Lecture 1: Introduction

Kai-Wei Chang
CS @ UCLA
kw@kwchang.net

Couse webpage: https://uclanlp.github.io/CS269-17/

Announcements

- Waiting list: Start attending the first few lectures as if you are registered. Given that some students will drop the class, some space will free up.

- We will use Piazza as an online discussion platform. Please sign up here: piazza.com/ucla/fall2017/cs269

Staff

- Instructor: Kai-Wei Chang
  - Email: ml17@kwchang.net
  - Office: BH 3732J
  - Office hour: 4:00 – 5:00, Tue (after class).
- TA: Md Rizwan Parvez
  - Email: wua4nw@virginia.edu
  - Office: BH 3809
  - Office hour: 12:00 – 2:00, Wed

This lecture

- Course Overview
  - What is NLP? Why it is important?
  - What types of ML methods used in NLP?
  - What will you learn from this course?
- Course Information
- What are the challenges?
- Key NLP components
- Key ML ideas in NLP

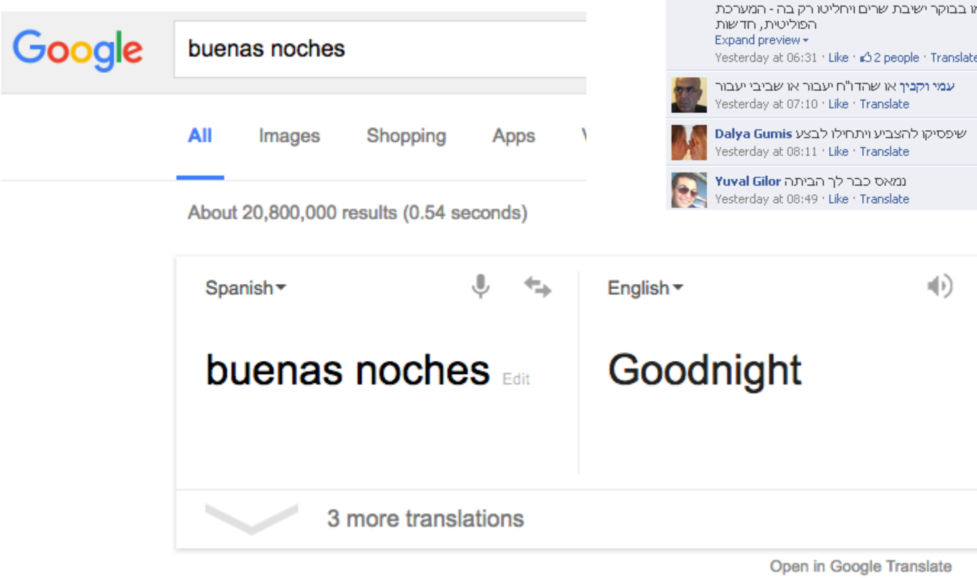
What is NLP

- **Wiki: Natural language processing** (**NLP**) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (**natural**) languages.

Go beyond the keyword matching

- Identify the structure and meaning of words, sentences, texts and conversations
- Deep understanding of broad language
- NLP is all around us

Machine translation



Facebook translation, image credit: Meedan.org

Statistical machine translation

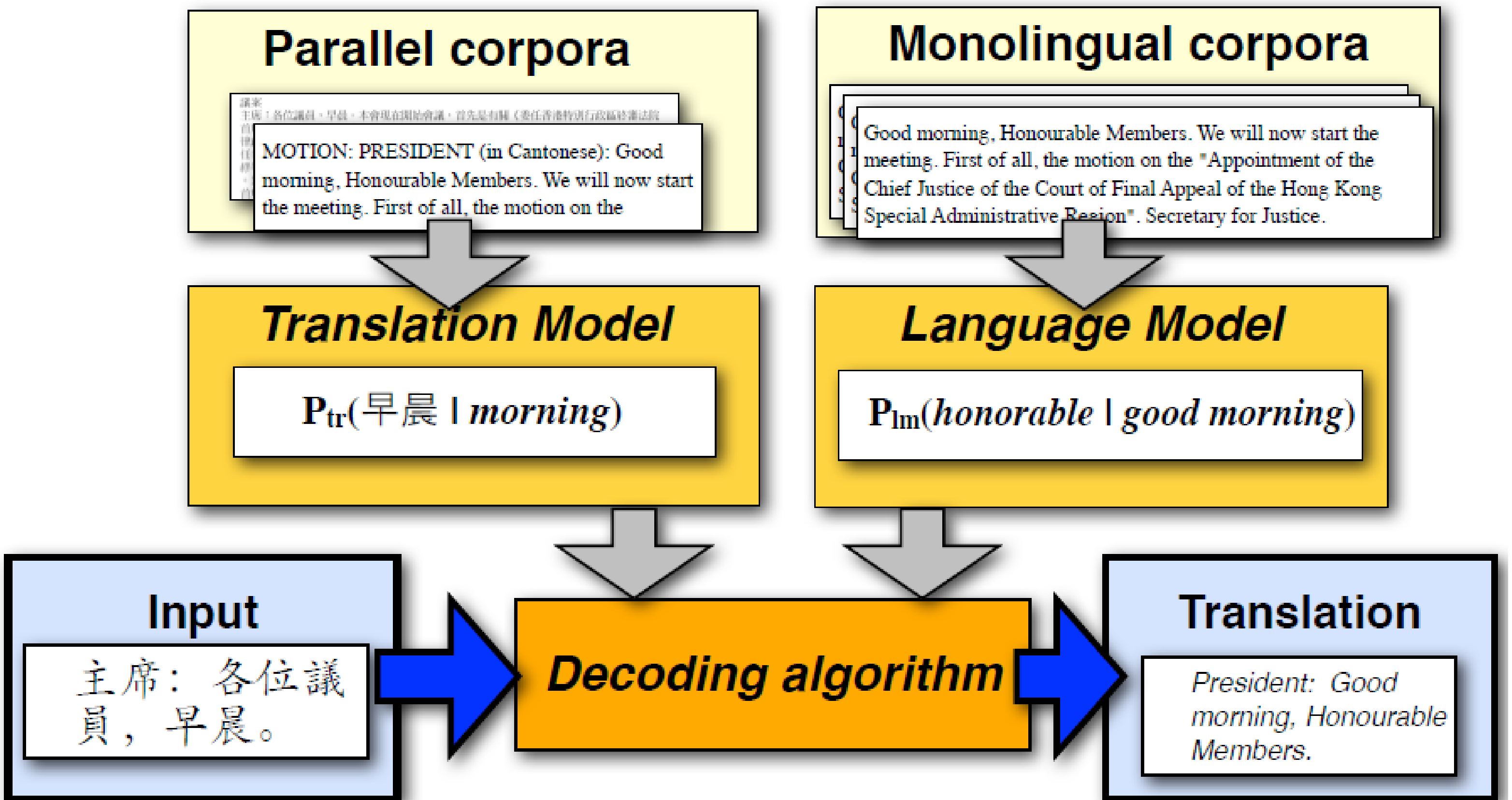
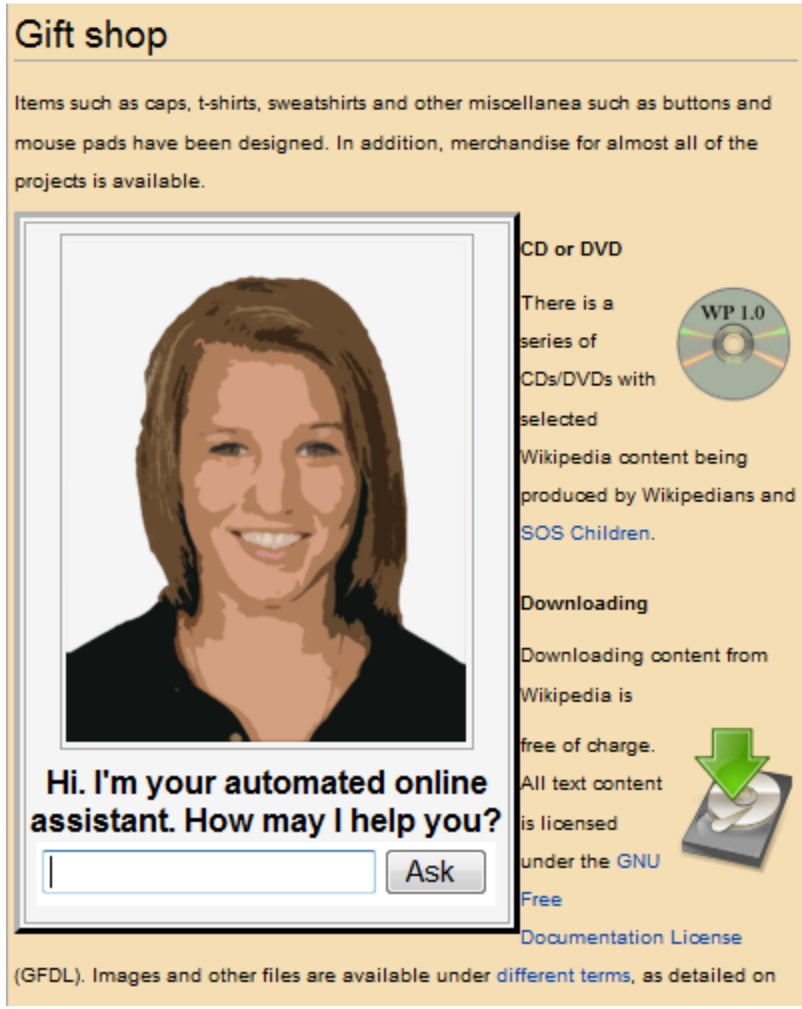
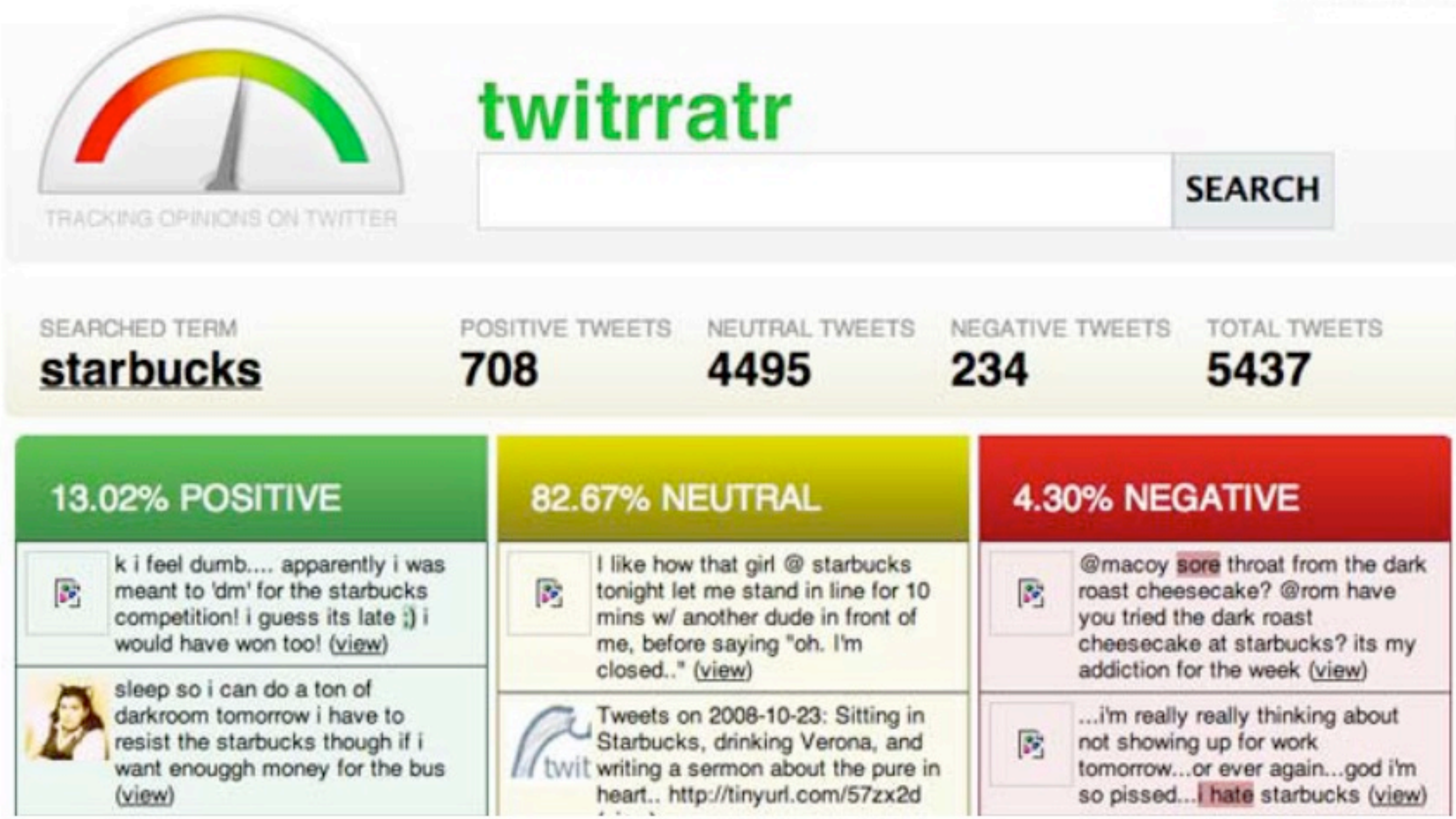


