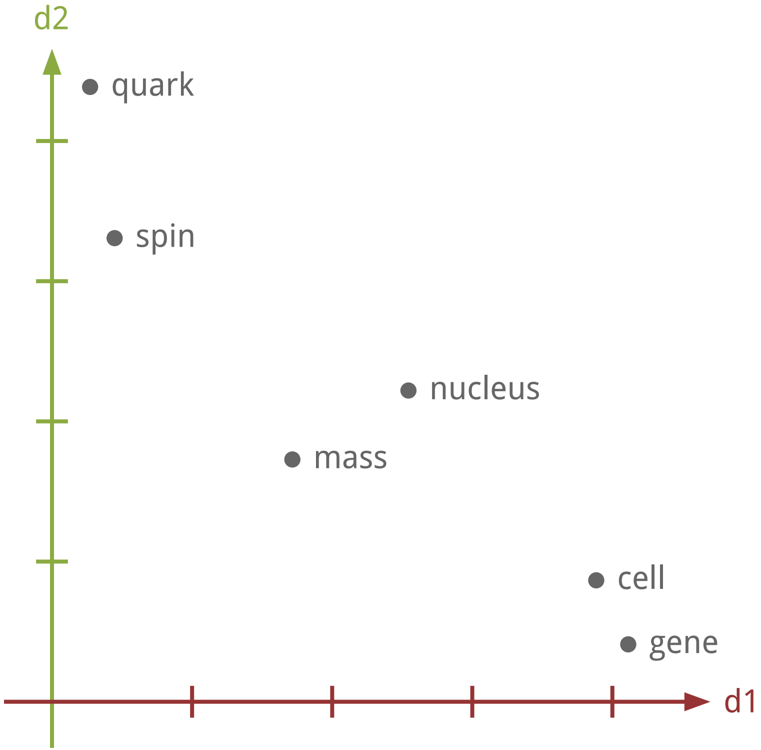
$$\underset{\theta}{\operatorname{argmax}} \sum_{z} p(y|z)\, p(z|x,\theta)$$

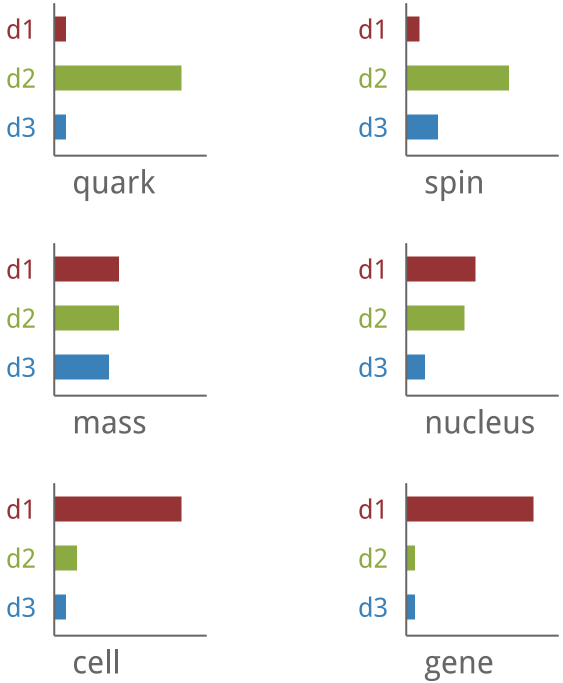
- Update using Stochastic Gradient Descent (SGD)
- Adaptive step size à la AdaGrad (Duchi et al., 2008)

Outline

- Introduction
- NLU yesterday and today
- Relation extraction
- Semantic parsing
- Vector-space semantics
- Sentiment analysis

