Natural Language
from 20,000 feet

Some slides from Paula Matuszek, Mary-Angela Papalaskari, Dan Weld, Jim Martin, Ralph Grishman, Yejin Choi, Jurafsky and Martin, Noriko Tomuro, Joyce Chai, Frank Ferraro

1

Bookkeeping

Final exam is Tuesday in class

- Project final paper due 12/10 at 11:59 PM
- Today:
  - Some NLP
  - Some exam review

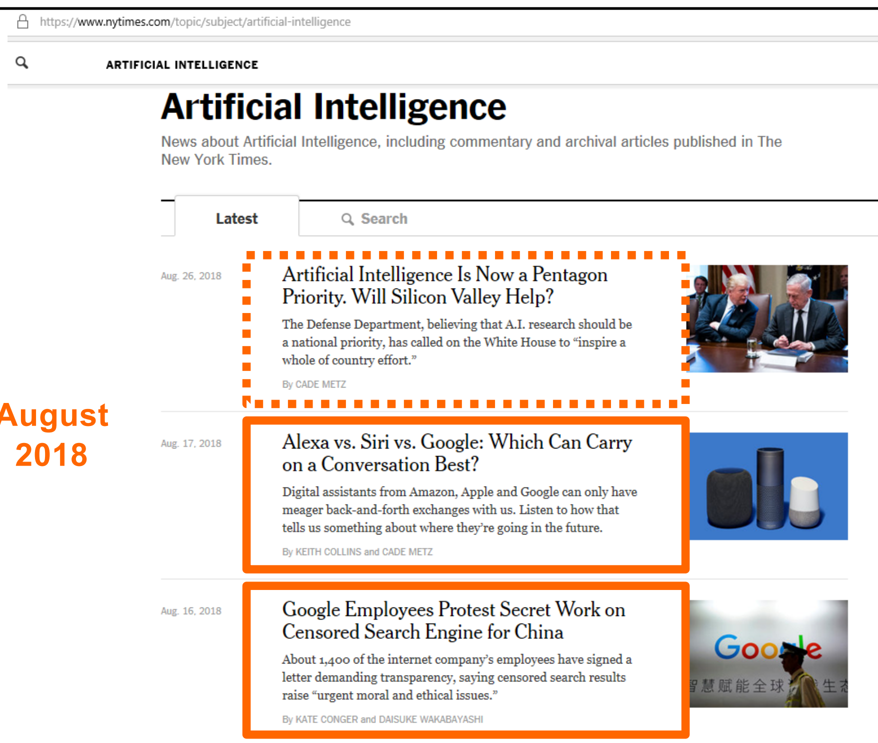
2

Speech and Language Processing

- Getting computers to do reasonably intelligent things with human language is the domain of Computational Linguistics (or Natural Language Processing or Human Language Technology)

3

August 2018


https://www.nytimes.com/topic/subject/artificial-intelligence

ARTIFICIAL INTELLIGENCE

Artificial Intelligence

News about Artificial Intelligence, including commentary and archival articles published in The New York Times.

Latest | Search

Aug. 26, 2018

Artificial Intelligence Is Now a Pentagon Priority. Will Silicon Valley Help?

The Defense Department, believing that A.I. research should be a national priority, has called on the White House to "inspire a whole of country effort."

By CADE METZ

Aug. 17, 2018

Alexa vs. Siri vs. Google: Which Can Carry on a Conversation Best?

Digital assistants from Amazon, Apple and Google can only have meager back-and-forth exchanges with us. Listen to how that tells us something about where they're going in the future.