Image credit: Julia Hockenmaier, Intro to NLP

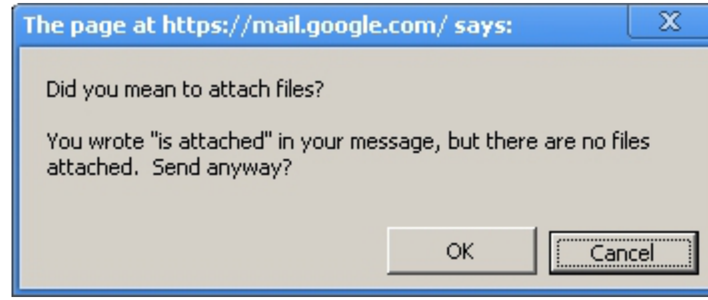
Dialog Systems

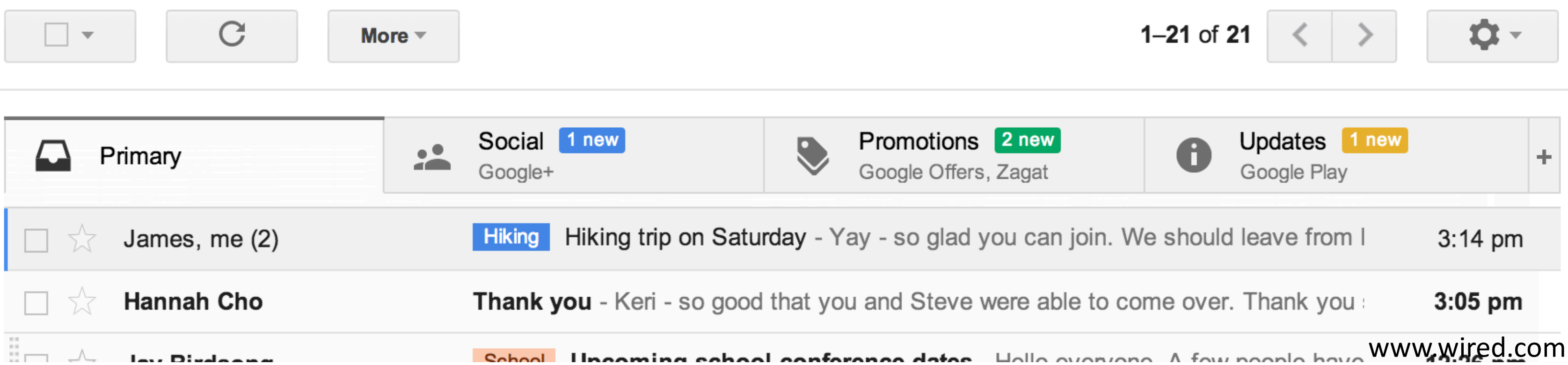


Sentiment/Opinion Analysis



Text Classification



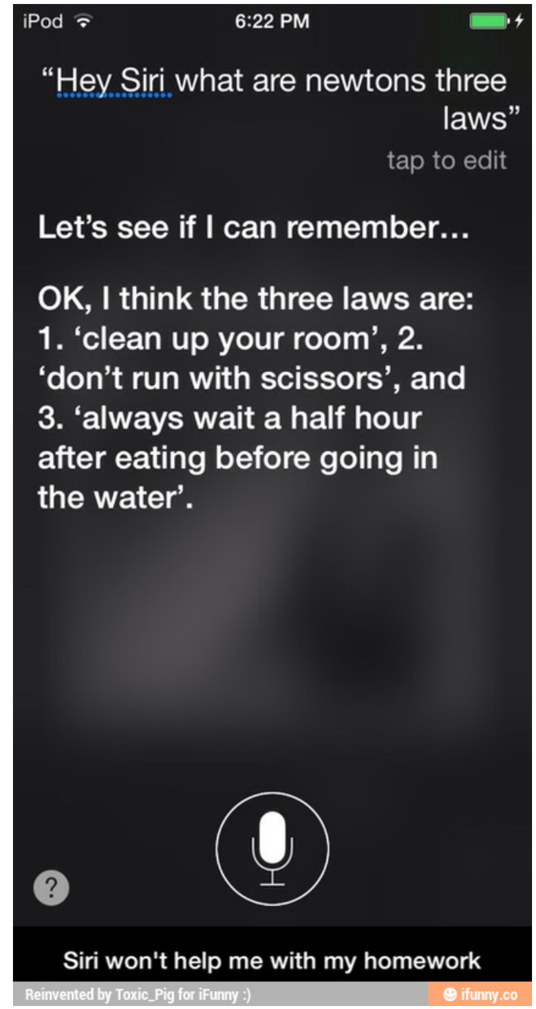


www.wired.com

- Other applications?

Question answering

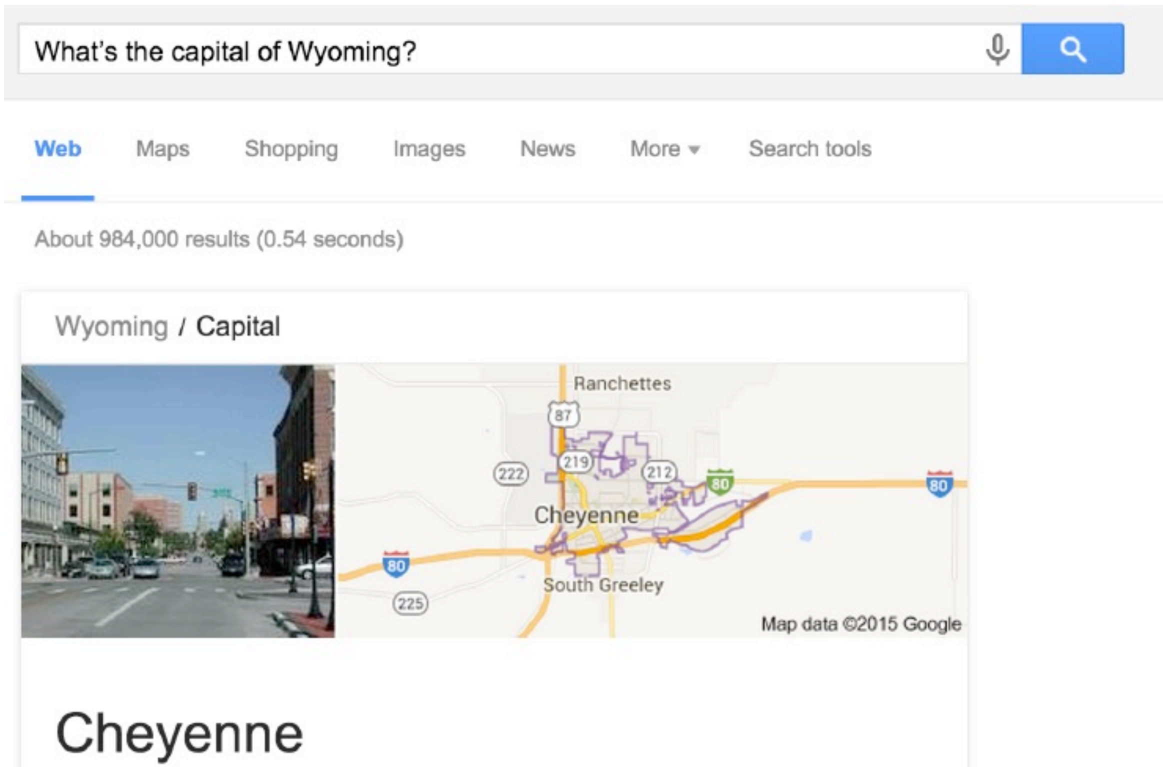
['Watson' computer wins at 'Jeopardy']