Vector-space models of meaning





Kinds of vector space models

"Distributional" models

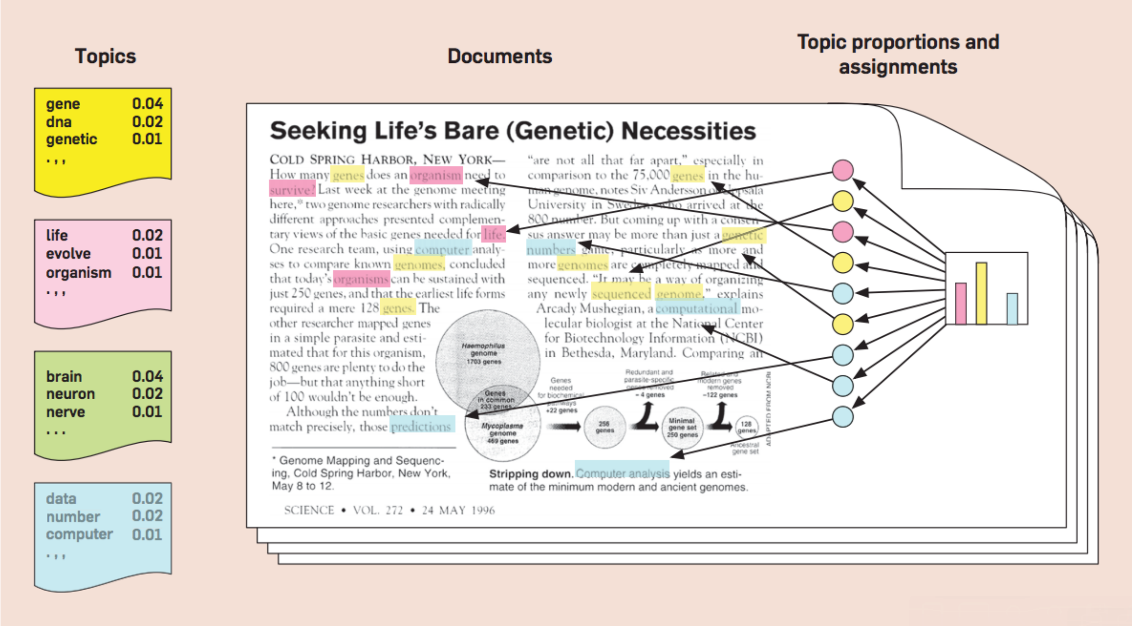
- Vectors based on the distribution of contexts in which word appears
- Examples: tf-idf (term frequency / inverse document frequency), LSA (latent semantic analysis), LDA (latent Dirichlet allocation)

"Distributed" models

- Vectors are the output of some neural network model
- Examples: NNLM (neural network language model), CBOW (continuous bag-of-words), skip-gram, RNN (recursive neural network), MV-RNN (matrix-vector RNN), RNTN (recursive neural tensor network)

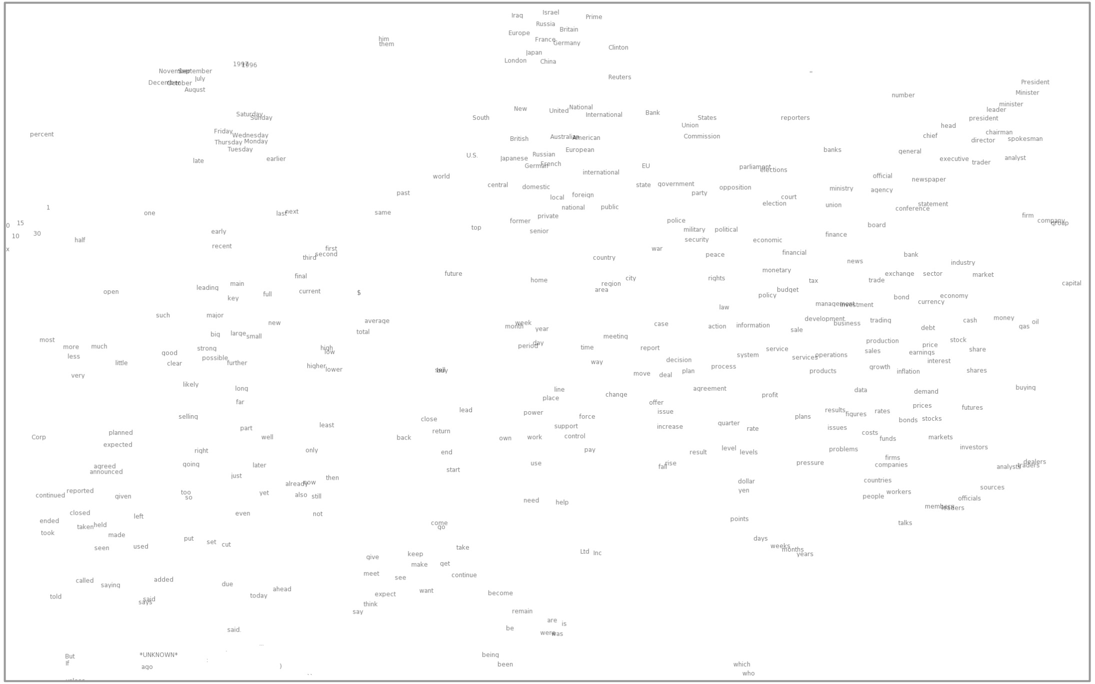
Latent Dirichlet allocation (LDA)