By KEITH COLLINS and CADE METZ

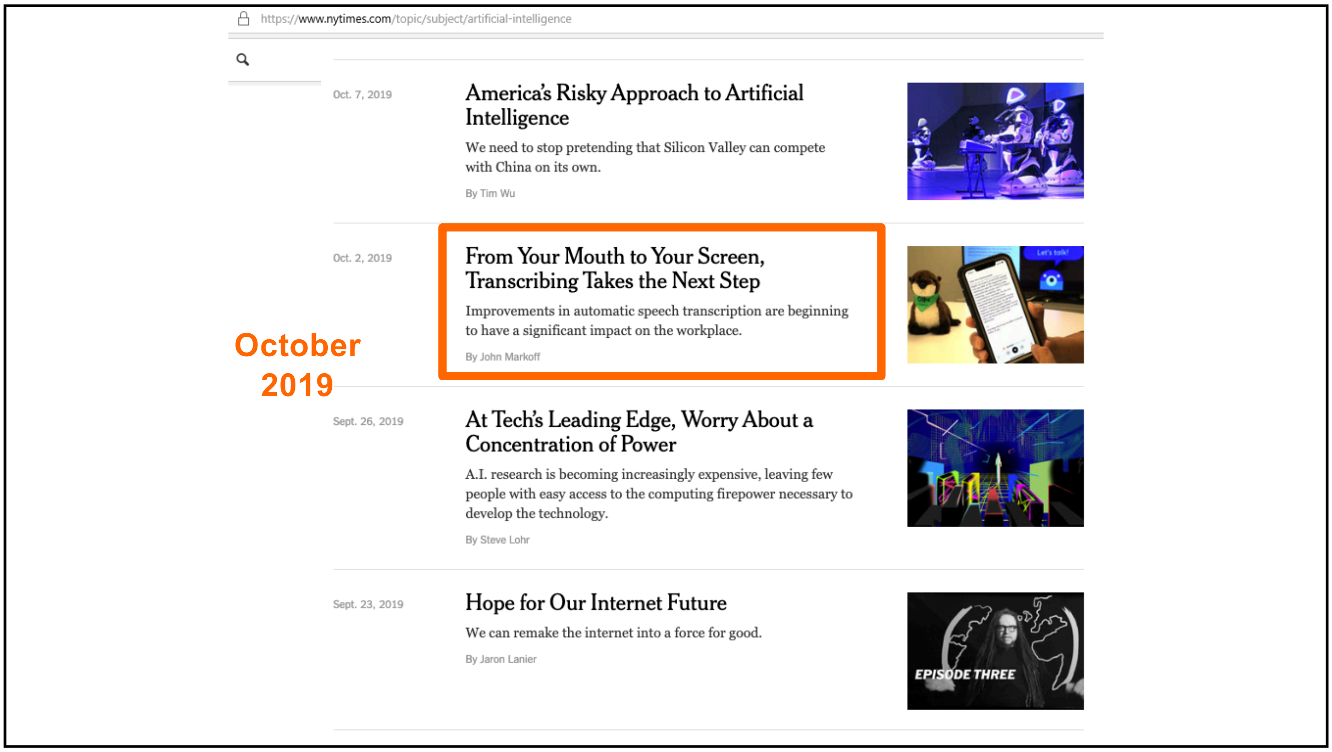
Aug. 16, 2018

Google Employees Protest Secret Work on Censored Search Engine for China

About 1,400 of the internet company's employees have signed a letter demanding transparency, saying censored search results raise "urgent moral and ethical issues."

By KATE CONGER and DAISUKE WAKABAYASHI

4



https://www.nytimes.com/topic/subject/artificial-intelligence

Oct. 7, 2019

America's Risky Approach to Artificial Intelligence

We need to stop pretending that Silicon Valley can compete with China on its own.

By Tim Wu

Oct. 2, 2019

From Your Mouth to Your Screen, Transcribing Takes the Next Step

Improvements in automatic speech transcription are beginning to have a significant impact on the workplace.

By John Markoff

October 2019

Sept. 26, 2019

At Tech's Leading Edge, Worry About a Concentration of Power

A.I. research is becoming increasingly expensive, leaving few people with easy access to the computing firepower necessary to develop the technology.

By Steve Lohr

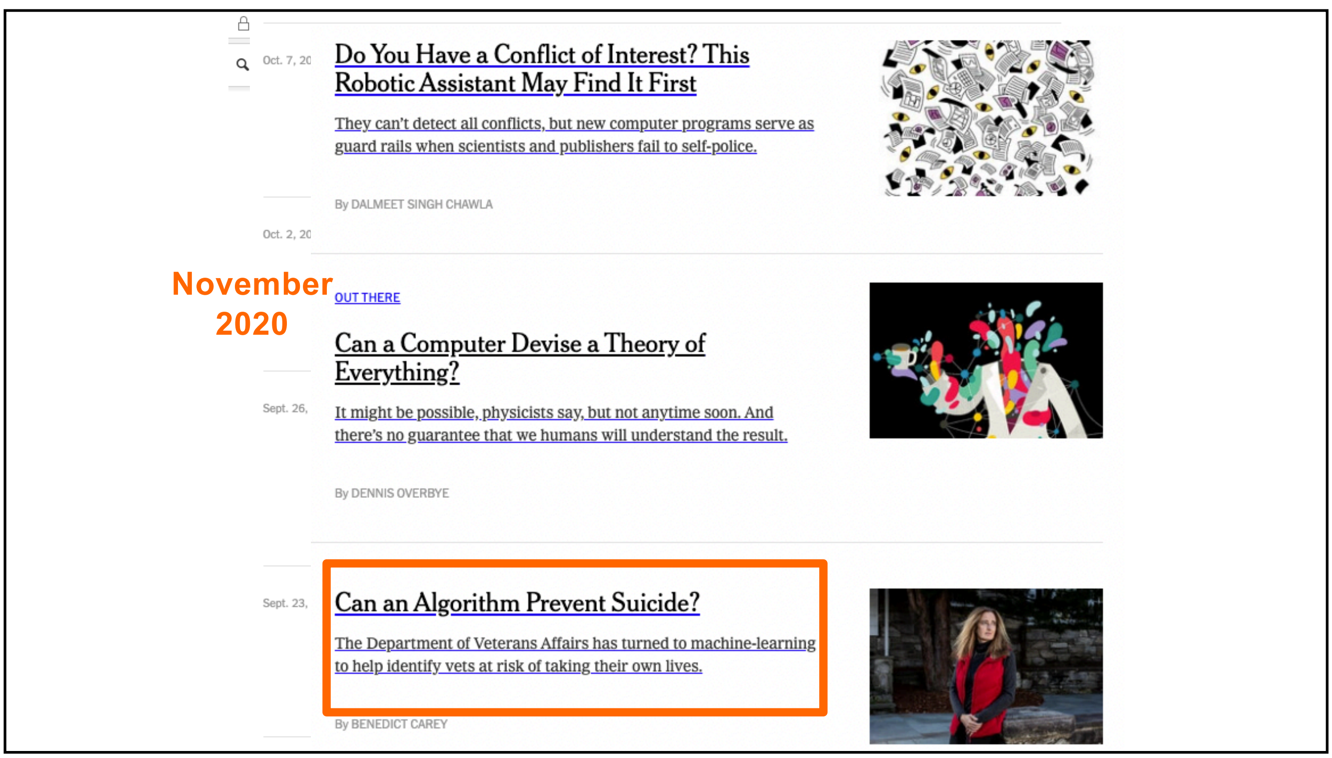
Sept. 23, 2019

Hope for Our Internet Future

We can remake the internet into a force for good.

By Jaron Lanier

5



Oct. 7, 20

Do You Have a Conflict of Interest? This Robotic Assistant May Find It First

They can't detect all conflicts, but new computer programs serve as guard rails when scientists and publishers fail to self-police.

By DALMEET SINGH CHAWLA

Oct. 2, 20

November 2020

OUT THERE

Can a Computer Devise a Theory of Everything?

Sept. 26,

It might be possible, physicists say, but not anytime soon. And there's no guarantee that we humans will understand the result.