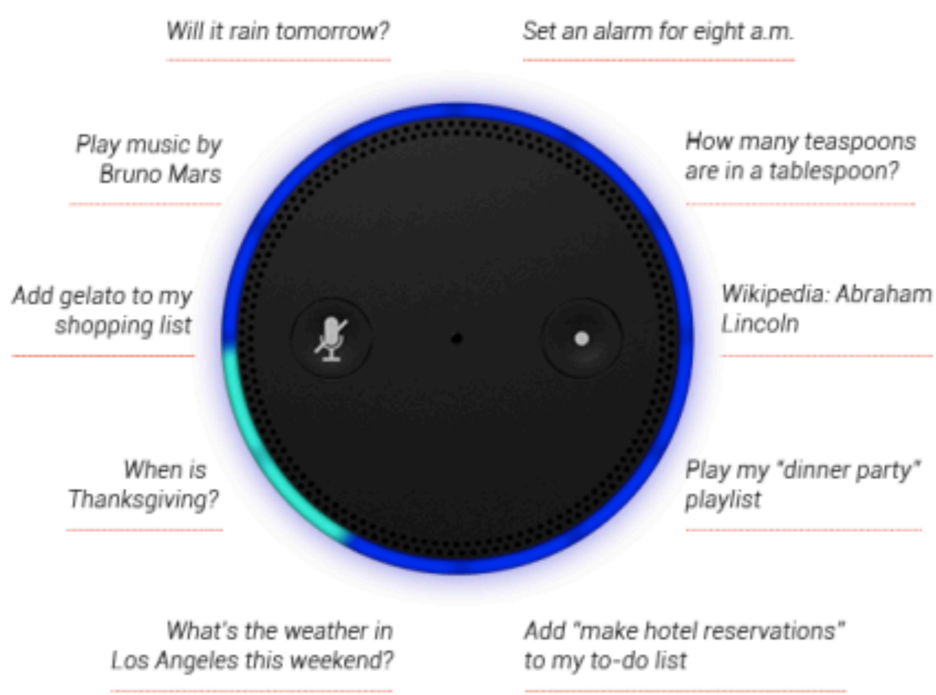
credit: ifunny.com

Question answering

❖ Go beyond search



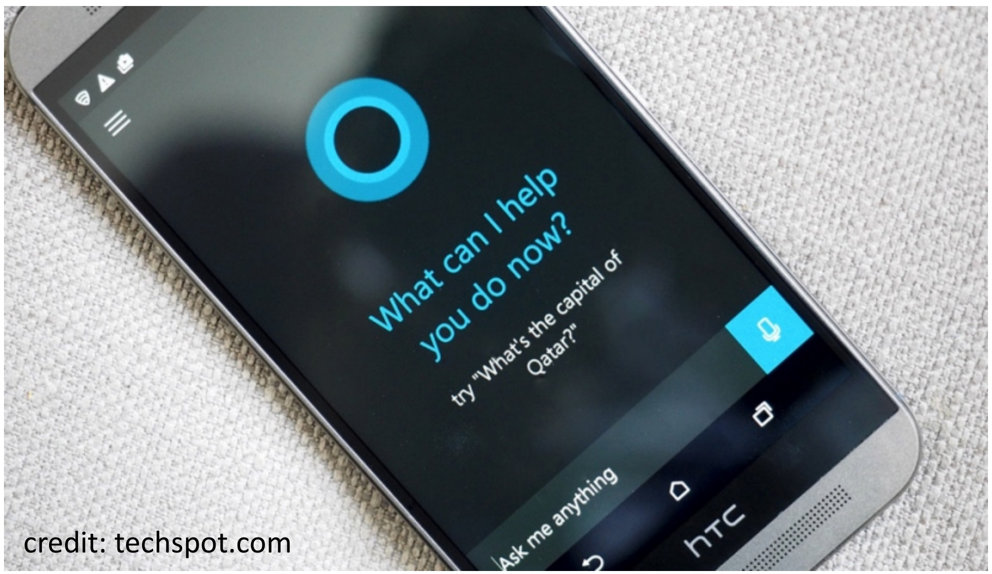
Natural language instruction



https://youtu.be/KkOCeAtKHIc?t=1m28s

Digital personal assistant

More on natural language instruction



credit: techspot.com

- Semantic parsing – understand tasks
- Entity linking – “my wife” = “Kellie” in the phone book

Information Extraction

- Unstructured text to database entries

New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

Yoav Artzi: Natural language processing

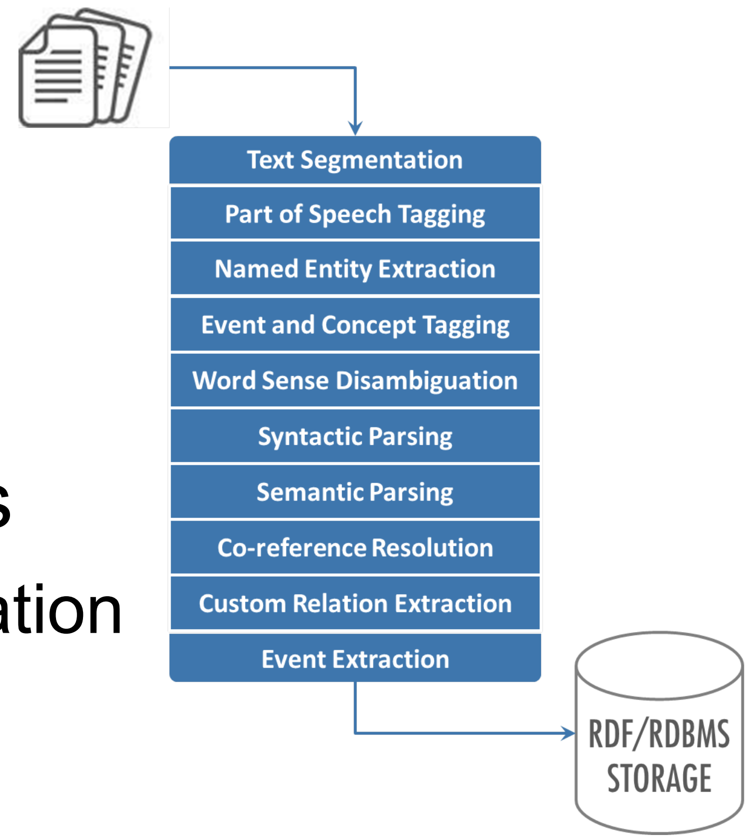
Language Comprehension

Christopher Robin is alive and well. **He** is the same person that you read about in the book, **Winnie the Pooh.** As **a boy**, **Chris** lived in a pretty home called **Cotchfield Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book

- Q: who wrote Winnie the Pooh?
- Q: where is Chris lived?

What will you learn from this course

- The NLP Pipeline
  - Key components for understanding text



- NLP systems/applications
  - Current techniques & limitation
- Build realistic NLP tools

What's not covered by this course

- Speech recognition – no signal processing
- Natural language generation
- Details of ML algorithms / theory
- Text mining / information retrieval

This lecture

- Course Overview
  - What is NLP? Why it is important?
  - What will you learn from this course?
- Course Information
- What are the challenges?
- Key NLP components

Overview

- New course, first time being offered
  - Comments are welcomed
  - target at first- or second- year PhD students
- Lecture + Seminar
- No course prerequisites, but I assume
  - programming experience (for the final project)
  - basic ML/AI background
  - basics of probability calculus, and linear algebra (HW0)

Grading

- Attendance & participations (10%)
  - Participate in discussion
- Paper summarization report (20%)
- Paper presentation (30%)
- Final project (40%)
  - Proposal (5%)
  - Final Paper (25%)
  - Presentation (10%)

Paper summarization

- 1 page maximum
- Pick one paper from recent ACL/NAACL/EMNLP/EACL
- Summarize the paper (use you own words)
  - Write a blog post using markdown or jupyter notebook:
  - https://einstein.ai/research/learned-in-translation-contextualized-word-vectors
  - https://github.com/uclanlp/reducingbias/blob/master/src/fairCRF_gender_ratio.ipynb

initialized methods for contextualizing word vectors through training on an intermediate task.

Encoders

A common approach to contextualizing word vectors is to use a recurrent neural network (RNN). RNNs are deep learning models that process vector sequences of variable length. This makes them suitable for processing sequences of word vectors. We use a specific kind of RNN called Long Short-Term Memory (LSTM) to better handle long sequences. At each step in processing, the LSTM takes in a word vector and outputs a new vector called the hidden vector. This process is often referred to as encoding the sequence, and the neural network that does the encoding is referred to as an encoder.

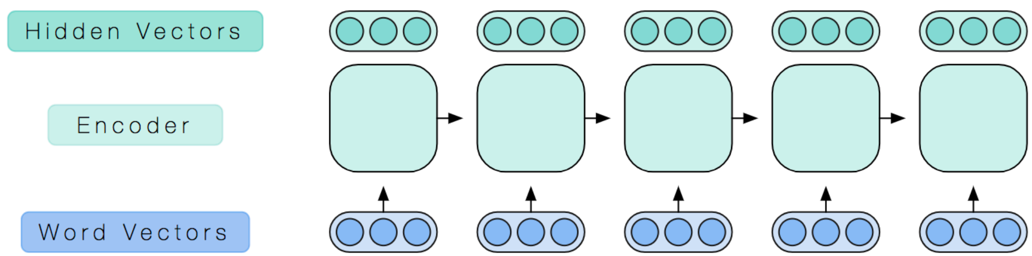


Figure 6: An LSTM encoder takes in a sequence of word vectors and outputs a sequence of hidden vectors.

Bidirectional Encoders

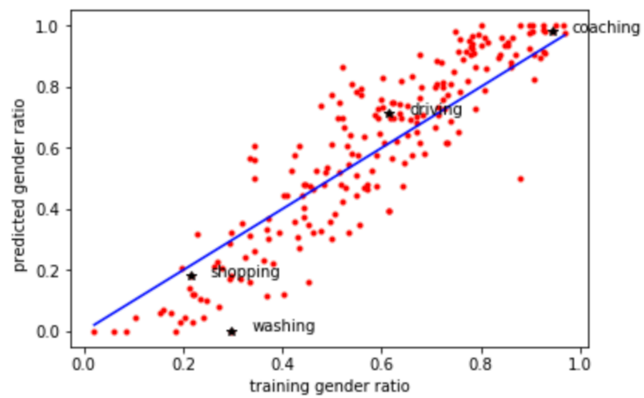
These hidden vectors do not incorporate information from words that appear later in the sequence, but this is easily remedied. We can run an LSTM backwards to get some backwards output vectors, and we can concatenate these with the output vectors from the forward LSTM to get a more useful hidden vector. We treat this pair of forward and backward LSTMs as a unit, and it is typically referred to as a bidirectional LSTM. It takes in a sequence of word vectors, runs a forward and a backward LSTM, concatenates outputs corresponding to the same input, and returns the resulting sequence of hidden vectors.

In [2]:
```
myutils.set_GPU(1)
```

2.1 Bias in vSRL

In [3]:
```
margin = 0.05
vSRL = 1
is_dev = 1
myutils.show_amplified_bias(margin, vSRL, is_dev)
```

```
start loading potential files
. . . . . . . . . . .
Finish loading dev potential files
```



imSitu is biased.

In the above figure, along the x-axis, we show the male favoring bias of imSitu verbs. Overall, the dataset is heavily biased toward male agents, with 64.6% of verbs favoring a male agent by an average bias of 0.707 (roughly 3:1 male). Nearly half of verbs are extremely biased in the male or female direction: 46.95% of verbs favor a gender with a bias of at least 0.7. Also it contains several activity labels revealing problematic biases. For example, "shopping" and "washing" are biased toward a female agent. Furthermore, several verbs such as "driving" and "coaching" are heavily biased toward a male agent.

Training on imSitu amplifies bias.

Also in this figure, along the y-axis, we show the ratio of male agents (% of total people) in predictions on an unseen development set. The mean bias ampli- fication in the development set is high, 0.050 on average, with 45.75% of verbs exhibiting amplification. Biased verbs tend to have stronger amplification: verbs with training bias over 0.7 in either the male or female direction have a mean amplification of 0.072. Several already problematic biases have gotten much worse. For example, serving, only had a small bias toward females in the training set, 0.402, is now heavily biased toward females, 0.122.

Paper presentation

- Each group has 2~3 students
- Read and understand 2~3 related papers
  - Cannot be the same as your paper summary
  - Can be related to your final project
  - Register your choice next week
- 30 min presentation/ Q&A
- Grading Rubric:
  40% technical understanding, 40% presentation, 20% interaction

Final Project

- Work in groups (3 students)
- Project proposal
  - 1 page maximum (template)
- Project report
  - Similar to the paper summary
  - Due before the final presentation
- Project presentation
  - in-class presentation (tentative)

Late Policy

- Submission site will be closed 1hr after the deadline.
- No late submission
  - unless under emergency situation

Cheating/Plagiarism

- No. Ask if you have concerns
- Rules of thumb:
  - Cite your references
  - Clearly state what are your contributions

Lectures and office hours

- Participation is highly appreciated!
  - Ask questions if you are still confusing
  - Feedbacks are welcomed
  - Lead the discussion in this class
  - Enroll Piazza

Topics of this class

- Fundamental NLP problems
- Machine learning & statistical approaches for NLP
- NLP applications
- Recent trends in NLP

What to Read?