Figure 1. The intuitions behind latent Dirichlet allocation. We assume that some number of "topics," which are distributions over words, exist for the whole collection (far left). Each document is assumed to be generated as follows. First choose a distribution over the topics (the histogram at right); then, for each word, choose a topic assignment (the colored coins) and choose the word from the corresponding topic. The topics and topic assignments in this figure are illustrative—they are not fit from real data. See Figure 2 for topics fit from data.



graphic from Blei 2012

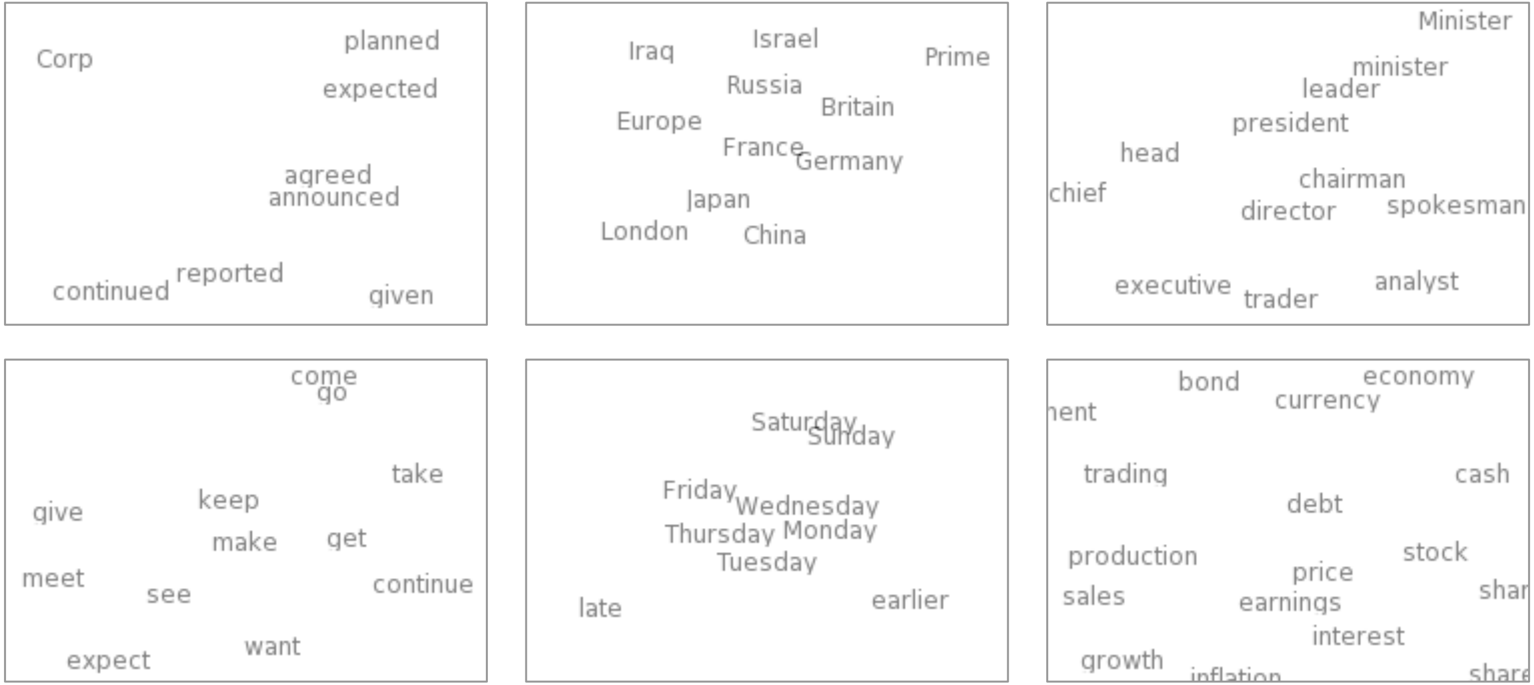
Neural word embeddings



100D word embeddings projected with t-SNE [Turian et al. 2010]

http://metaoptimize.com/projects/wordreprs/

Neural word embeddings



100D word embeddings projected with t-SNE [Turian et al. 2010]

http://metaoptimize.com/projects/wordreprs/

The word2vec model

Work by Mikolov & others at Google (2013)

