Machine Learning Method for Natural Language Processing

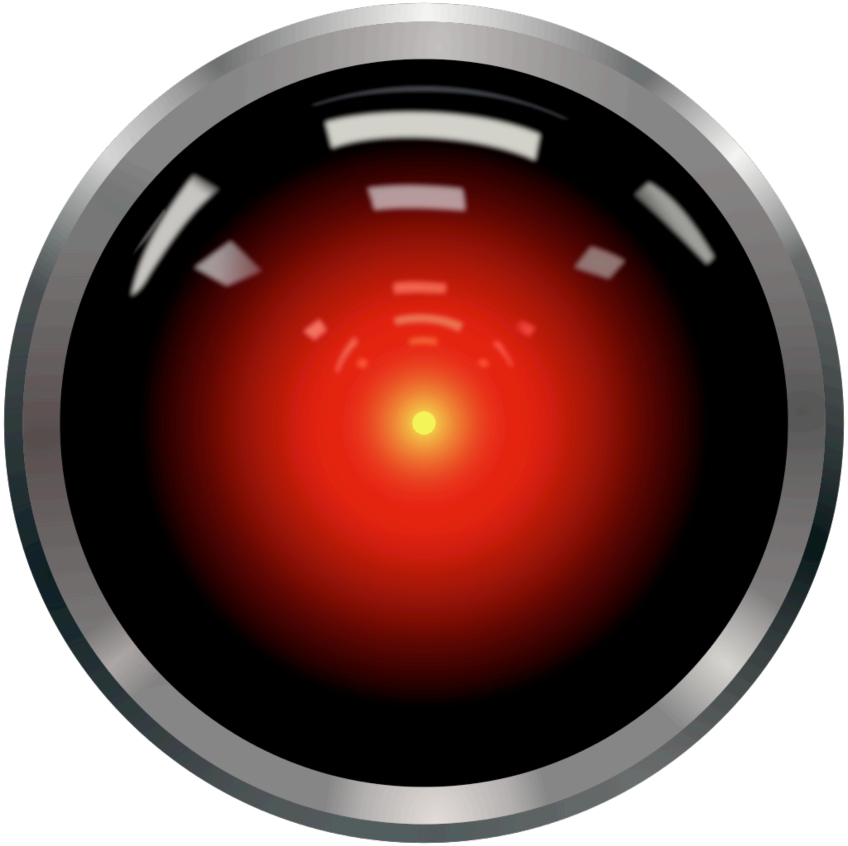
Dan Goldwasser

Purdue University

dgoldwas@purdue.edu

CS 590NLP

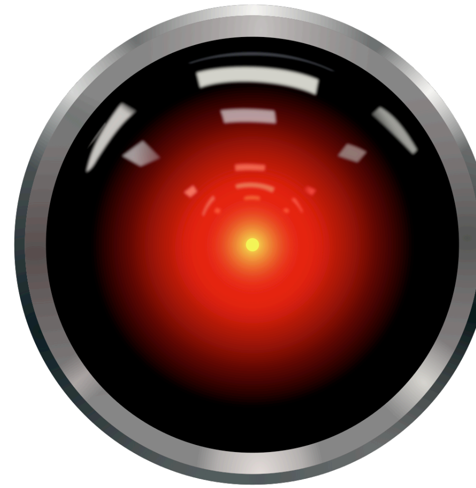
2001



Maybe it’s better to start with where we were supposed to be 16 years ago, **according to 1960’s sci-fi**

Dave : *Open the pod bay doors, HAL.*

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



Clearly unrealistic! Hal understood:

(1) the meaning of the words
(2) grounded their meaning in a physical environment
(3) understood the situation and intent of the speaker

Now consider the "perfect" personal assistant on your smartphone stopping you from sending that *ANGRY* email to your boss.

Realistic?

NLP research is about pushing the limits of realistic applications

NLP in Practice

- Personal assistants that interact in natural language
  - *Simple voice activation turned into complex language analysis*
  - Several recent high profile applications
  - *Interaction is often humorous and enjoyable*

Purdue *in the News*

Google

purdue news

Web News Shopping Images Videos More Search tools

About 73,100 results

About 73,100 results (0.31 seconds)

Purdue's Discovery Park launches global soundscapes re...
Purdue Newsroom - Oct 29, 2014
WEST LAFAYETTE, Ind. - **Purdue** University ecologist Bryan Pijanowski gained international attention for an Earth Day effort to capture ...

Purdue volleyball loses to Illinois
wlfi.com - Oct 25, 2014
CHAMPAIGN, Ill. (**Purdue** Sports) — The No. 13 Boilermaker volleyball team battled No. 10 Illinois to the wire in three sets of a four-set loss on ...
Illinois Volleyball block beats Indiana, No. 13 **Purdue**
Daily Illini - Oct 26, 2014

NLP in Practice

- **Information Extraction**
  - Parse unstructured text into structured information
  - Now a standard part of most email services

Hi Dan,
I just wanted to let you know that we scheduled the meeting for Monday 9:30, at the Lawson Commons. It will take two hours.