- Natural Language Processing
  ACL, NAACL, EACL, EMNLP, CoNLL, Coling, **TACL**
  aclweb.org/anthology
- Machine learning
  ICML, NIPS, ECML, AISTATS, ICLR, **JMLR**, **MLJ**
- Artificial Intelligence
  AAAI, IJCAI, UAI, **JAIR**

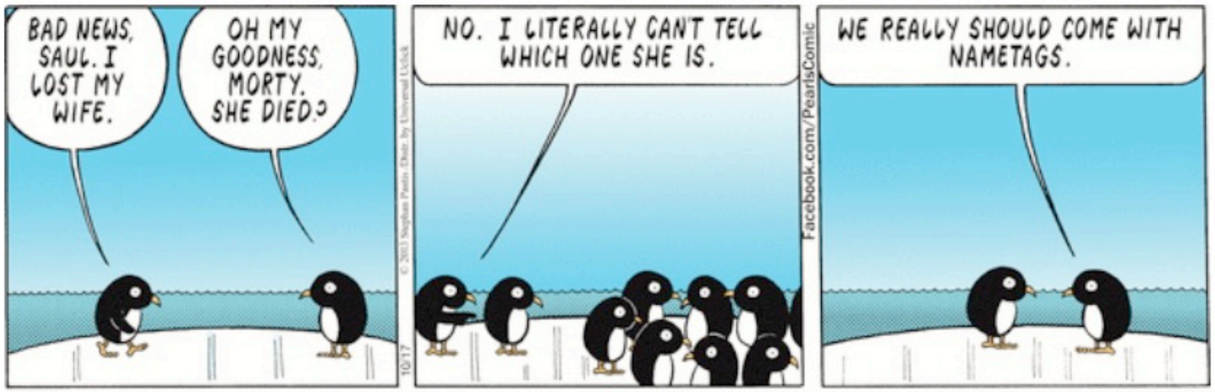
Questions?

This lecture

- Course Overview
  - What is NLP? Why it is important?
  - What will you learn from this course?
- Course Information
- What are the challenges?
- Key NLP components
- Key ML ideas in NLP

Challenges – ambiguity

- Word sense ambiguity



credit: A. Zwicky

Challenges – ambiguity

- Word sense / meaning ambiguity

Credit: http://stuffsirisaid.com

Challenges – ambiguity

- PP attachment ambiguity

San Jose cops kill man with knife

Close

Text Paper Translate Listen

San Jose cops kill man with knife

Ex-college football player, 23, shot 9 times allegedly charged police at fiancee's home

By Hamed Aleaziz and Vivian Ho

A man fatally shot by San Jose police officers while allegedly charging at them with a knife was a 23-year-old former football player at De Anza College in Cupertino who was distraught and depressed, his family said Thursday.

Police officials said two officers opened fire Wednesday afternoon on Phillip Watkins outside his fiancee's home because they feared for their lives. The officers had been drawn to the home, officials said, by a 911 call reporting an armed home invasion that, it turned out, had been made by Watkins himself.

But the mother of Watkins' fiancee, who also lives in the home on the 1300 block of Sherman Street, said she witnessed the shooting and described it as excessive. Faye Buchanan said the confrontation happened shortly after she called a suicide intervention hotline in hopes of getting Watkins medical help.

Watkins' 911 call came in at 5:01 p.m., said Sgt. Heather Randol, a San Jose police spokeswoman. "The caller stated there was a male breaking into his home armed with a knife," Randol said. "The caller also stated he was locked in an upstairs bedroom with his children and requested help from police."

She said Watkins was on the sidewalk in front of the home when two officers got there. He was holding a knife with a 4-inch blade and ran toward the officers in a threatening manner, Randol said.

"Both officers ordered the suspect to stop and drop the knife," Randol said. "The suspect continued to charge the officers with the knife in his hand. Both officers, fearing for their safety and defense of their life, fired at the suspect."

On the police radio, one officer said, "We have a male with a knife. He's walking toward us."

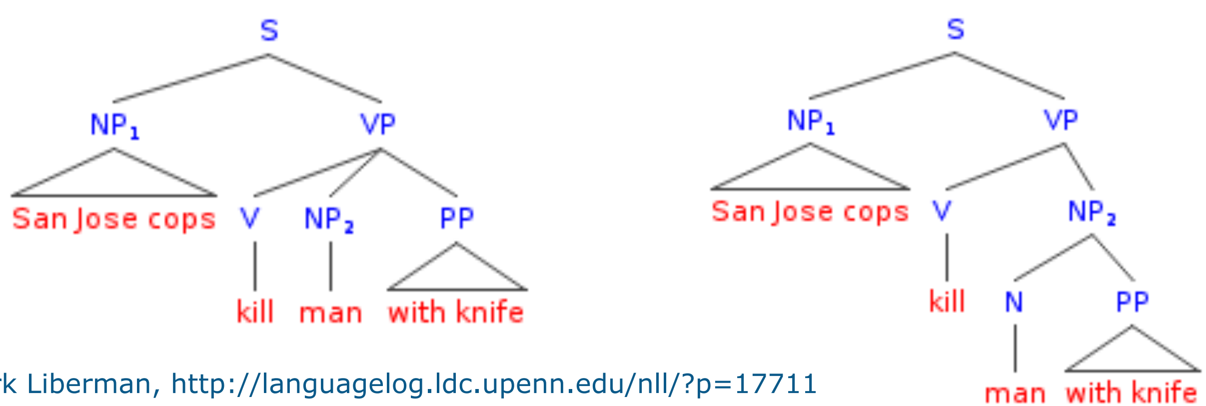
"Shots fired! Shots fired!" an officer said moments later.

A short time later, an officer reported, "Male is down. Knife's still in hand."

Buchanan said she had been prompted to call the

Shoot continues on D8

Back Continue



Credit: Mark Liberman, http://languagelog.ldc.upenn.edu/nll/?p=17711

Challenges -- ambiguity

- Ambiguous headlines:
  - Include your children when baking cookies
  - Local High School Dropouts Cut in Half
  - Hospitals are Sued by 7 Foot Doctors
  - Iraqi Head Seeks Arms

  - Safety Experts Say School Bus Passengers Should Be Belted
  - Teacher Strikes Idle Kids

Challenges – ambiguity

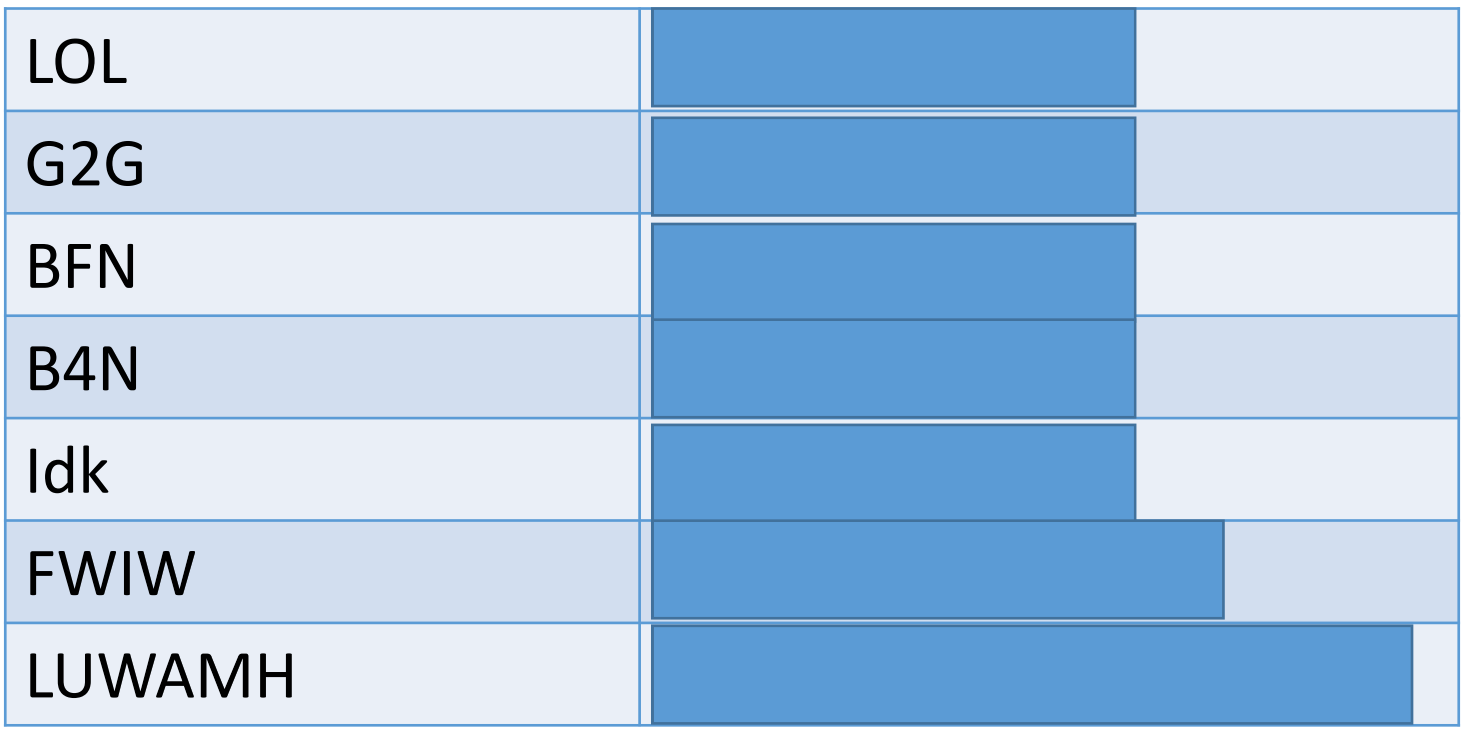
- Pronoun reference ambiguity

Dr. Macklin often brings his dog Champion to visit with the patients. He just loves to give big, wet, sloppy kisses!

Credit: http://www.printwand.com/blog/8-catastrophic-examples-of-word-choice-mistakes