"Skip-gram" model: a simple neural embedding

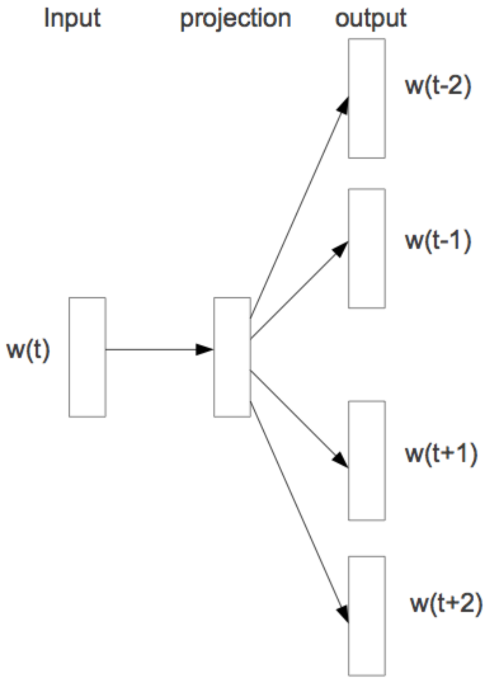
Learn word vectors that predict nearby words:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_{t+j}|w_t)$$

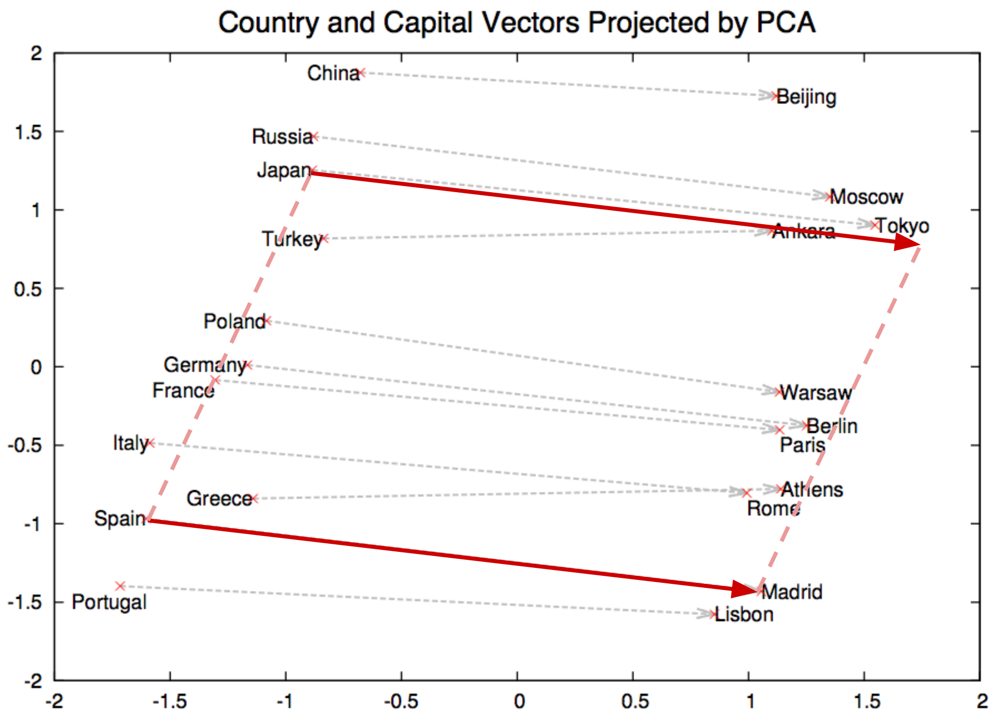
Highly efficient & scalable

Train 1000D vectors on 33B words in 1 day!

Get it at https://code.google.com/p/word2vec/



Relationships learned by word2vec



Mikolov et al. 2013

Madrid
– Spain
+ Japan

= Tokyo

Solving analogies using word2vec

Newspapers			
New York	New York Times	Baltimore	Baltimore Sun
San Jose	San Jose Mercury News	Cincinnati	Cincinnati Enquirer
NHL Teams			
Boston	Boston Bruins	Montreal	Montreal Canadiens
Phoenix	Phoenix Coyotes	Nashville	Nashville Predators
NBA Teams			
Detroit	Detroit Pistons	Toronto	Toronto Raptors
Oakland	Golden State Warriors	Memphis	Memphis Grizzlies
Airlines			
Austria	Austrian Airlines	Spain	Spainair
Belgium	Brussels Airlines	Greece	Aegean Airlines
Company executives			
Steve Ballmer	Microsoft	Larry Page	Google
Samuel J. Palmisano	IBM	Werner Vogels	Amazon

Mikolov et al. 2013

Vector compositionality in word2vec

The four closest tokens to the sum of two vectors:

Czech + currency	Vietnam + capital	German + airlines	Russian + river	French + actress
koruna Check crown Polish zolty CTK	Hanoi Ho Chi Minh City Viet Nam Vietnamese	airline Lufthansa carrier Lufthansa flag carrier Lufthansa Lufthansa	Moscow Volga River upriver Russia	Juliette Binoche Vanessa Paradis Charlotte Gainsbourg Cecile De

Mikolov et al. 2013

But why!?! My interpretation:

Vectors implicitly represent (log of) probability distributions over contexts.
Sum of vectors represents product (conjunction) of context distributions.
Which words are probable both near “Czech” and near “currency”?