By DENNIS OVERBYE

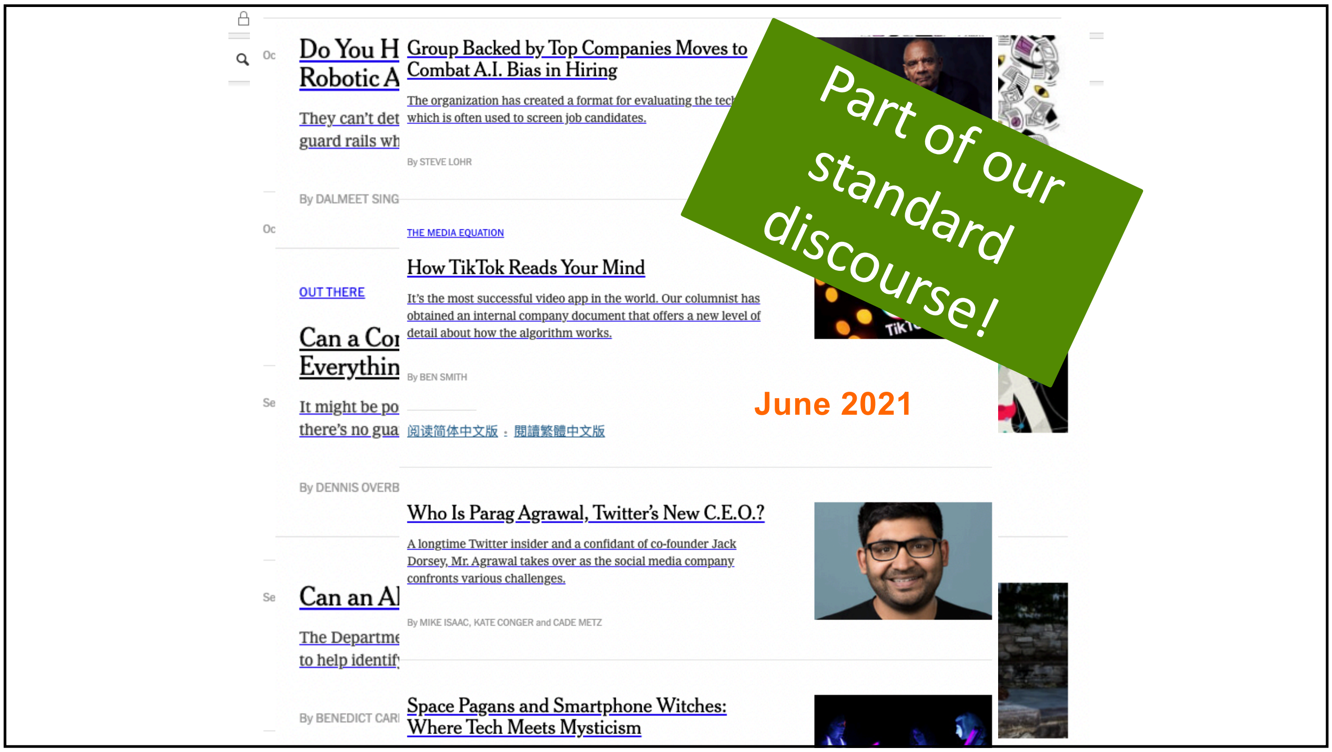
Sept. 23,

Can an Algorithm Prevent Suicide?

The Department of Veterans Affairs has turned to machine-learning to help identify vets at risk of taking their own lives.

By BENEDICT CAREY

6



7

Applications

- Applications of NLP can be broken down into Small and Big
- Small applications include many things you never think about:
  - Hyphenation
  - Spelling correction
  - OCR
  - Grammar checkers
- Big applications include:
  - Machine translation
  - Question answering
  - Conversational speech recognition

8

Applications

- There’s another kind: Medium
  - Speech recognition in closed domains
  - Question answering in closed domains
  - Question answering for factoids
  - Information extraction from news-like text
  - Generation and synthesis in closed/small domains.

9

NLP Research

- In between the linguistics and the big applications are a host of hard problems.
  - Robust Parsing
  - Word Sense Disambiguation
  - Semantic Analysis
  - Etc
- Not too surprisingly, solving these problems involves:
  - Choosing the right logical representations
  - Managing hard search problems
  - Dealing with uncertainty
  - Using machine learning to train systems to do what we need

10

Why Study NLP?

- A hallmark of human intelligence.
- To interact with computing devices using human (natural) languages
  - Building intelligent robots (AI)
  - Enabling voice-controlled operation
- To access (large amount of) information and knowledge stored in the form of human languages quickly
  - Text is the largest repository of human knowledge and is growing quickly.
    - Emails, news articles, web pages, IM, scientific articles, insurance claims, customer complaint letters, transcripts of phone calls, technical documents, government documents, patent portfolios, court decisions, contracts, …

12

NL and NLP

- “Natural” languages = human languages
  - English, Russian, Wolof, …
- Natural Language Processing: any form of dealing with NL computationally
- Many, many sub-areas; important from an AI perspective, 2 are most crucial:
  - **Natural Language Understanding**: understanding the meaning (semantics) of spoken or written text
  - **Natural Language Generation:** Producing meaningful, relevant language

13

Layers of Language

- **Phonology**: The noises you make and understand
- **Morphology**: What you know about the structure of the words in your language, including their derivational and inflectional behavior
- **Syntax**: What you know about the order and constituency of the utterances you make
- **Semantics**: What does in all mean?
  - What is the connection between language and the world?
- **Discourse**: Dealing with larger chunks of language; dealing with language in context

15

Fundamental NLP Tasks

- **Speech recognition:** speech signal to words
- **Parsing:** decompose sentence into units
  - Part of speech tagging: Eg, noun vs. verb
- **Semantic role labeling:** for a verb, find the units that fill pre-defined "semantic roles" (eg, Agent, Patient or Instrument)
  - Example: "John hit the ball with the bat"

17

Fundamental NLP Tasks

- Semantics: map sentence to corresponding “meaning representation” (e.g., logic)
  - give(John, Book, Sally)
  - Quantification: Everybody loves somebody
- Word Sense Disambiguation
  - orange juice vs. orange coat
  - Where do word senses come from?
- Co-reference resolution:
  - The dog found a cookie; He ate it.
- Implicit “text” - what is left unsaid?
  - Joe entered the restaurant, ordered, ate and left. The owner said “Come back soon!”
  - Joe entered the restaurant, ordered, ate and left. The owner said “Hey, I’ll call the police!”