Challenges – language is not static

- Language grows and changes
  - e.g., cyber lingo

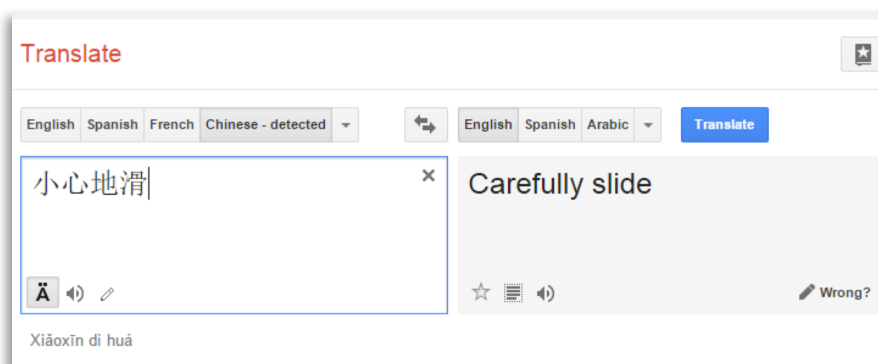


Challenges--language is compositional

Challenges--language is compositional

小心:
Carefully
Careful
Take
Care
Caution

地滑:
Slide
Landslip
Wet Floor
Smooth



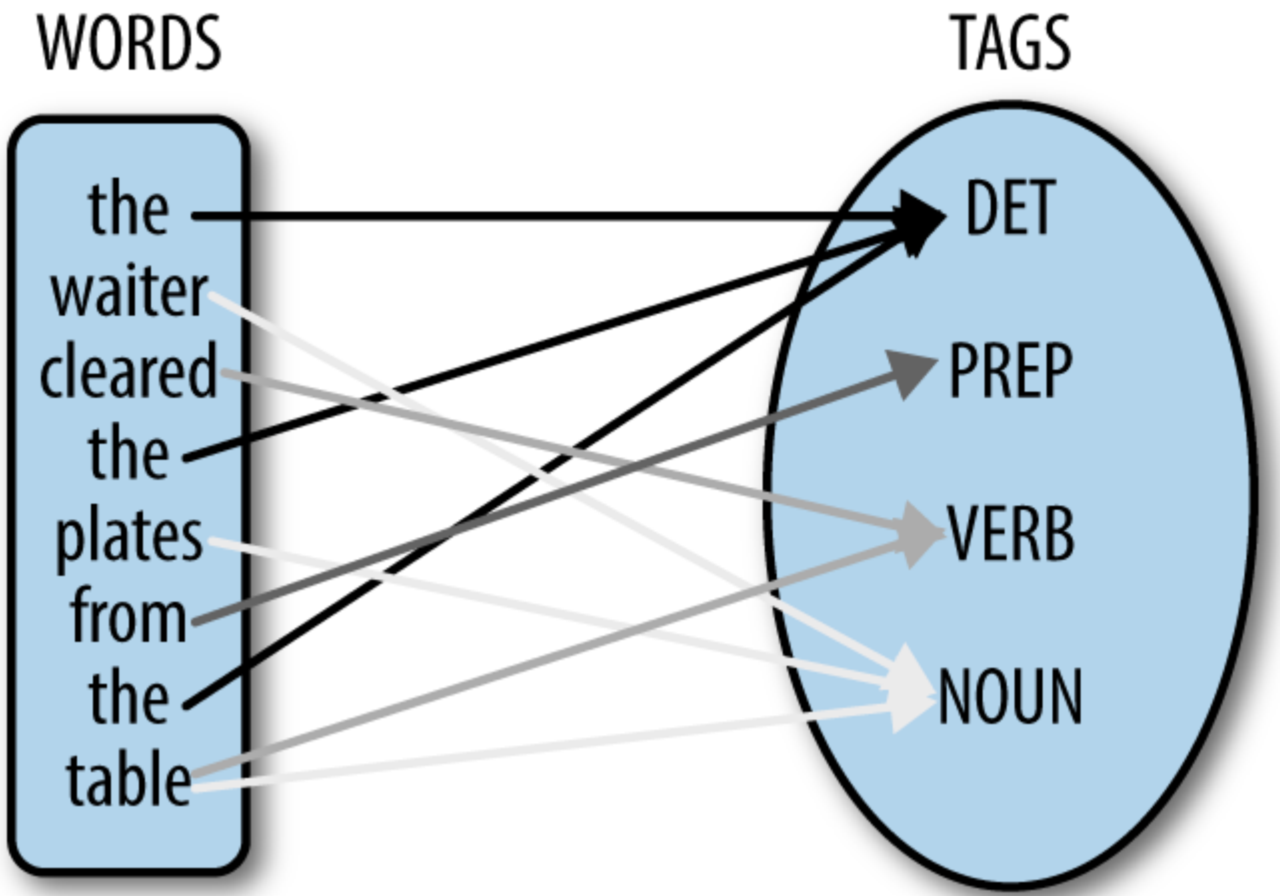
Challenges – scale

- Examples:
  - Bible (King James version): ~700K
  - Penn Tree bank ~1M from Wall street journal
  - Newswire collection: 500M+
  - Wikipedia: 2.9 billion word (English)
  - Web: several billions of words

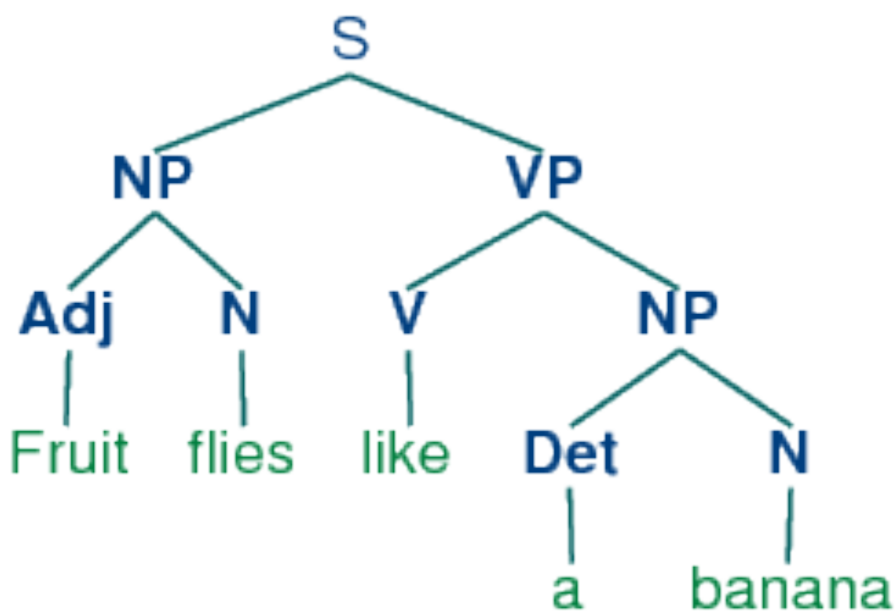
This lecture

- Course Overview
  - What is NLP? Why it is important?
  - What will you learn from this course?
- Course Information
- What are the challenges?
- Key NLP components
- Key ML ideas in NLP

Part of speech tagging



Syntactic (Constituency) parsing



Syntactic structure => meaning

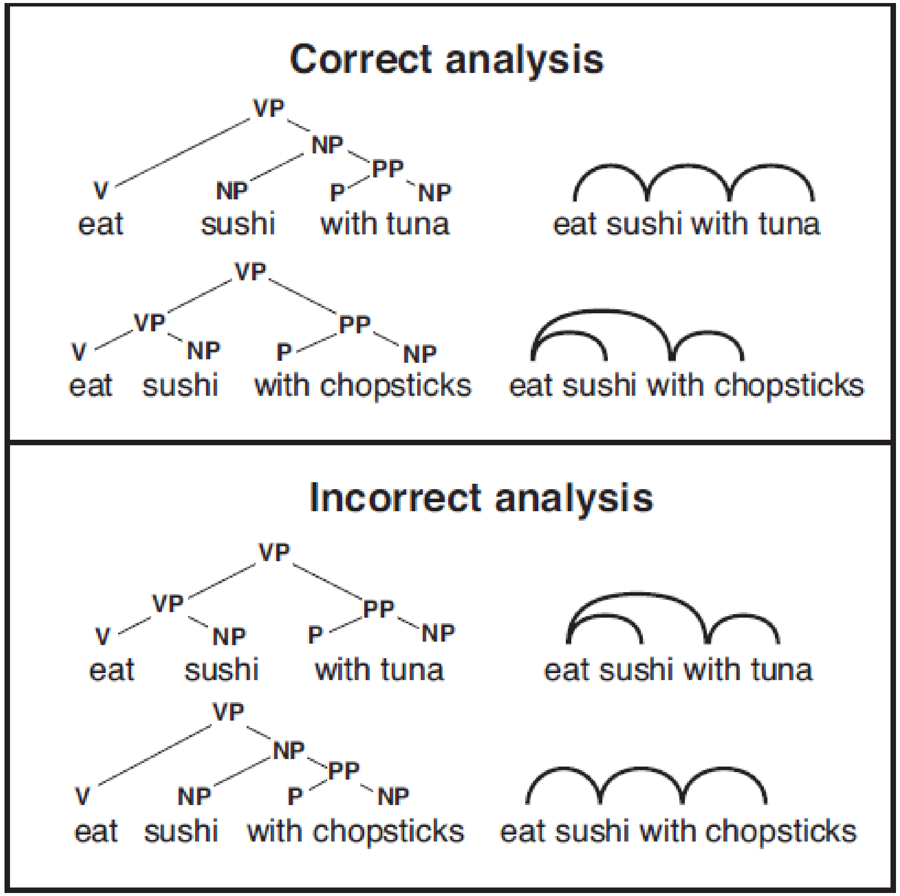
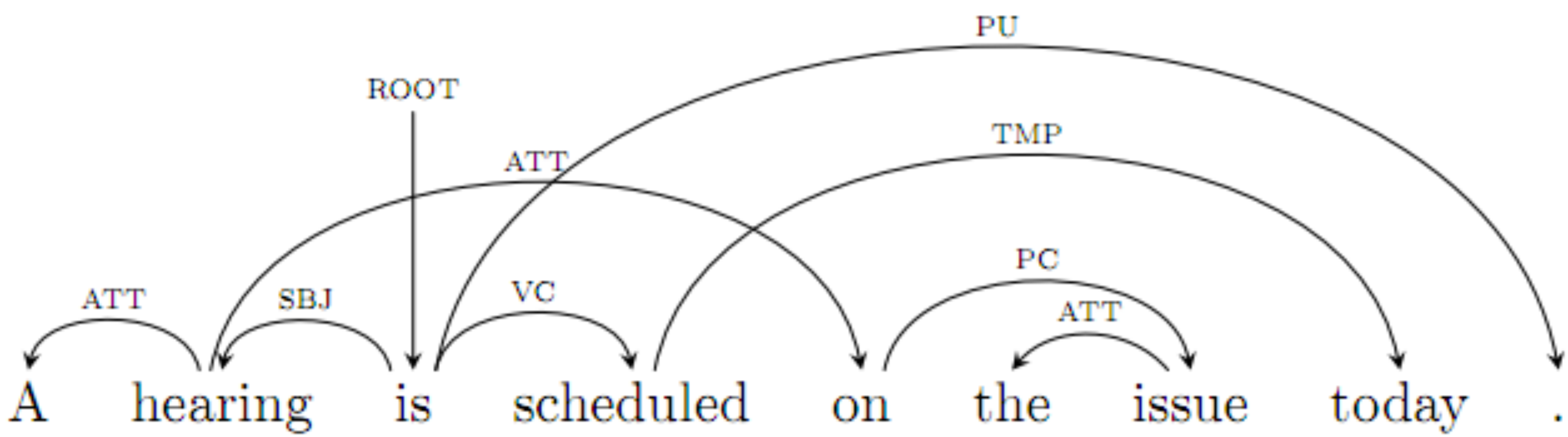


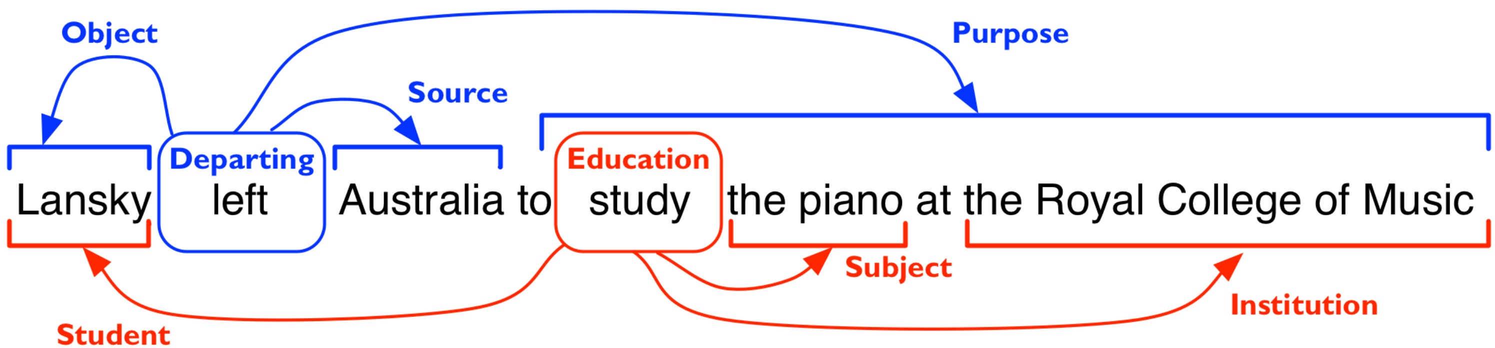
Image credit: Julia Hockenmaier, Intro to NLP

Dependency Parsing



Semantic analysis

- Word sense disambiguation
- Semantic role labeling



Credit: Ivan Titov

Q: [Chris] = [Mr. Robin] ?