Outline

- Introduction
- NLU yesterday and today
- Relation extraction
- Semantic parsing
- Vector-space semantics
- Sentiment analysis

Sentiment analysis

Traditional approach: count +/– sentiment words

best movie of the year … a triumph 

slick and entertaining, despite a weak script 

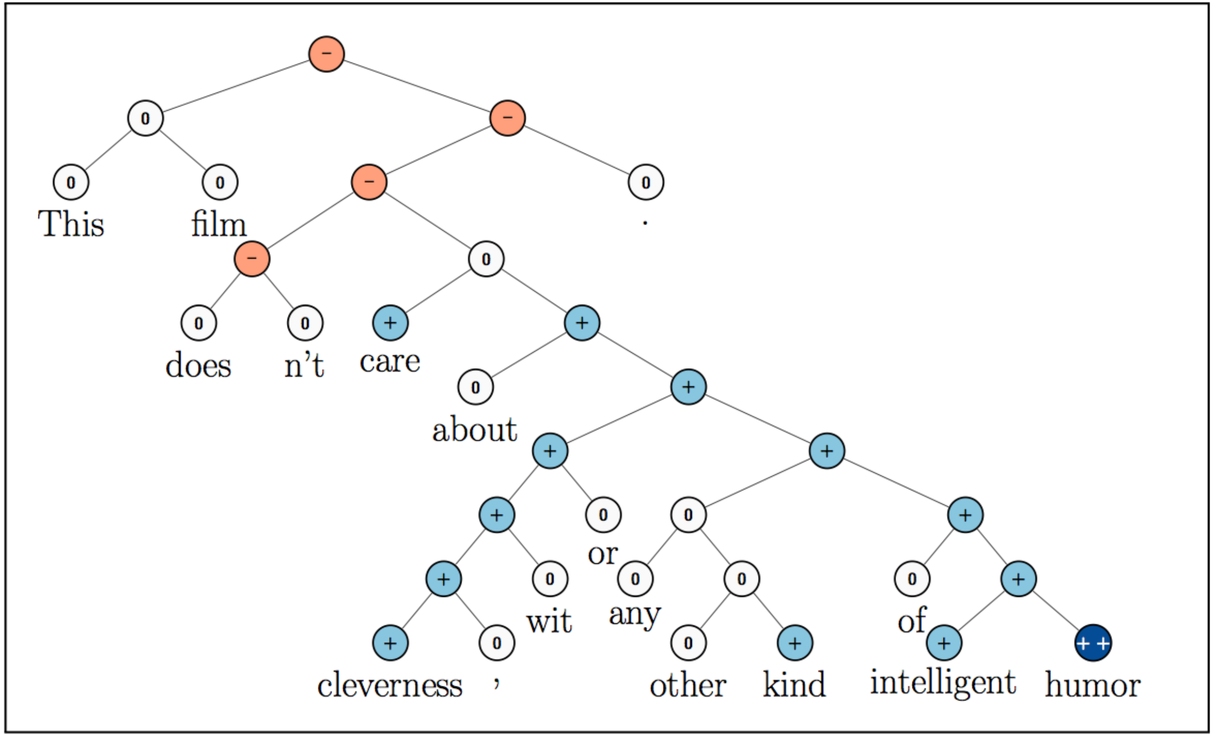
an abysmal failure 

But it’s hard to account for role of semantic composition

not an abysmal failure 

fun, sweet, and earnest, but ultimately unsatisfying 

Sentiment and compositionality



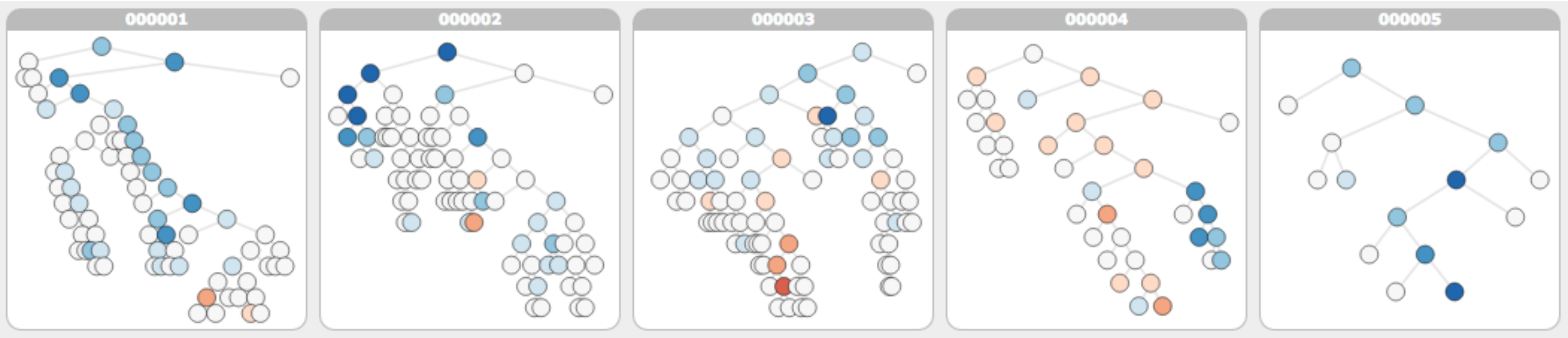
Socher et al. 2013