18

Applied NLP

- Machine translation
- Spelling/grammar correction
- Information Retrieval/extraction
- Data mining
- Document classification
- Question answering
- Conversational agents

19

You See It Daily

- Question answering: Siri, OK Google, Cortana, Alexa
- spelling/grammar correction
- Automated response systems
- To get input for
  - Information Retrieval
  - Data mining
  - Document classification
- Machine translation

20

Human Languages

- You know ~50,000 words of primary language, each with several meanings
- Six year old knows ~13000 words
- First 16 years we learn 1 word every 90 min of waking time
- Mental grammar generates sentences
  - virtually every sentence is novel!
- 3 year olds already have 90% of grammar
- ~6000 human languages – none of them simple!

Adapted from Martin Nowak 2000 – Evolutionary biology of language – Phil.Trans. Royal Society London

21

Human spoken language

- Most complicated mechanical motion of the body
  - Movements must be accurate to within half mm
  - synchronized within hundredths of a second
- We can understand up to 50 phonemes/sec (normal speech 10-15ph/sec)
  - but if sound is repeated 20 times /sec we hear continuous buzz!
- All aspects of language processing are involved and manage to keep apace

Adapted from Martin Nowak 2000 – Evolutionary biology of language – Phil.Trans. Royal Society London

22

Main Challenge in NLP: Ambiguity

- Lexical:
  - Label (noun or verb)?
  - London (Jack or capital of the UK)?
- Syntactic (examples from newspaper headlines):
  - Prepositional Phrase Attachment: Ban on Nude Dancing on Governor's Desk
  - Word Sense Disambiguation: Iraqi Head Seeking Arms
  - Syntactic Ambiguity (what's modifying what): Juvenile Court to Try Shooting Defendant
- Semantic ambiguity:
  - "snake poison"
- Rampant metaphors:
  - "prices went up"

23

Some Ambiguous Headlines

- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Kids Make Nutritious Snacks
- Bush Wins on Budget, but More Lies Ahead
- Hospitals are Sued by 7 Foot Doctors

Source: Marti Hearst, i256, at UC Berkeley

24

Dealing with Ambiguity

- Four possible approaches:
  1. Formal approaches -- Tightly coupled interaction among processing levels; knowledge from other levels can help decide among choices at ambiguous levels.
  2. Pipeline processing that ignores ambiguity as it occurs and hopes that other levels can eliminate incorrect structures.
  3. **Probabilistic approaches based on making the most likely choices**
  4. Don’t do anything, maybe it won’t matter

Source: Jurafsky & Martin “Speech and Language Processing”

25

What Are Words?

bat

http://www.freepngimg.com/download/bat/9-2-bat-png-hd.png

26

What Are Words?

bats

http://www.freepngimg.com/download/bat/9-2-bat-png-hd.png

27

What Are Words?

Fledermaus

flutter mouse

http://www.freepngimg.com/download/bat/9-2-bat-png-hd.png

28

What Are Words?

pişirdiler

They cooked it.

29

What Are Words?

pişmişlermişlerdi

They had it cooked it.

30

What Are Words?

):

31

What Are Words?

my leg is hurting nasty):

32

What Are Words?

add two cups (a pint): bring to a boil

http://www.dummies.com/wp-content/uploads/88513.image4.jpg

33

What Are Words?

- Hard to get agreement
- (Human) Language-dependent
- White-space separation is a sometimes okay (for written English longform)
- Social media? Spoken vs. written? Other languages?

34

Why Language is Hard

- NLP is *AI-complete*
- Abstract concepts are difficult to represent
- LOTS of possible relationships among concepts
- Many ways to represent similar concepts
- Tens of hundreds or thousands of features/dimensions

35

Why Language is Easy

- Highly redundant
- Relatively crude methods provide fairly good results
- Lots of subject matter experts!

36

Some of the Tools

- A mixed bag, at various levels...
  - Tokenizers
  - Regular Expressions and Finite State Automata
  - Part of Speech taggers
  - Grammars
  - Parsers
  - N-Grams
  - Semantic Analysis

37

What will it take?

- Models of computation (state machines)
- Formal grammars
- Knowledge representation
- Search algorithms
- Dynamic programming
- Logic
- Machine learning
- Probability theory

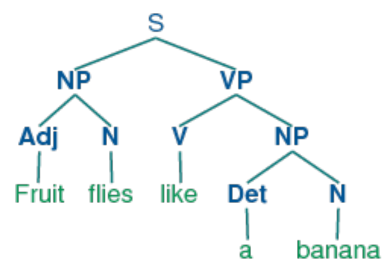
38

A Few Key Problems and Tools

39

Parts of Speech Tagging

- Part-of-Speech (POS) taggers identify nouns, verbs, adjectives, noun phrases, etc.
- More recent work uses machine learning to create taggers from labeled examples



40

Named Entities (NE) Tagging

- Persons, places, companies
  - "Proper nouns"
  - One of most common information extraction tasks
  - Combination of rules and dictionary
- Example rules:
  - Capitalized word not at beginning of sentence
  - Two capitalized words in a row
  - One or more capitalized words followed by Inc
  - Dictionaries of common names, places, major corporations.
    - Sometimes called "gazetteer"

41

Reference Resolution

- Discourse Knowledge — what have we just said?
  - Paula is here. She is ready.
- Domain/World Knowledge
  - U: I would like to register in a CMSC Course.
  - S: Which number?
  - U: 647.
  - S: Which section?
  - U: Which section is in the evening?
  - S: section 1.
  - U: Then that one.