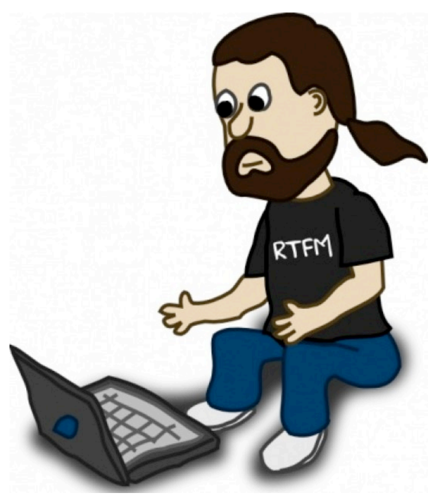
Event: Meeting
Date: Monday, Sep 15
Start: 9:30am
End: 11:30
Location: Lawson Commons

NLP in Practice

- **Sentiment Analysis**
  - Meaningful interpretation of product reviews
  - Identify the product aspects users care about
  - *Deception detection*

I just bought ***company-A*** newest laptop. The display is ***awesome***, the speakers are ***not that great***. I'm ***happy*** with the performance, but I think they charge too much for it!

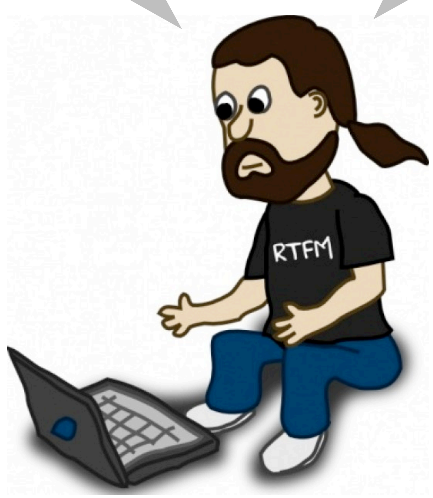
Display: Positive
Speakers: Negative
Performance: Positive
Price: Negative



NLP can be Very Challenging!

Dude, I just watched this horror flick! Selling points: nightmares scenes, torture scenes, terrible monsters that was so bad a##!

Don’t buy the popcorn it was terrible, the monsters selling it must have wanted to torture me, it was so bad it gave me nightmares!



NLP can be Very Challenging!

Kiel first encountered Moore's James Bond in 1977's "The Spy Who Loved Me," where his silent hitman Jaws repeatedly menaced Bond with his sharp metal teeth. Although repeatedly thwarted by the British spy, Jaws proved resilient and even sort of likable: Near the movie's end he survived a brush with a killer shark by biting the creature.

Jaws was such a popular character that the producers of the Bond series brought him back two years later for "Moonraker," which was set partly in space. He and Bond battled each other in a opening skydiving sequence and in a memorable scene atop an aerial tram in Rio de Janeiro. Later, Jaws switched allegiances to Bond upon learning that his employer, the villain Drax, planned to exterminate him.

Opinion: Why Jaws was best 'Bond' villain ever

In later years, Kiel turned his hand to writing and producing as well as acting, including in the 1991 movie "The Giant of Thunder Mountain," according to the IMDB.

He also had a small role in "Happy Gilmore," the 1996 Adam Sandler golf comedy.

"Richard Kiel was one of the nicest, funniest guys I've ever met. I'll never forget hanging out with him & how good he was to everyone," Sandler tweeted Thursday.

Kiel's family posted this message on Facebook:

"It is with very heavy hearts that we announce that Richard has passed away, just three days shy of his 75th birthday. Richard had an amazing joy for life and managed to live every single day to the fullest. Though most people knew of him through his screen persona, those who were close to him knew what a kind and generous soul he was.

"His family was the most important thing in his life and we are happy that his last days were spent surrounded by family and close friends. Though his passing was somewhat unexpected, his health had been declining in recent years. It is nice to think that he can, once again, stand tall over us all."

Summarization: *Capture the essence of a long text*

Richard Kiel, actor who played Bond villain "Jaws" dies at 74

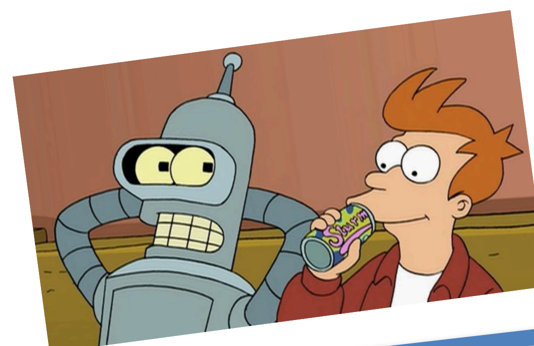
Summarization is still an open problem

Representative of difficult problems in NLP:

- *Word-level decisions* ***VS*** *high level tasks that require deeper understanding*