Christopher Robin is alive and well. **He** is the same person that you read about in the book, **Winnie the Pooh.** As **a boy**, **Chris** lived in a pretty home called **Cotchfield Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book

Slide modified from Dan Roth

Co-reference Resolution

Christopher Robin is alive and well. **He** is the same person that you read about in the book, **Winnie the Pooh**. As a **boy**, **Chris** lived in a pretty home called **Cotchfield Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book

This lecture

- Course Overview
  - What is NLP? Why it is important?
  - What will you learn from this course?
- Course Information
- What are the challenges?
- Key NLP components
- Key ML ideas in NLP

Machine learning 101

Input

$\mathbf{x} \in \mathcal{X}$

An item **x** drawn from an input space $\mathcal{X}$

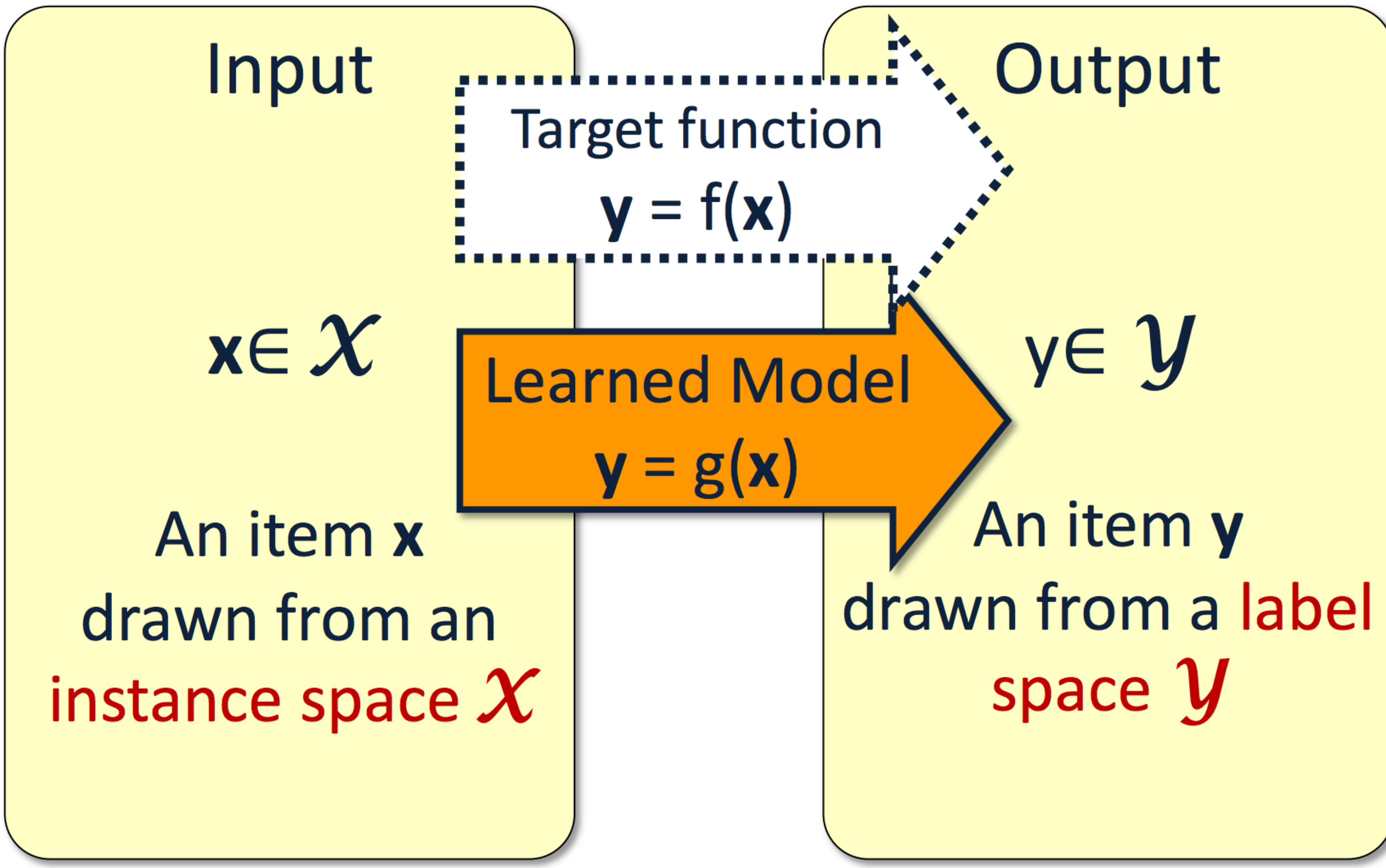
System

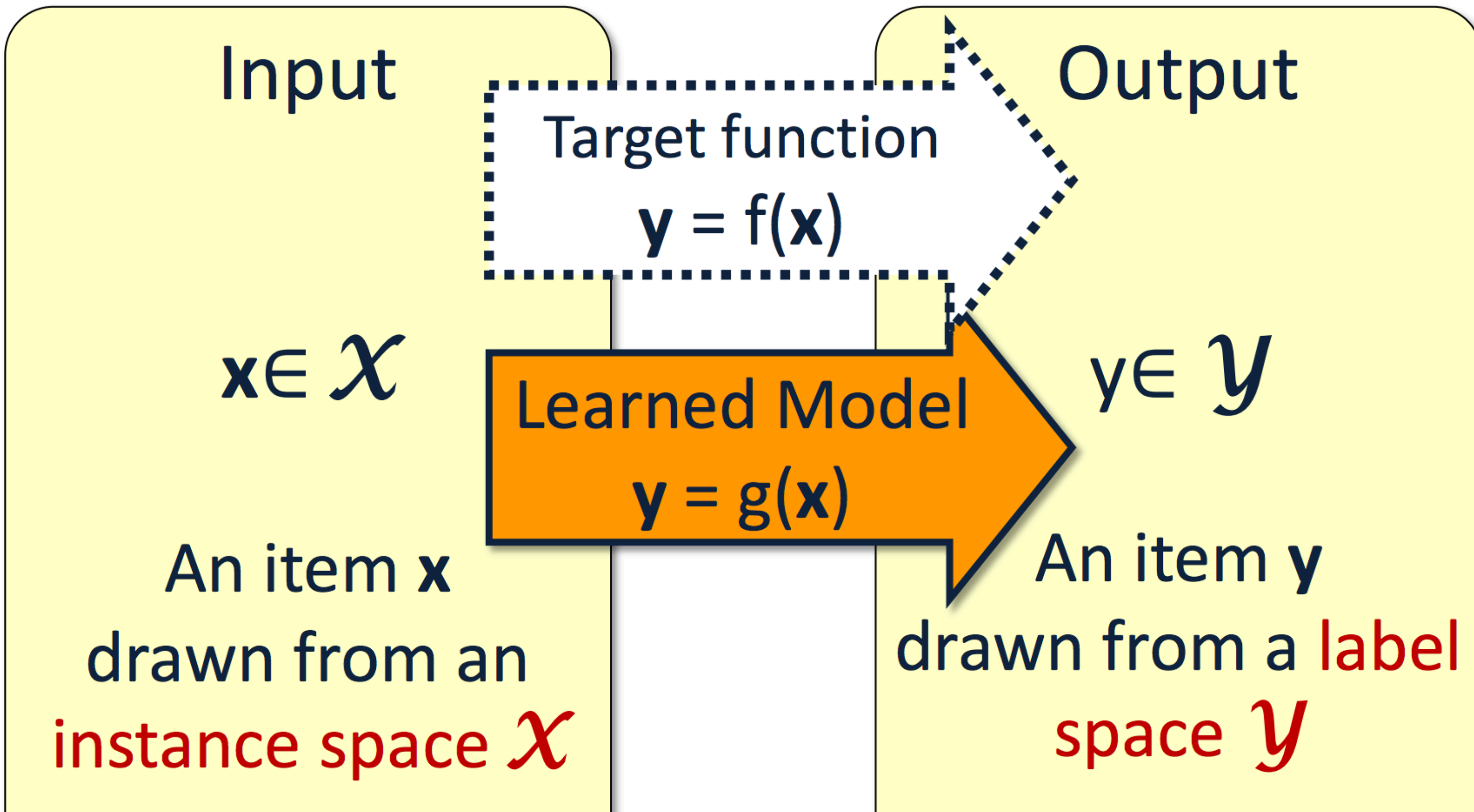
$y = f(\mathbf{x})$

Output

$y \in \mathcal{Y}$

An item **y** drawn from an output space $\mathcal{Y}$


Input
$\mathbf{x} \in \mathcal{X}$
An item **x**
drawn from an
instance space $\mathcal{X}$
Target function
$\mathbf{y} = f(\mathbf{x})$
Learned Model
$\mathbf{y} = g(\mathbf{x})$
Output
$y \in \mathcal{Y}$
An item **y**
drawn from a label
space $\mathcal{Y}$



Perceptron, decision tree, support vector machine
K-NN, Naïve Bayes, logistic regression….

Classification is generally well-understood

- Theoretically: generalization bound
  - # examples to train a good model
- Algorithmically:
  - Efficient algorithm for large data set
    - E.g., take a few second to train a linear SVM on data with millions instances and features
  - Algorithms for non-linear model
    - E.g., Kernel methods

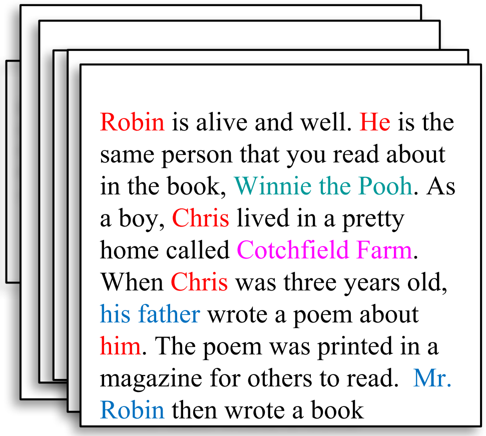
Is this enough to solve all real-world problems?

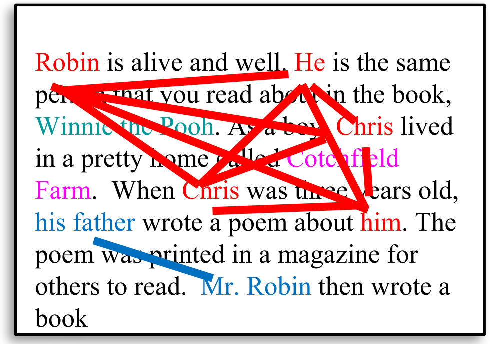
Reading Comprehension

(ENGLAND, June, 1989) - Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now

Challenges





- Modeling challenges

  Structured prediction models

  - How to model a complex decision?
- Representation challenges

  Deep learning models

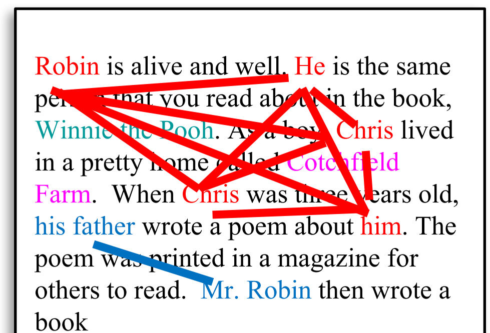
  - How to extract features?
- Algorithmic challenges

  Inference / learning algorithms

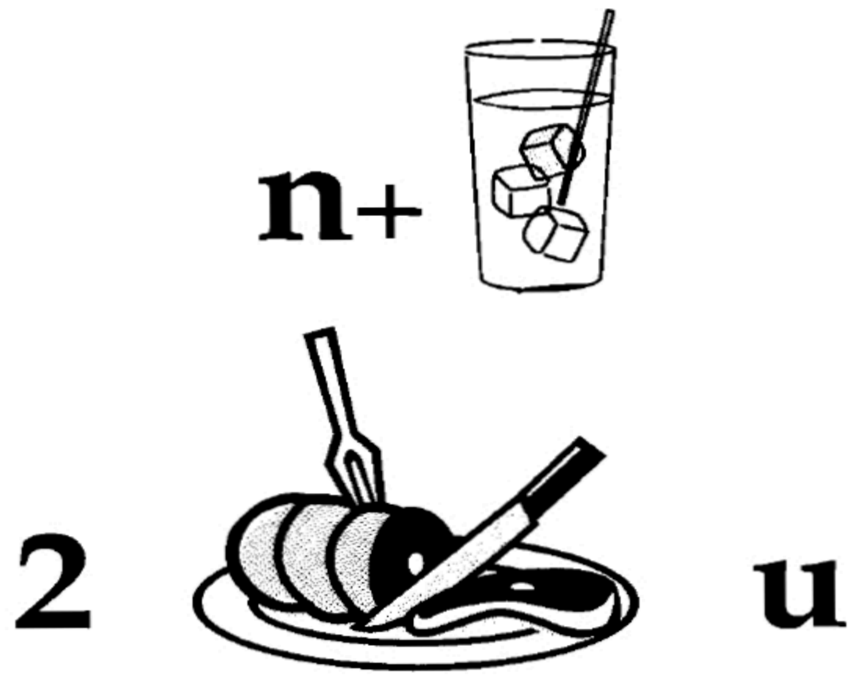
  - Large amount of data and complex decision structure

Modeling Challenges

- How to model a complex decision?
- Why this is important?