42

Word Sense Resolution

- Many words have several meanings or **senses**
- We need to resolve which of the senses of an ambiguous word is invoked in a particular use of the word
- I made her duck. (meanings?)

43

Word Sense Resolution

- Many words have several meanings or **senses**
- We need to resolve which of the senses of an ambiguous word is invoked in a particular use of the word
- I made her duck. (meanings?)
  1. I cooked waterfowl for her benefit (to eat)
  2. I cooked waterfowl belonging to her
  3. I created the (plaster?) duck she owns
  4. I caused her to quickly lower her head or body
  5. I waved my magic wand and turned her into undifferentiated waterfowl
- Again, discourse and world knowledge

Duck example Jurafsky & Martin "Speech and Language Processing"

44

Phonological Ambiguity

- I mate or duck
- I'm eight or duck
- Eye maid; her duck
- Aye mate, her duck
- I maid her duck
- I'm aid her duck
- I mate her duck
- I'm ate her duck
- I'm ate or duck
- I mate or duck

Sound like
"I made her duck"

Duck example Jurafsky & Martin "Speech and Language Processing"

45

Semantics

- What kinds of things can we not do well with the tools we have already looked at?
  - Retrieve information in response to unconstrained questions: e.g., travel planning
  - Accurate translations?
  - Play the "chooser" side of 20 Questions
  - Read a newspaper article and answer questions about it
- These tasks require that we also consider **semantics**: the meaning of our tokens and their sequences

46

Semantics

- How do you map a sentence to a semantic representation?
  - What are the semantic primitives?
- Schank: 12 (or so) primitives
- The basis of natural language is conceptual.
- The conceptual level is interlingual, while the sentential level is language-specific.
  - De-emphasize syntax
  - Focus on semantic expectations as the basis for NLU

47

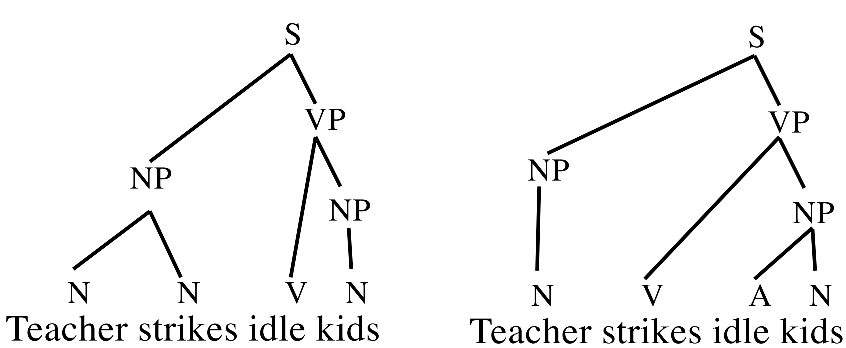
Syntax is about Sentence Structures

- Sentences have structures and are made up of constituents.
  - The constituents are phrases.
  - A phrase consists of a head and modifiers.
- The category of the head determines the category of the phrase
  - e.g., a phrase headed by a noun is a noun phrase

48

Alternative Parses

Teacher strikes idle kids



49

Issues with Statistical Parsing

- Statistical parsers still make plenty of errors
- Tree banks are language specific
- Tree banks are genre specific
  - Train on WSJ ➔ fail on the Web
  - standard distributional assumption
- Unsupervised, un-lexicalized parsers exist
  - But performance is substantially weaker

50

Big Applications

- POS tagging, parsing and word sense disambiguation are all medium-sized enabling applications.
  - They don’t actually do anything that anyone actually cares about.
  - MT and QA are things people seem to care about.

51

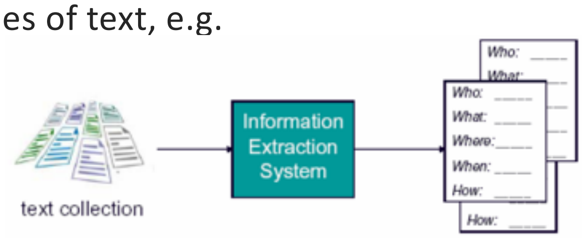
How Difficult is Morphology?

- Examples from Woods et. al. 2000
  - delegate (de + leg + ate) take the legs from
  - caress (car + ess) female car
  - cashier (cashy + er) more wealthy
  - lacerate (lace + rate) speed of tatting
  - ratify (rat + ify) infest with rodents
  - infantry (infant + ry) childish behavior

52

Information Extraction (IE)

- Identify specific pieces of information (data) in an unstructured or semi-structured text
- Transform unstructured information in a corpus of texts or web pages into a structured database (or templates)
- Applied to various types of text, e.g.
  - Newspaper articles
  - Scientific articles
  - Web pages
  - etc.



Source: J. Choi, CSE842, MSU

53

But "Reading" the Web is Tough

- Traditional IE is narrow
- IE has been applied to small, homogenous corpora
- No parser achieves high accuracy
- No named-entity taggers
- No supervised learning

- How about semi-supervised learning?

55

What makes IE from the Web Different?

Less grammar, but more formatting & linking

Newswire

Apple to Open Its First Retail Store
in New York City

MACWORLD EXPO, NEW YORK--July 17, 2002--
Apple's first retail store in New York City will open in Manhattan's SoHo district on Thursday, July 18 at 8:00 a.m. EDT. The SoHo store will be Apple's largest retail store to date and is a stunning example of Apple's commitment to offering customers the world's best computer shopping experience.