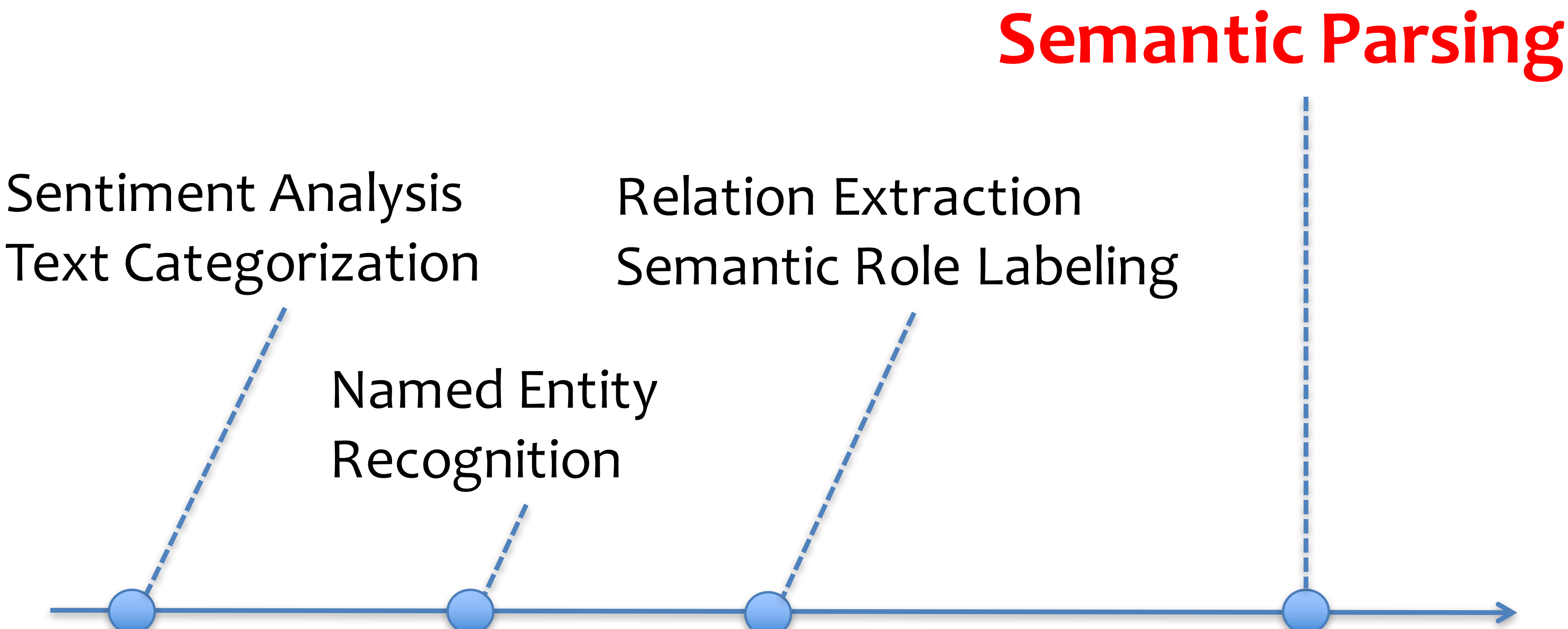
The Meaning of *Meaning*

NLP is about $\underbrace{\text{extracting structured information}}_{\textit{extracting meaning from text}}$ from unstructured data



Meaning

The Meaning of Meaning



The food was OK, ... quite disappointing. [**NEGATIVE**]

The president ... Obama ***PERSON***

[Bank ***A0****] [gave* ***Pr****] [the*

$density(argmax(x,city(x)\wedge loc(x,IN),population(x,y)))$

Allows direct interaction with machines. E.g., NL Db access

More than Words

Natural language is inherently ambiguous

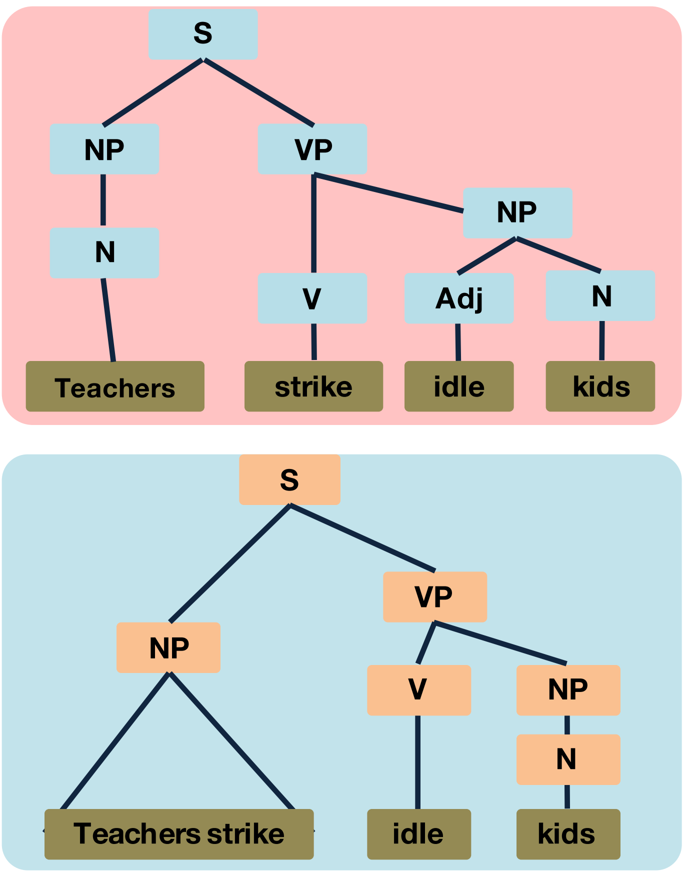
I just watched this horror flick! .. nightmares ..torture .., terrible ..monsters

Horror movie + nightmare = Good

don't buy the popcorn it was terrible.. monsters …torture ..nightmares!

Food + nightmare = Bad

"Teachers strike idle kids"



More than Words

- Natural language is inherently ambiguous

Teacher strikes idle kids

Hospital sued by 7 foot doctors

Local High School Dropouts Cut in Half

Reading Comprehension

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

Let’s answer some questions..

Adapted from J.Eisner

Reading Comprehension

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

Did John get :

(a) Something to eat?

(b) Spare tire

Adapted from J.Eisner

Reading Comprehension

John stopped at the ***donut store*** on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

Is the store:

(a) Is it run by donuts
(b) A shopping center for donuts
(c) Made of donuts
(d) Sells donuts

Adapted from J.Eisner

Reading Comprehension

John stopped at the donut store on his way home from work. **He thought a coffee was good every few hours.** But it turned out to be too expensive there.