The Stanford Sentiment Treebank

10K sentences from movie reviews, with 215K phrases

5-level sentiment labels collected from Mechanical Turk

See http://nlp.stanford.edu/sentiment/treebank.html



Socher et al. 2013

Recursive neural tensor networks (RNTNs)

Semantics of words are 30D vectors

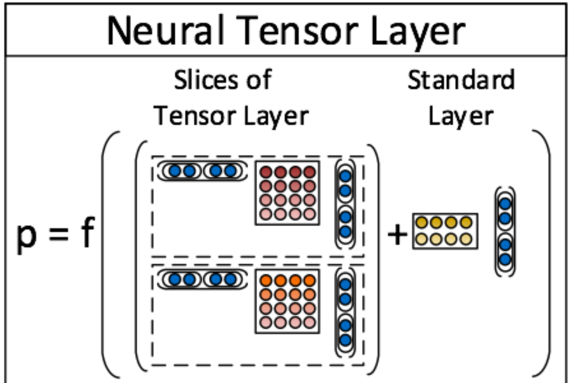
Semantics of phrases via tensor product

Trained via backprop in neural network

Goal: minimize KL divergence to SST labels

85% accuracy on +/– sentence sentiment

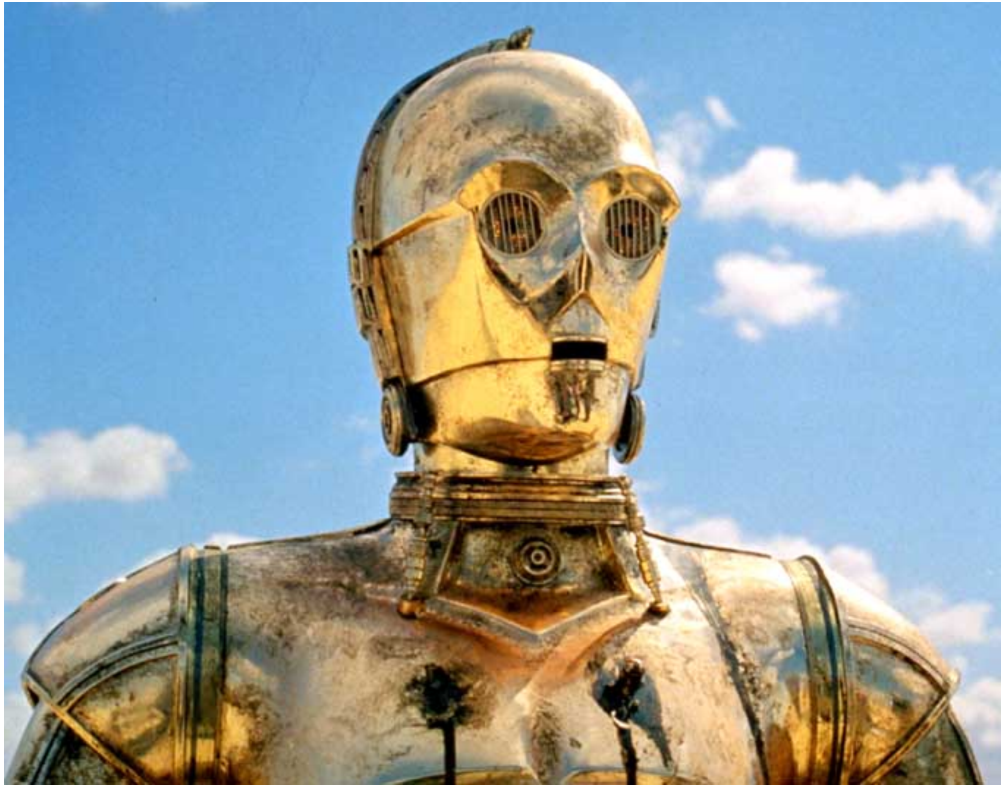
Try it! http://nlp.stanford.edu/sentiment/