Language is structural

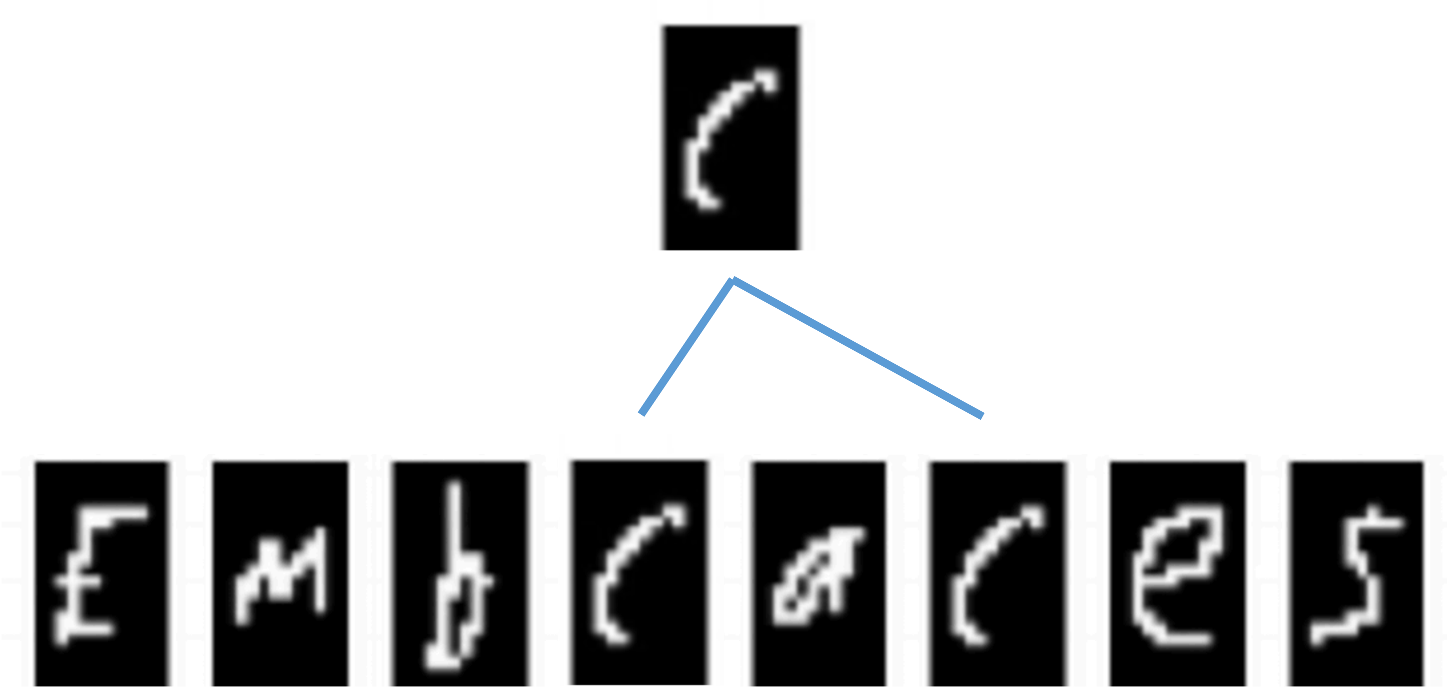


Hand written recognition

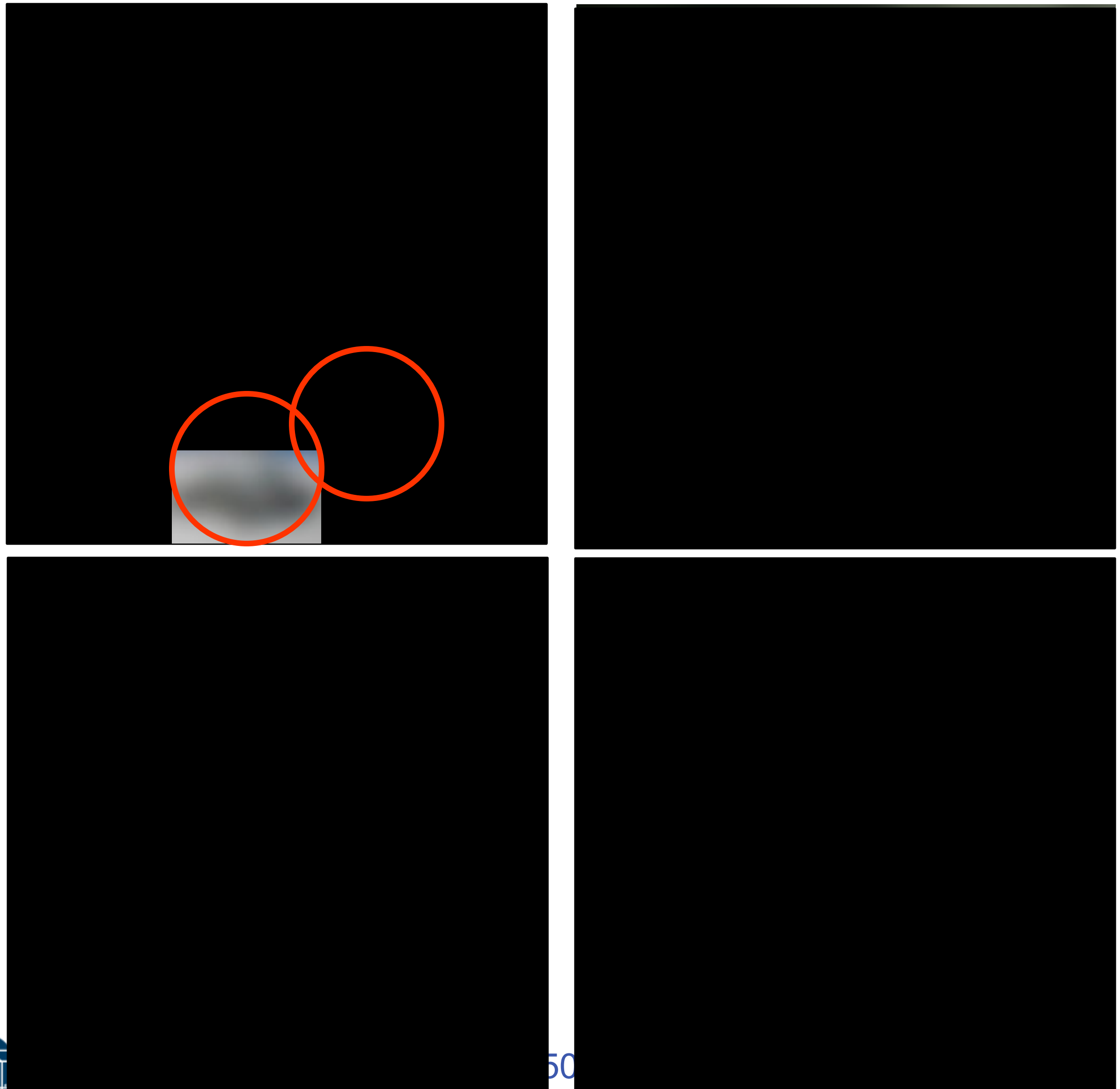
- What is this letter?

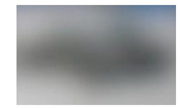
Hand written recognition

- What is this letter?

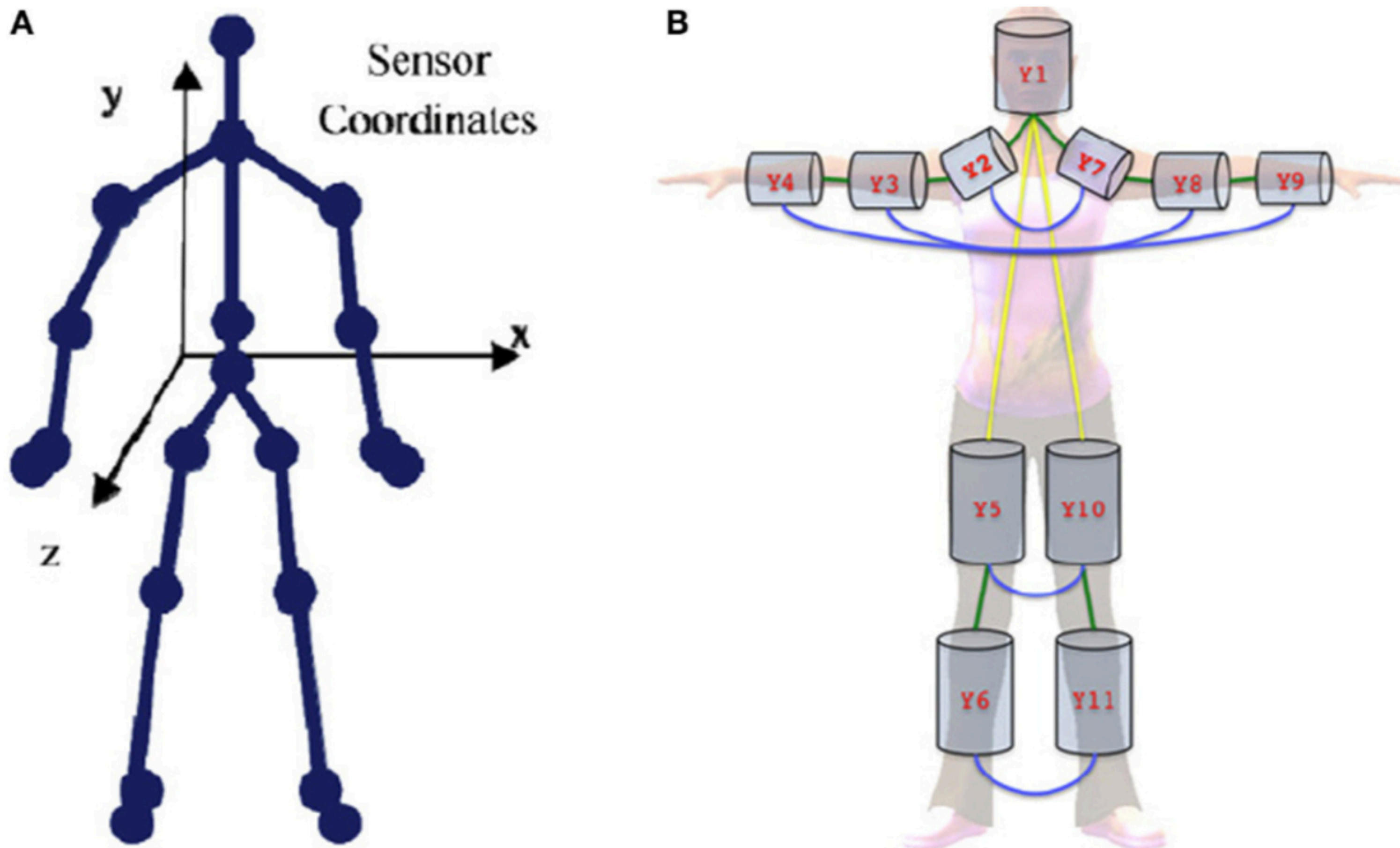


Visual recognition





Human body recognition



Bridge the gap

- Simple classifiers are not designed for handle complex output
- Need to make multiple decisions jointly
- Example: POS tagging:

can you can a can as a canner can can a can

Example from Vivek Srikumar

Make multiple decisions jointly

- Example: POS tagging:

can you can a can as a canner can can a can

- Each part needs a label
  - Assign tag (V., N., A., …) to each word in the sentence
- The decisions are mutually dependent
  - Cannot have verb followed by a verb
- Results are evaluated jointly