(a) Is the coffee good every few hours?
- Which coffee?

(b) Did he think about it every few hours?

Similarly: *"In America a woman has a baby every 15 minutes. Our job is to find that woman and stop her."*

Adapted from J.Eisner

Reading Comprehension

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be **too expensive** there.

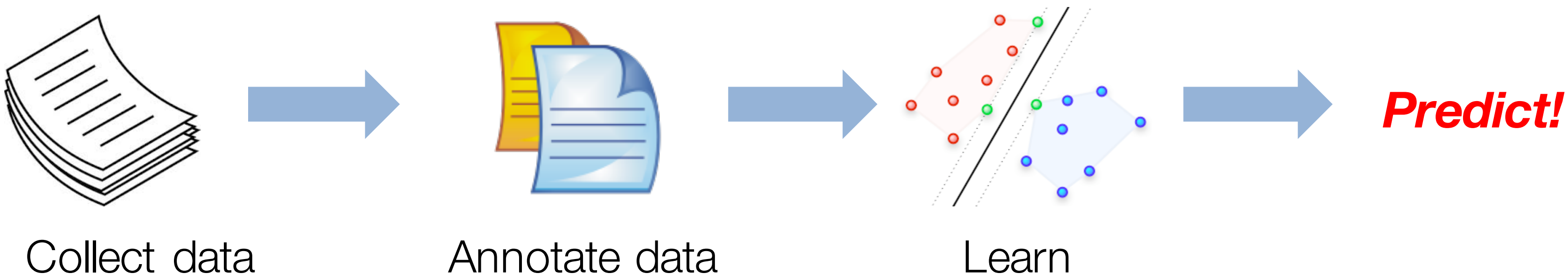
What is too expensive?

In this example,
resolving ambiguity requires a global view!

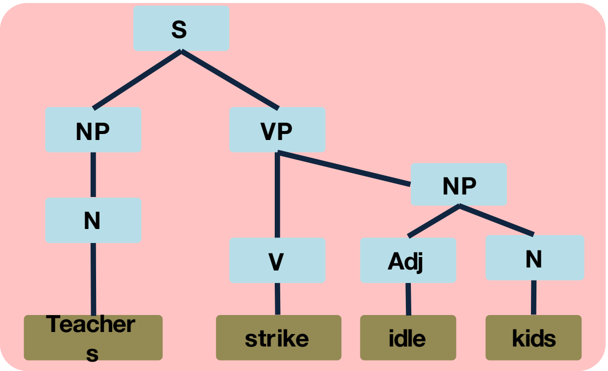
Adapted from J.Eisner

Dealing with Ambiguity

- *Machine learning* is an effective tool for resolving ambiguities
  - Build statistical prediction models based on annotated data

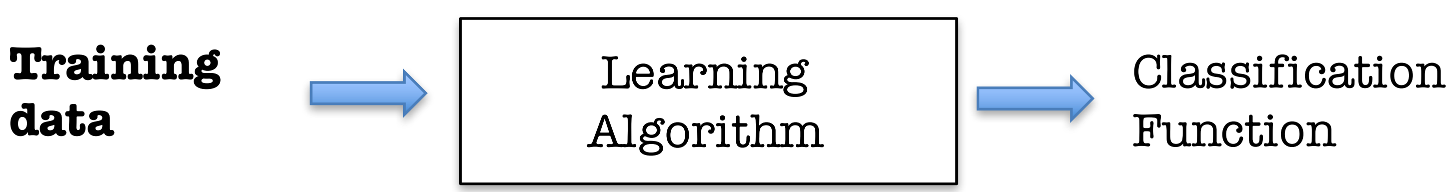


- *Not a magical solution*: learning can be difficult *(e.g., twitter posts do not look like WSJ or NYT articles)*
  - ***Annotating data for high level tasks is difficult!***

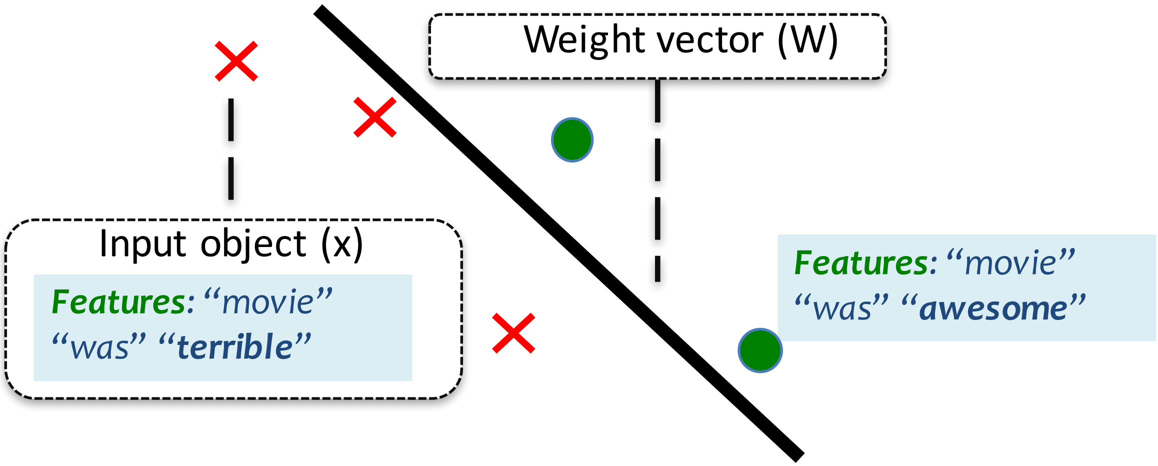


Classification

- **Classification**: mapping data into categories
  - *Determine if an English sentence is grammatical*
  - *Positive or negative sentiment? Mentions Purdue?*
- Can't we just write code?
- Provide **labeled examples** and let a classifier distinguish between the two classes