"Fourteen months after opening our first retail store, our 31 stores are attracting over 100,000 visitors each week," said Steve Jobs, Apple's CEO. "We hope our SoHo store will surprise and delight both Mac and PC users who want to see everything the Mac can do to enhance their digital lifestyles."

The directory structure, link structure, formatting & layout of the Web is its own new grammar.

Web

Slides from Cohen & McCallum

56

Machine Translation

- The automatic translation of texts between languages is one of the oldest non-numerical applications in Computer Science.
- In the past 15 years or so, MT has gone from a niche academic curiosity to a robust commercial industry.

57

Google Translate

Document classification

Machine translation

58

Sentiment Analysis

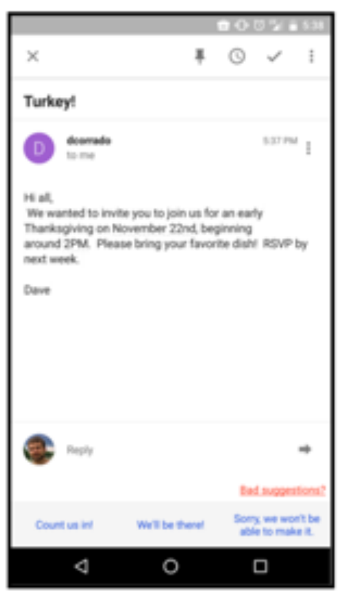
- The field of sentiment analysis deals with categorization (or classification) of opinions expressed in textual documents

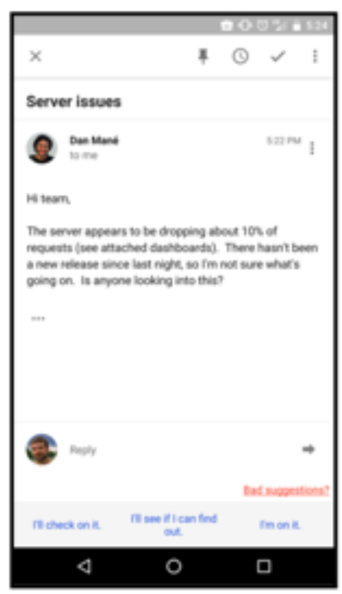
> The TV is wonderful. Great size, great picture, easy interface. It makes a cute little song when you boot it up and when you shut it off. I just want to point out that the 43" does not in fact play videos from the USB .This is really annoying because that was one of the major perks I wanted from a new TV. Looking at the product description now, I realize that the feature list applies to the X758 series as a whole, and that each model's capabilities are listed below. Kind of a dumb oversight on my part, but it's equally stupid to put a description that does not apply on the listing for a very specific model.

Green color represents positive tone, red color represents negative tone, and product features and model names are highlighted in blue and brown, respectively.

59

Natural Language Generation





https://cdn.arstechnica.net/wp-content/uploads/2015/11/Screen-Shot-2015-11-02-at-9.11.40-PM-640x543.png

60

Document Labeling/Classification

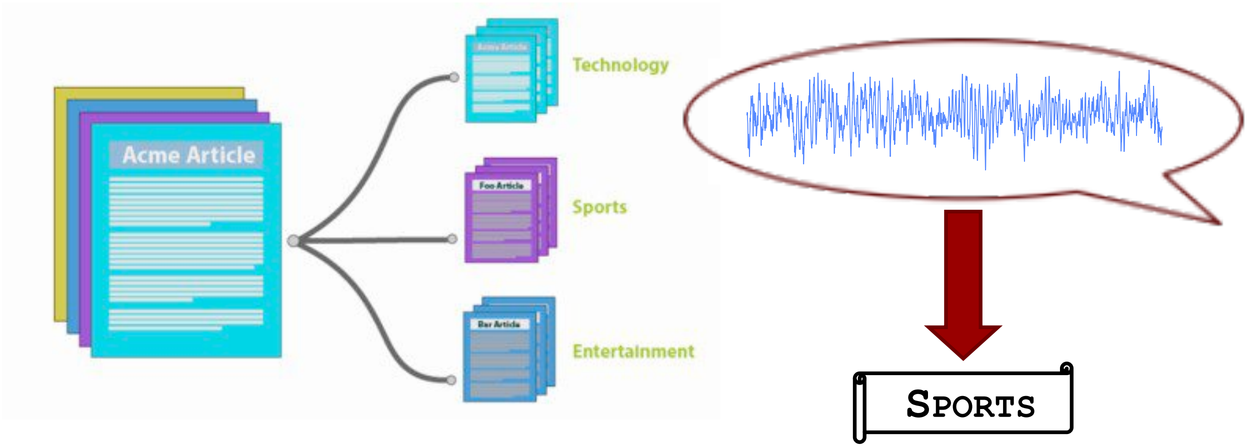


Image: https://www.kdnuggets.com/2015/01/text-analysis-101-document-classification.html

61

Conclusions

- NLP is harder than it might seem naively
- Many subtasks
- Statistical NLP is the dominant paradigm
  - supervised learning
  - corpus-based statistics (language models)
  - Some important limitations in practice!
- NL “understanding” has received very little attention

62

Our NLP/NLU Class: Course Goals

- Be introduced to some of the core problems and solutions of NLP (big picture)
- Learn different ways that success and progress can be measured in NLP
- Relate to statistics, machine learning, and linguistics
- Implement NLP programs

63

Our NLP/NLU Class: Course Topics

- Probability, classification, and the efficacy of simple counting methods
- Language modeling (n-gram models, smoothing heuristics, maxent/log-linear models, and distributed/vector-valued representations)
- Sequences of latent variables (e.g., hidden Markov models, some basic machine translation alignment)
- Trees and graphs, as applied to syntax and semantics
- Some discourse-related applications (coreference resolution, textual entailment)
- Special and current topics (e.g., fairness and ethics in NLP)
- Modern, neural approaches to NLP, such as recurrent neural networks and transformers (e.g., BERT or GPT-2)