Structured prediction problems

- Problems that
  - have multiple interdependent output variables
  - and the output assignments are evaluated jointly
- Need a joint assignment to all the output variables
  - We called it joint inference, global infernece or simply inference

A General learning setting

- Input: $x \in X$
- Truth: $y^* \in Y(x)$
- Predicted: $h(x) \in Y(x)$
- Loss: $loss(y, y^*)$

I	can	can	a	can
Pro	Md	Vb	Dt	Nn
Pro	Md	Md	Dt	Vb
Pro	Md	Md	Dt	Nn
Pro	Md	Nn	Dt	Md
Pro	Md	Nn	Dt	Vb

Goal: make joint prediction to minimize a joint loss

find $h \in H$ such that $h(\mathrm{x}) \in Y(X)$
minimizing $E_{(x,y)\sim D}\left[loss\left(y, h(x)\right)\right]$ based on N
samples $(x_n, y_n) \sim D$

Combinatorial output space

- Input: $x \in X$
- Truth: $y^* \in Y(x)$
- Predicted: $h(x) \in Y(x)$
- Loss: $loss(y, y^*)$

I	can	can	a	can
Pro	Md	Vb	Dt	Nn
Pro	Md	Md	Dt	Vb
Pro	Md	Md	Dt	Nn
Pro	Md	Nn	Dt	Md
Pro	Md	Nn	Dt	Vb

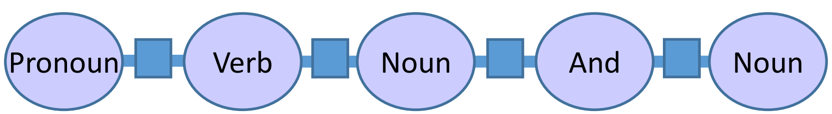
POS tags: 45

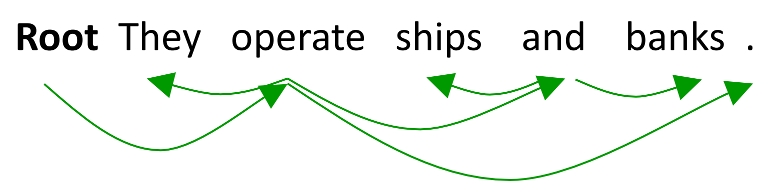
How many possible outputs for sentence with 10 words? $45^{10} = 3.4 \times 10^{16}$

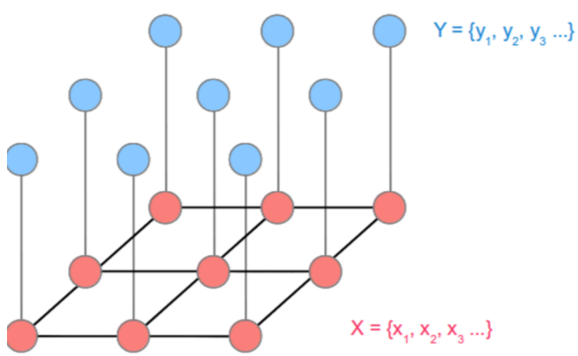
Observation: Not all sequences are valid, and we don't need to consider all of them

Representation of interdependent output variables

- A compact way to represent output combinations
  - Abstract away unnecessary complexities
  - We know how to process them
    - Graph algorithms for linear chain, tree, etc.







Algorithms/models for structured prediction

- Many learning algorithms can be generalized to the structured case
  - Perceptron → Structured perceptron
  - SVM → Structured SVM
  - Logistic regression → Conditional random field (a.k.a. log-linear models)
- Can be solved by a reduction stack
  - Structured prediction → multi-class → binary

Representation Challenges

- How to obtain features?

Robin is alive and well. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book

Representation Challenges

- How to obtain features?
  1. Design features based on domain knowledge
     - E.g., by patterns in parse trees

       When Chris was three years old, his father wrote a poem about him.

     - By nicknames

       Christopher Robin is alive and well. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm.

  - Need human experts/knowledge

Representation Challenges

- How to obtain features?
  1. Design features based on domain knowledge
  2. Design feature templates and then let machine find the right ones
     - E.g., use all words, pairs of words, …

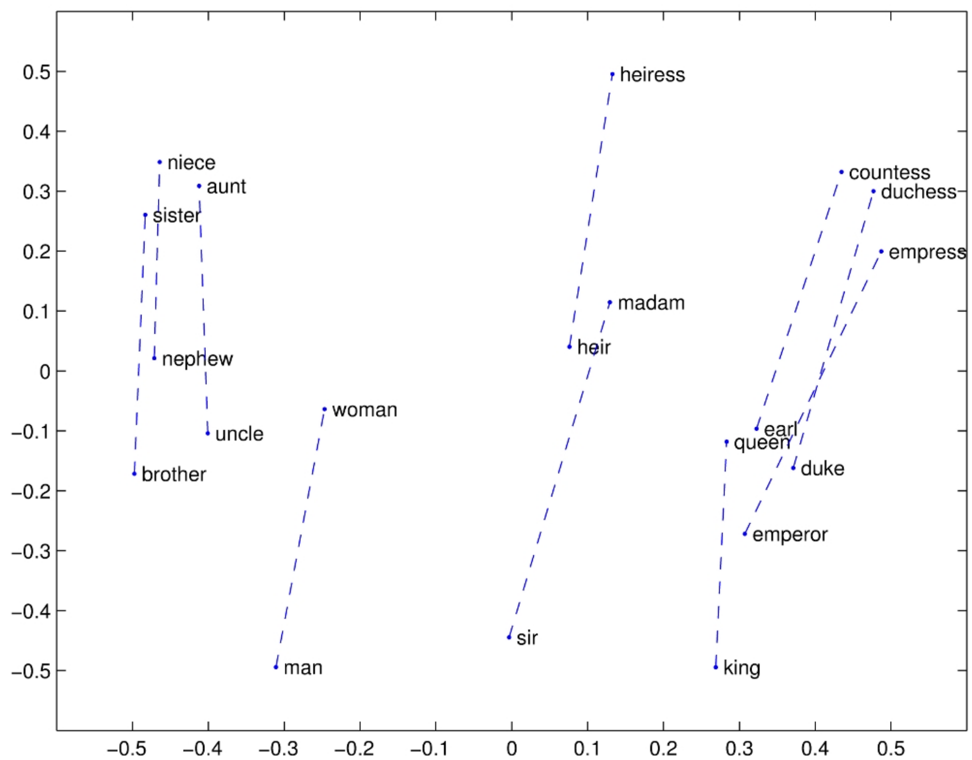
Robin is alive and well. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book

Representation Challenges

- How to obtain features?
  1. Design features based on domain knowledge
  2. Design feature templates and then let machine find the right ones
- Challenges:
  - # featuers can be very large
    - # English words: 171K (Oxford)
    - # Bigram: $(171K)^2 \sim 3\times 10^{10}$, # trigram?
  - For some domains, it is hard to design features

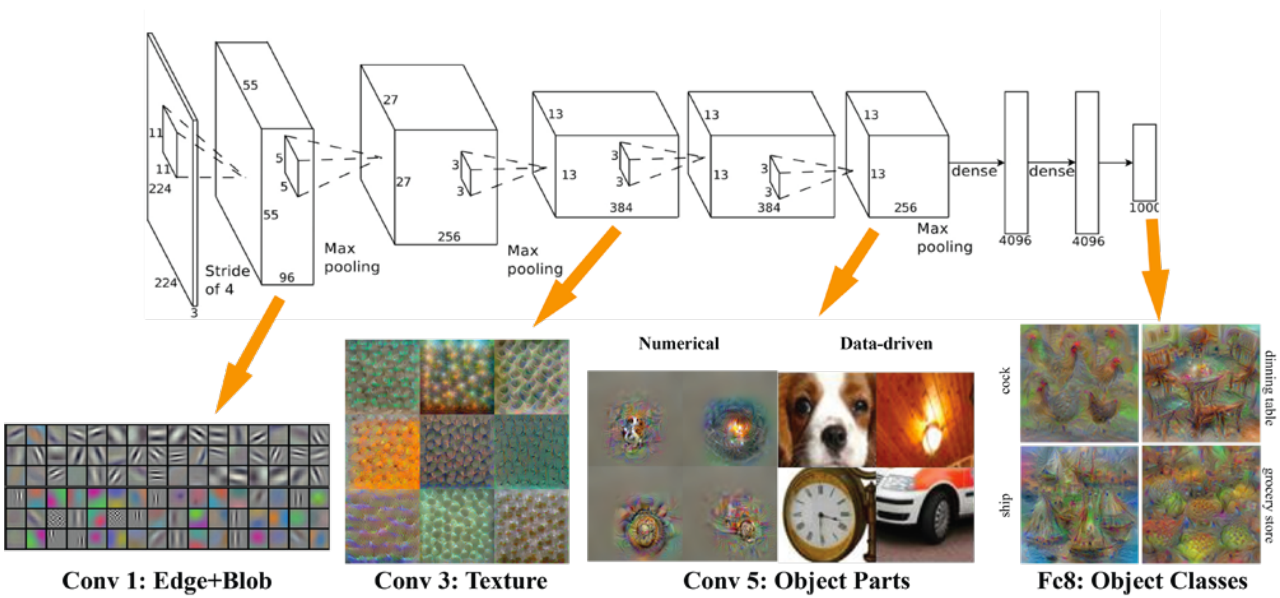
Representation learning

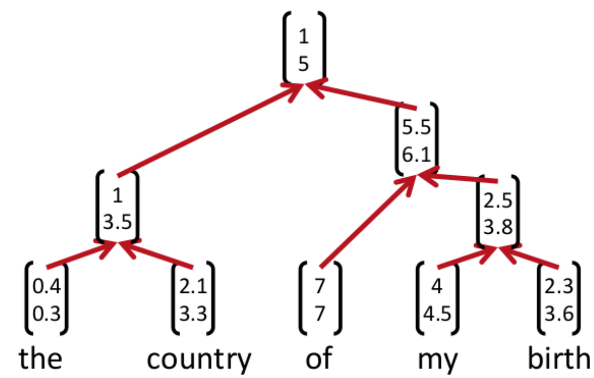
- Learn compact representations of features
  - Combinatorial (continuous representation)



Representation learning

- Learn compact representations of features
  - Combinatorial (continuous representation)
  - Hieratical/compositional





What will learn from this course

- Structured prediction
  - Models / inference/ learning
- Representation (deep) learning
  - Input/output representations
- Combining structured models and deep learning