Classification 101



- **Input Representation**: feature functions $\phi(x)$
  - E.g. *Bag-of-words, N-grams features*
- **Prediction** : $f_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x})$, $f_{\mathbf{w}}(\mathbf{x}) > 0$
- **Learning**: given training data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \frac{\lambda}{2} \|\mathbf{w}\|^2 + \sum_i \ell\left(-y_i f_{\mathbf{w}}(\mathbf{x}_i)\right)$$

Machine Learning for NLP

- ***Text Categorization, Sentiment Analysis***

Sentiment Analysis   *The movie was awesome!* → ***positive/ Negative?***

Text categorization   *The trend in Wall Street …* → ***Sports/ Finance ?***

  – What are the features?

Deep Learning for NLP

- Consider these two examples –
  - *"The movie was not good, bad!"*
  - *"The movie was not bad, good!"*

- What is the problem here?
- How would you fix it?

Deep Learning for NLP

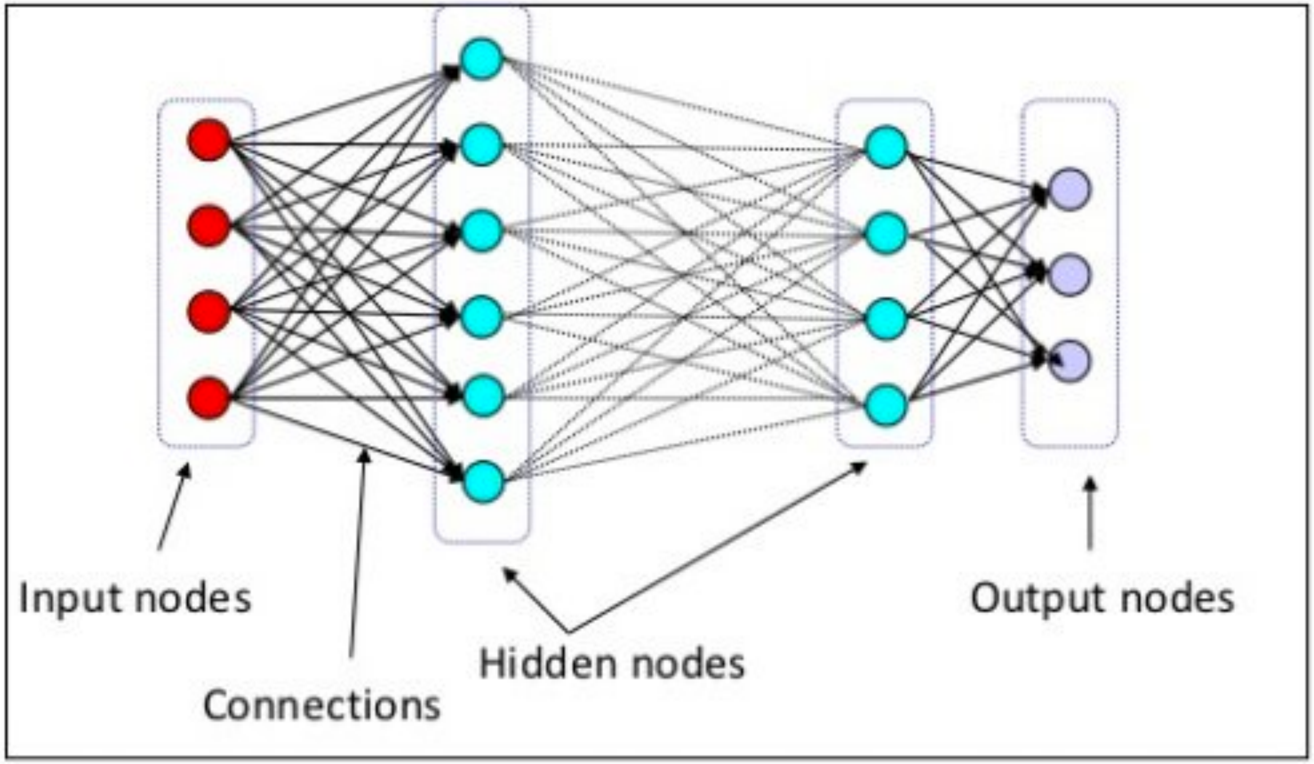
- **The issue:**
  - *linear classifiers over BoW features might not be expressive enough*

- **The solution:**
  - *Learn non-linear classifier*
  - Don’t use BoW features

- Deep learning addresses these two issues directly

Deep Learning for NLP

- Learn a non-linear classifier



Deep Learning for NLP

- Don't use BoW features
  - Deep learning has popularized the use of distributed representations in NLP

	Cucumber	Tomato
Long	X	
Round		X
Green	X	
Red		X

Watermelon = Long + Round + Green + Red

Machine Learning for NLP

- Text Categorization, Sentiment Analysis
- **Named Entity Recognition** (LOC,PER, ORG)

Barak Obama visited Mount Sinai Hospital in New York.