64

Exam Review

65

Exam Topics

- Knowledge
  - Knowledge-Based Agents
  - Knowledge Representation
  - First-Order Logic
  - Inference
- Planning
  - State spaces
  - PO Planning
  - Probabilistic Planning
- Machine Learning
  - Decision Trees
  - Classification
  - Reinforcement Learning
  - Clustering
- Applications
  - Robotics
  - Natural Language

66

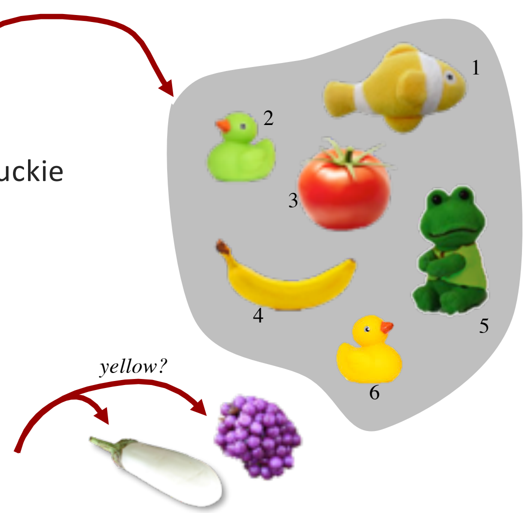
Machine Learning

- Optimize a performance criterion using example data or past experience
- Many varieties…
  - Classification
  - Regression
  - Unsupervised learning
  - Reinforcement learning
- **The Big Idea:** given some data, you learn a model of how the world works that lets you predict new data

67

Machine Learning Terminology

- What we have:
- **Data:** examples of our problem
  - Processed to produce **features**
    - Can't give a computer a rubber duckie
  - Turned into a feature **vector**
  - Sometimes labeled, sometimes not
- What we want:
- A **prediction** over new data



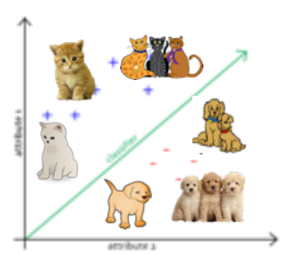
68

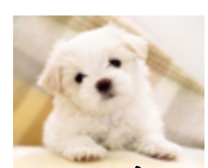
Machine Learning

- Supervised vs. Unsupervised
  - What is classification?
  - What is clustering?
  - Exploitation v. Exploration
  - K-Means, EM, and failure modes

Classification

- Classification or concept learning (aka “induction”)
  - Given a set of examples of some concept/class/category:
  - Determine if a given example is an instance of the concept (class member) or not
  - If it is: **positive example**
  - If it is not: **negative example**
  - Or we can make a probabilistic prediction (e.g., using a Bayes net)

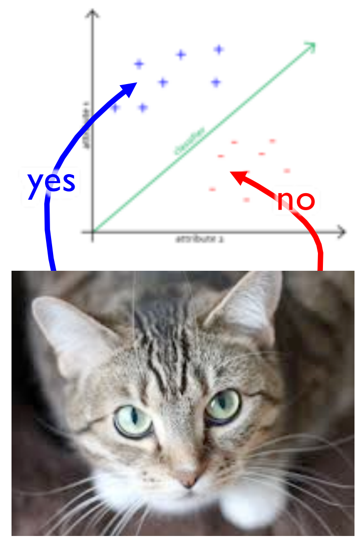




cat?

More on the Classification Problem

- Extrapolate from **examples** to make accurate **predictions** about future data points
  - Examples are called **training data**
- Predict into **classes**, based on attributes (“**features**”)
  - Example: it has tomato sauce, cheese, and no bread. Is it pizza?
  - Example: does this image contain a cat?



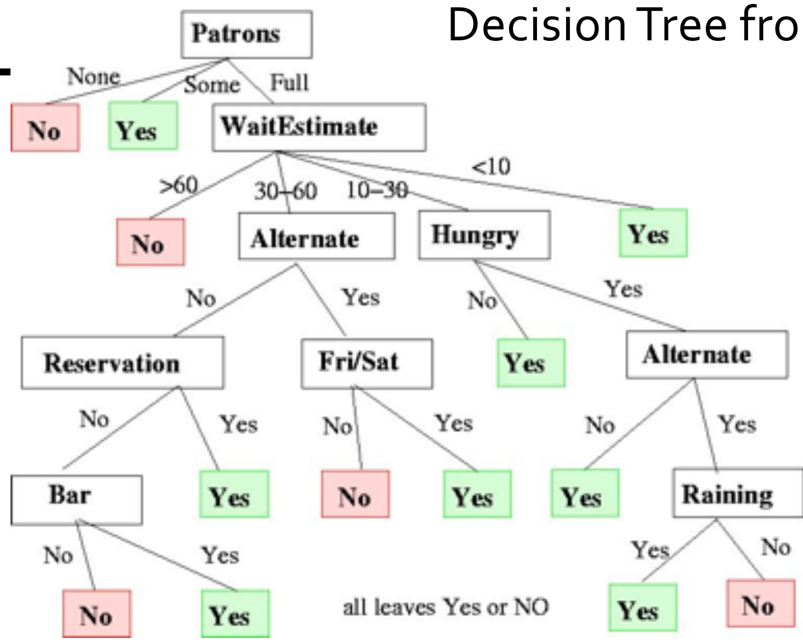
71

Decision Trees

- Goal: Build a tree to classify examples as positive or negative instances of a concept using supervised learning from a training set
- A decision tree is a tree where:
  - Each **non-leaf** node is an attribute (feature)
  - Each **leaf** node is a classification (+ or -)
    - Positive and negative data points
  - Each **arc** is one possible value of the attribute at the node from which the arc is directed
- Generalization: allow for >2 classes
  - e.g., {sell, hold, buy}