$$p = f\left(\begin{bmatrix} b \\ c \end{bmatrix}^T V^{[1:2]} \begin{bmatrix} b \\ c \end{bmatrix} + W \begin{bmatrix} b \\ c \end{bmatrix} \right)$$

Socher et al. 2013

Are we there yet?



THE END

Questions?

Outline

- Introduction
- NLU yesterday and today
- Relation extraction
- Semantic parsing
- Vector-space semantics
- Sentiment analysis
- Vector-space semantic parsing [bonus topic!]

Vector space semantic parsing

OK, VSMs seem cool, but what about semantic composition?

Idea: combine VSMs with combinatory categorial grammar (CCG)

Nouns → vectors; adjs & dets → matrices; tr. verbs & preps → tensors

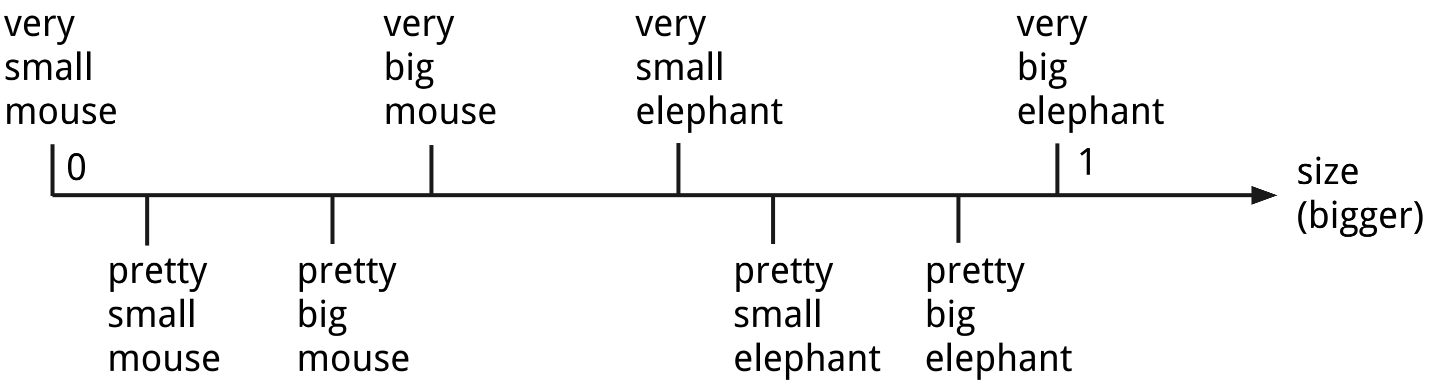
Semantic composition via matrix / tensor multiplication

$$
\dfrac{
\dfrac{\text{the}}{\begin{array}{c} NP/N \\ \lambda x.x \end{array}}
\quad
\dfrac{
\dfrac{\text{red}}{\begin{array}{c} N/N \\ \lambda x.A_{red}x \end{array}}
\quad
\dfrac{\text{ball}}{\begin{array}{c} N \\ v_{ball} \end{array}}
}{N : A_{red}v_{ball}}
}{NP : A_{red}v_{ball}}
\quad
\dfrac{
\dfrac{\text{on}}{\begin{array}{c} (NP\backslash NP)/NP \\ \lambda x.\lambda y.A_{on}x + B_{on}y \end{array}}
\quad
\dfrac{
\dfrac{\text{the}}{\begin{array}{c} NP/N \\ \lambda x.x \end{array}}
\quad
\dfrac{\text{table}}{\begin{array}{c} N \\ v_{table} \end{array}}
}{NP : v_{table}}
}{NP\backslash NP : \lambda y.A_{on}v_{table} + B_{on}y}
$$

$$
\overline{\qquad NP : A_{on}v_{table} + B_{on}A_{red}v_{ball} \qquad}
$$

Krishnamurthy & Mitchell 2013

Adjective-adverb-noun composition



"elephant"	$\begin{pmatrix} 1.6 \\ -0.1 \end{pmatrix}$	"mouse"	$\begin{pmatrix} -0.1 \\ 1.6 \end{pmatrix}$
"small"	$\begin{pmatrix} 0.22 & 0 \\ 0 & 1.7 \end{pmatrix}$	"big"	$\begin{pmatrix} 1.7 & -1.1 \\ 0 & 0.22 \end{pmatrix}$
"very small"	$\begin{pmatrix} 0.25 & -.12 \\ -1.34 & 2.3 \end{pmatrix}$	"very big"	$\begin{pmatrix} 2.3 & -1.34 \\ -0.12 & 0.25 \end{pmatrix}$