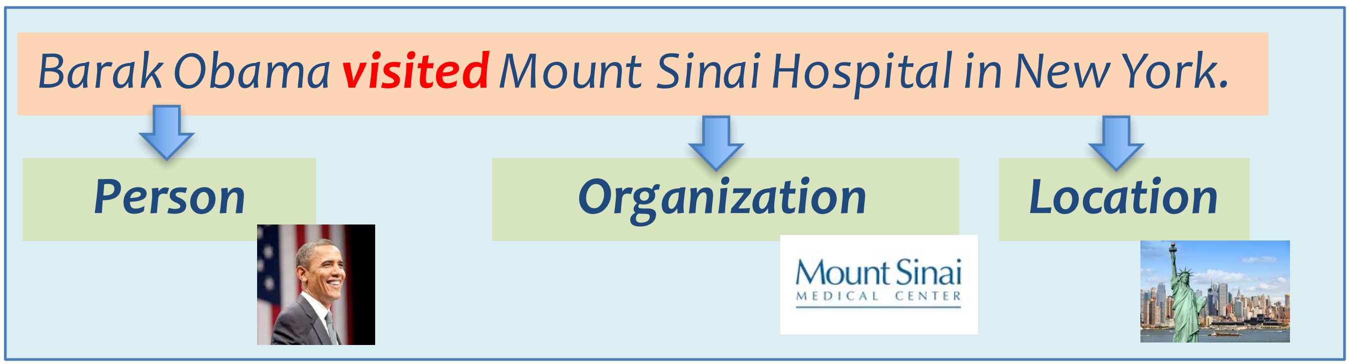
Person ***Organization*** ***Location***

Can **you** predict NER?

What are good features for NER?

Machine Learning for NLP

- Text Categorization, Sentiment Analysis
- Named Entity Recognition
- **Grounding and *Semantic Parsing*** (*relations*)



How many classifiers do we need to train for this task?

- **Supervision bottleneck**
  – Difficult to collect data for such diverse tasks

Semantic Role Labeling

Also known as – *"Who did what to whom, when, where, why,…"*

I *left* my pearls to my daughter in my will .

[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM\text{-}LOC}$.

- ***A0*** Leaver
- ***A1*** Things left
- ***A2*** Benefactor
- ***AM-LOC*** Location

Basic Approach*: Local Classifiers make predictions (A0,A1,..) over words*

Are these decisions related? *"If A2 is present, A1 must also be present."*

How to express the constraints on the decisions? How to "enforce" them?

Slide taken from Dan Roth SRL slides

Semantic Role Labeling

Also known as – *"Who did what to whom, when, where, why,…"*

- Typical pipeline of decisions:
  - **Identify Argument candidates**
    - Train binary classifier

`[I] [left] [my pearls] [to] [[my] [daughter]] [in] [my will]`

  - **Classify Argument candidates**
    - Train multiclass classifier – Arg0, Arg1,..
  - **Global Inference**
    - Find global optimal subset of decisions (=candidates)
    - Use the scoring function defined by the classifier
    - Enforce linguistics constrains to ensure "legal" output

Global Inference

<u>Key issues:</u>

- How can you *translate* **domain knowledge into constraints**
- How can you **learn** to make predictions with constraints

other	0.05
per	**0.85**
loc	0.10

other	0.10
per	**0.60**
loc	0.30

other	0.05
per	0.50
loc	**0.45**

Dole 's wife, Elizabeth , is a native of N.C.

E_1 $\xrightarrow{R_{12}}$ E_2 $\xrightarrow{R_{23}}$ E_3

irrelevant	0.05
spouse_of	**0.45**
born_in	0.50

irrelevant	0.10
spouse_of	0.05
born_in	**0.85**

Structured Prediction

- Natural Language Decisions are Structured
  - Global decisions in which several local decisions play a role but there are mutual dependencies on their outcome.
- Essential to make coherent decisions
  - Consider decision interdependencies
  - Joint (or Global) Inference.
- Most interesting NLP problems require predicting multiple interdependent variables.
  - Not just “standard” classification
  - These are typically called **Structured Output Problems** – and will be the focus of this class.

Slide taken from Dan Roth SRL slides

A little bit about language..

Traditionally Computational Linguistics studies language at **different levels of analysis**

- **Morphology**
  - How words are constructed
- **Syntax**
  - Structural relation between words
- **Semantics**
  - The meaning of words and of combinations fo words
- **Pragmatics.**
  - How a sentence is used? What's its purpose
- **Discourse** (distinguished of subfield of Pragmatics)
  - Relationships between sentences; global context.

Morphology

- **Morphology**:
- How words are constructed; prefixes & Suffixes
- The simple cases are:

  kick, kicks, kicked, kicking
- But other cases may be

  sit, sits, sat, sitting
- Not as simple as adding /deleting certain endings:

  gorge, gorgeous

  good, goods

  arm, army
- This might be very different in other languages…

Syntax

- **Syntax**: Structural relationship between words.
- The main issues here are structural ambiguities, as in:

  I saw the Grand Canyon flying to New York.

- or

  Time flies like an arrow.

- The sentence can be interpreted as a
  - **Metaphor**: time passes quickly, but also
  - **Declaratively**: Insects have an affinity for arrows
  - **Imperative**: measure the time of the insects.

- **Key issue**: *syntax is not enough for representing meaning*.

Semantics

- **Semantics**: The meaning of words and of combinations of words. Some key issue here:
- Lexical ambiguities:

```
I walked to the bank  {of the river / to get money}.
The bug in the room {was planted by spies/ flew out}.
```

- Compositionality: The meaning of phrases/sentences as a function of the meaning of words in them

John kicked the bucket after drinking poison

Pragmatics/Discourse

- **Pragmatics**: How a sentence is used; **its purpose**.
- E.g.: Rules of conversation:
  - Can you tell me what time it is
  - Could I have the salt

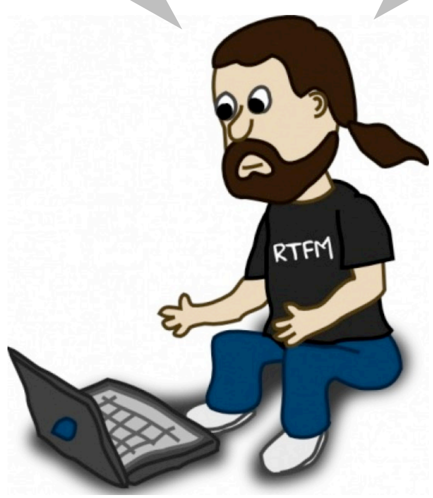
- **Discourse**: Relations between sentences; global context.
- An important example here is the problem of co-reference:

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

Question: which level of analysis?

Dude, I just watched this horror flick! Selling points: nightmares scenes, torture scenes, terrible monsters that was so bad a##!

Don’t buy the popcorn it was terrible, the monsters selling it must have wanted to torture me, it was so bad it gave me nightmares!



About this class

- **This is an advanced research class**
  - *You should play an active part!*
  - *Use this class to advance **your** research interests*
- **Expectations**
  - Attend lectures and actively participate
  - Complete assignments (3 programming assignments)
  - Present and critique papers
  - Final project
  - Final exam

Goals

At the end of this class you should:

- Understand NLP research papers
- Understand structured prediction papers
- Define and implement learning models for NLP tasks using ML methods
- **Conduct your own research and publish papers**

Topics

- ***Brief* NLP and Computational linguistics overview**
  - "100 things you wanted to know" Tutorial will be online. **Read it!**
- **Statistical NLP basics**
  - Language models, Binary and Multiclass classification
  - Useful things you can do with classification: sentiment analysis,..
  - Home assignment: **text classification**
- **Sequence Labeling problems**
  - HMM, CRF, Structured Perceptron and structured SVM
  - Useful things you can do with sequence labeling: NER, POS,..
  - Home assignment: **sequence labeling**

Topics

- **Complex structured prediction problem**
  - Inference: dynamic programming, ILP, approximate inference
  - **Useful things**: parsing, semantic interpretation, dialog analysis
- **Using less supervision**
  - Different training regimes, constrained driven learning, indirect supervision, learning with latent variables
  - **Useful things**: up to you!
  - Assignments: **project proposals**
- **Advanced topics**
  - **Up to you**: can be advanced applications, cool directions in ML
  - **Useful things**: technical aspects of your projects.
  - Assignment: **project presentations and submissions**

Other guidelines

- Working in groups
  - **Encouraged**, both for home assignments, final projects and paper presentations.
    - Groups should have 2-3 members
  - You can collaborate freely. **NO CHEATING**
- **Late policy**: 24 hours total for entire semester
  - **Start Early!**
- Drop by during office hours to discuss your papers and projects!
  - **Mandatory** if you are presenting
- Course website:

https://www.cs.purdue.edu/homes/dgoldwas/Teaching/ml4nlp_fall2017/
  - Sign up for the Piazza page for this class!

Part 1 : ML methods

- **Introduction to ML4NLP**
- **Classification**
  - **NLP Side**: sentiment classification, text categorization, ..
  - **ML side**: generative/discriminative classification, large margin classifiers, multiclass classification, Neural nets
- **Sequences**
  - **NLP Side:** POS tagging, chunking, NER, ..
  - **ML Side:** generative/discriminative tagging, large-margin extensions to sequences
- **General formulation for structured prediction**
  - **NLP side:** Information extraction, Semantic Role Labeling
  - **ML Side:** Training strategies (joint, independent, hybrid)

Part 2 : NLP Applications

- **Syntax**
  - Constituency Parsing
  - Dependency parsing
- **Semantics**
  - Lexical semantics
  - Sentence level
  - Grounded/logical semantics
- **Discourse**
  - Co-reference resolution
  - Discourse parsing

Part 3: Advanced ML/NLP

- Pretty open-ended but likely to include:
  - **Representation learning**
    - Traditional language models, neural language models, sentence embedding and beyond.
  - **Deep learning methods in NLP**
    - Neural nets, recurrent nets, recursive nets, CNN
    - Deep Structured Prediction
  - **Reinforcement Learning**

Grading

- **Final Exam**: 35%
- **Homework assignments**: 30%
  - Three programming assignments
- **Paper presentation:** 15%
  - Once in the semester
- **Final project**: 20%
  - Two deadlines: (1) proposal (2) submission+presentation

Assignments

- You will have to implement three assignments.
- **Key idea**: master the tools and algorithms used in NLP

- These will cover -
  - Machine learning warm up
  - Deep learning implementation
  - Advanced structured prediction

Paper Presentation + Review

- You will have to **Present** and **Review** a paper
  - Different papers!
- **Basic idea**: how can you go beyond the "vanilla" algorithms, deal with new problems and come up with new solutions.
- **Paper Presentations**
  - Short presentation, 10-15 minutes.
    - Papers will appear on the website, **choose**!
  - Motivate, provide context, explain the technical approach and evaluation
- **Paper Review**
  - Short review, 1 page
    - Same list of papers, SUBMISSION BEFORE THE PAPER IS PRESENTED.
  - Summarize the paper, identify strong/weak points