References

Berant, Jonathan, Andrew Chou, Roy Frostig, and Percy Liang. Semantic Parsing on Freebase from Question-Answer Pairs. EMNLP-2013.

Krishnamurthy, Jayant, and Tom M. Mitchell. Vector space semantic parsing: A framework for compositional vector space models. ACL-2013.

Kwiatkowski, Tom, Eunsol Choi, Yoav Artzi, and Luke Zettlemoyer. Scaling semantic parsers with on-the-fly ontology matching. EMNLP-2013.

Liang, Percy, Michael I. Jordan, and Dan Klein. Learning dependency-based compositional semantics. Computational Linguistics, 2013.

Matuszek, Cynthia, Even Herbst, Luke S. Zettlemoyer, and Dieter Fox. Learning to parse natural language commands to a robot control system. 13th International Symposium on Experimental Robotics, 2012.

References

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. NIPS-2013.

Mintz, Mike, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. ACL-2009.

Socher, Richard, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. EMNLP-2013.

Turian, Joseph, Lev Ratinov, and Yoshua Bengio. Word representations: a simple and general method for semi-supervised learning. ACL-2010.

Zettlemoyer, Luke S. and Collins, Michael. Learning to map sentences to logical form: structured classification with probabilistic categorial grammars. UAI-2005.

BACKUP SLIDES

The Geo880 dataset (880 examples)

what cities in texas have the highest number of citizens?
what is the area of the state with the smallest population density?
what state is des moines located in?
what are the major cities in states through which the mississippi runs?
what are the major cities in the smallest state in the us?
what is the capital of ohio?
what is the population of denver?
what is the biggest city in nebraska?
what are the major cities in new mexico?
what is the capital of california?
what is the capital of utah?
what are the population of mississippi?
where is mount whitney?
what is the population of the state with the largest area?
what is the capital of iowa?
what is the most populous state through which the mississippi runs?
how many states border on the state whose capital is boston?
which states does the longest river cross?
what is the capital of new york?
what is the smallest city in arkansas?
how many people live in mississippi?

what is largest capital?
how many states are in the usa?
how many big cities are in pennsylvania?
what state contains the highest point in the us?
where is san jose?
how many cities are in montana?
what states border michigan?
name the rivers in arkansas.
what rivers are in nevada?
could you tell me what is the highest point in the state of oregon?
what state borders new york?
which states border hawaii?
what is the population of atlanta ga?
which state is the smallest?
what is the largest city in missouri?
how much population does texas have?
give me the number of rivers in california?
how many states does iowa border?
what states border states that the ohio runs through?
which states border texas?
what is the population of dallas?

The WebQuestions dataset (5810 examples)

what is the name of justin bieber brother?
what character did natalie portman play in star wars?
what state does selena gomez?
what country is the grand bahama island in?
what kind of money to take to bahamas?
what character did john noble play in lord of the rings?
who does joakim noah play for?
where are the nfl redskins from?
where did saki live?
how old is sacha baron cohen?
what two countries invaded poland in the beginning of ww2?
what time zone am i in cleveland ohio?
who did draco malloy end up marrying?
which countries border the us?
where is rome italy located on a map?
what is nina dobrev nationality?
what country does iceland belong to?
which kennedy died first?
what books did beverly cleary right?
who did the philippines gain independence from?
where to fly into bali?

what movies does taylor lautner play in?
what year lebron james came to the nba?
what did the german revolution lead to?
how much did adriana lima gain during pregnancy?
what does thai mean?
which wife did king henry behead?
who was ishmael's mom?
what was malcolm x trying to accomplish?
where are the netherlands on a world map?
what is the president of brazil?
what are the major cities in france?
what city did esther live in?
what sport do the toronto maple leafs play?
what is saint nicholas known for?
when is the new series of the only way is essex starting?
what is cher's son's name?
what is martin cooper doing now?
what party was andrew jackson?
what is medicare a?
what county is the city of hampton va in?
what is the name of the first harry potter novel?