Final Project

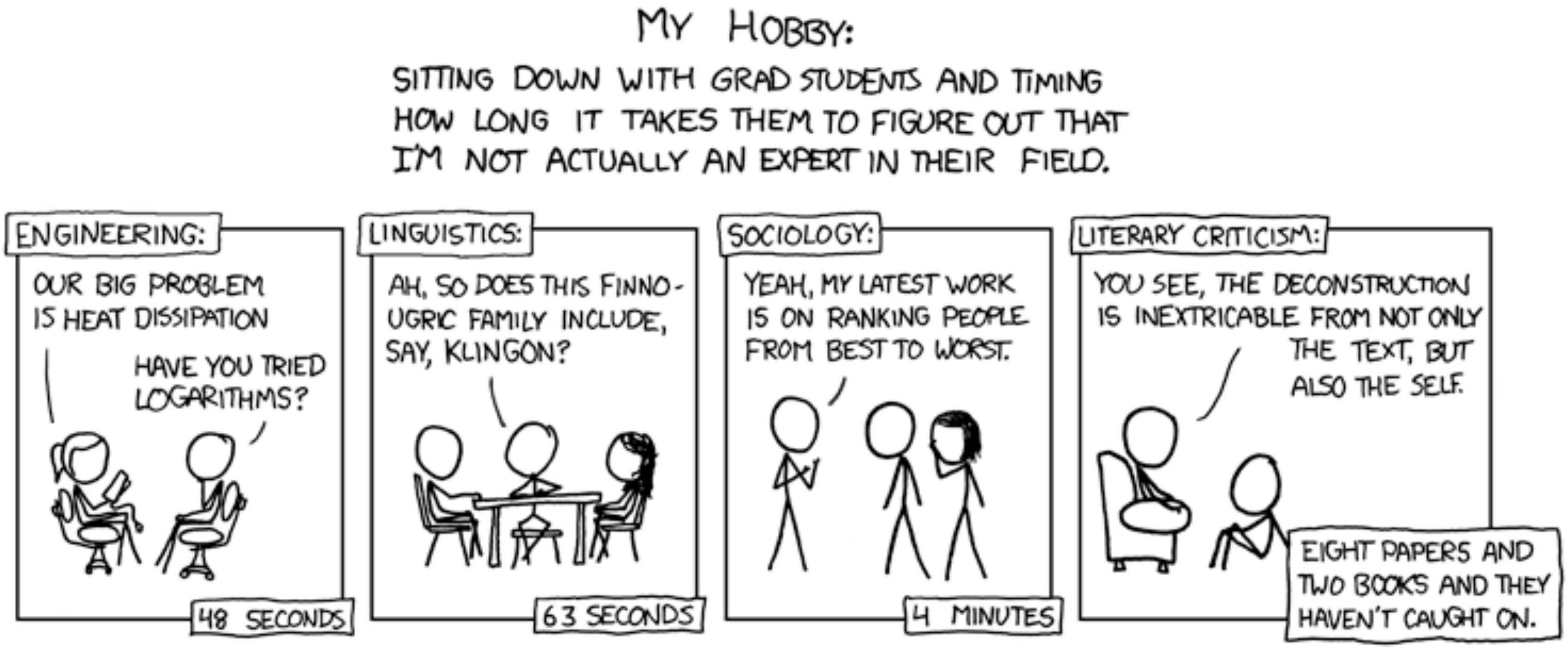
- You will have to submit a final project
  - Includes a project proposal, implementation and report.
- **Basic idea:** demonstrate your ability to come up with interesting solutions to new problems.
  - **Find a topic your care about!**
- Language is everywhere, you can be creative!
- You can also work on an existing problem, if you have a new approach to try
- **Be ambitious. Be reasonable.**
  - What **not** to do: *reduce to BoW classifier*.

Final Project

- **Proposal**:
- Define the problem
  - Related work
  - Basic intuitions and preliminary model
  - Datasets and experimental settings
  - **No more than 5 pages!**
- **Submission + Presentation:**
  - Short presentations in class
  - Short report describing your findings

Final Project Ideas

- Spelling correction
- Automatic essay grading
- Question answering system
- Language grounding (Combining vision and NLP)
- Legal text analysis
- Text generation (from events to headlines)
- Conversation analysis (who is winning a debate?)
- Metaphor and non-literal language


MY HOBBY:
SITTING DOWN WITH GRAD STUDENTS AND TIMING HOW LONG IT TAKES THEM TO FIGURE OUT THAT I'M NOT ACTUALLY AN EXPERT IN THEIR FIELD.
ENGINEERING:
OUR BIG PROBLEM IS HEAT DISSIPATION
HAVE YOU TRIED LOGARITHMS?
48 SECONDS
LINGUISTICS:
AH, SO DOES THIS FINNO-UGRIC FAMILY INCLUDE, SAY, KLINGON?
63 SECONDS
SOCIOLOGY:
YEAH, MY LATEST WORK IS ON RANKING PEOPLE FROM BEST TO WORST.
4 MINUTES
LITERARY CRITICISM:
YOU SEE, THE DECONSTRUCTION IS INEXTRICABLE FROM NOT ONLY THE TEXT, BUT ALSO THE SELF.
EIGHT PAPERS AND TWO BOOKS AND THEY HAVEN'T CAUGHT ON.

Questions?