72

Decision Tree from Inspection

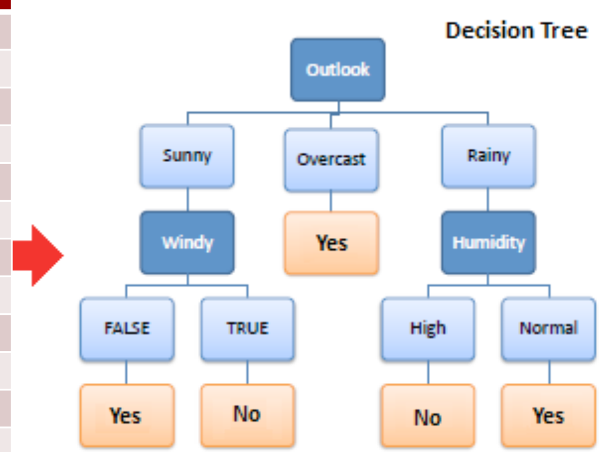


Problem from R&N, table from Dr. Manfred Kerber @ Birmingham, with thanks – www.cs.bham.ac.uk/~mmk/Teaching/AI/l3.html

73

Exercise: draw a decision tree

Outlook	Temp	Humidity	Windy	Play golf?
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



www.saedsayad.com/decision_tree.htm

74

Choosing the Attribute to Split On

- **Information gain**: how much entropy decreases (homogeneity increases) when a dataset is split on an attribute.
  - High homogeneity → high likelihood samples will have the same class
- Constructing a decision tree is all about finding attribute that returns the highest information gain (i.e., the most homogeneous branches)

75

Knowledge Representation

- Ontologies
  - What would an ontology of "living things" look like?
    - Graphically? As a formal representation?
- Semantic Nets
  - Give an eight-node, nine-arc network about food
    - Graphically? As a formal representation?
- Types of relationships
  - Predicates: return true or false (a truth value)
  - Functions: return a value
  - Common types: is-a, part-of, kind-of, member-of
  - Keep individuals (e.g., Einstein) and groups (e.g., scientists) straight

76

Representation, Reasoning, and Logic

- Point of knowledge representation is to express knowledge in a **computer usable** form
  - Needed for agents to act on it!
- **Logics** are formal languages for representing information such that conclusions can be drawn
- **Syntax** defines how symbols can be put together to form the sentences in the language
- **Semantics** define the "meaning" of sentences;
  - i.e., define truth of a sentence in a world (given an interpretation)
- Knowledge is stored in a Knowledge Base, or KB

77

YOUR MISSION

- Prove that the Wumpus is in (1,3) and there is a pit in (3,1), given the observations shown and these rules:

Rules

- If there is no stench in a cell, then there is no wumpus in any adjacent cell
- If there is a stench in a cell, then there is a wumpus in some adjacent cell
- If there is no breeze in a cell, then there is no pit in any adjacent cell
- If there is a breeze in a cell, then there is a pit in some adjacent cell
- If a cell has been visited, it has neither a wumpus nor a pit
  - **FIRST** write the propositional rules for the relevant cells
  - **NEXT** write the proof steps and indicate what inference rules you used in each step

PL Proofs

A = *Agent*
B = *Breeze*
G = *Glitter, Gold*
OK = *Safe square*
P = *Pit*
S = *Stench*
V = *Visited*
W = *Wumpus*

V12 S12 -B12	V22 -S22 -B22		
V11 -S11 -B11	V21 B21 -S21		

INFERENCE RULES

Modus Ponens
$A, A \rightarrow B$
ergo B

And Introduction
A, B
ergo $A \wedge B$

And Elimination
$A \wedge B$
ergo A

Double Negation
$\neg\neg A$
ergo A

Unit Resolution
$A \vee B, \neg B$
ergo A

Resolution
$A \vee B, \neg B \vee C$
ergo $A \vee C$

78

First-Order Logic

- First-order logic (FOL) models the world in terms of
  - **Objects**, which are things with individual identities
  - **Properties** of objects that distinguish them from other objects
  - **Relations** that hold among sets of objects
  - **Functions**, which are a subset of relations where there is only one "value" for any given "input"
- Examples:
  - Objects: students, lectures, companies, cars ...
  - Relations: brother-of, bigger-than, outside, part-of, has-color, occurs-after, owns, visits, precedes, ...
  - Properties: blue, oval, even, large, ...
  - Functions: father-of, best-friend, second-half, one-more-than ...

79

Translating English to FOL

- **Every gardener likes the sun.**
  - $\forall x\ gardener(x) \Rightarrow likes(x,Sun)$
- **You can fool some of the people all of the time.**
  - $\exists x\ \forall t\ \ person(x) \wedge time(t) \Rightarrow can\text{-}fool(x,t)$
- **You can fool all of the people some of the time.**
  - $\forall x\ \exists t\ (person(x) \Rightarrow time(t) \wedge can\text{-}fool(x,t))$
  - $\forall x\ (person(x) \Rightarrow \exists t\ (time(t) \wedge can\text{-}fool(x,t))$

  Equivalent
- **All purple mushrooms are poisonous.**
  - $\forall x\ (mushroom(x) \wedge purple(x)) \Rightarrow poisonous(x)$

80

Translating English to FOL

- **No purple mushroom is poisonous.**
  - $\neg\exists x$ purple(x) $\wedge$ mushroom(x) $\wedge$ poisonous(x)
  - $\forall x$ (mushroom(x) $\wedge$ purple(x)) $\Rightarrow$ $\neg$poisonous(x)

  Equivalent
- **There are exactly two purple mushrooms.**
  - $\exists x$ $\exists y$ mushroom(x) $\wedge$ purple(x) $\wedge$ mushroom(y) $\wedge$ purple(y) ^ $\neg$(x=y) $\wedge$ $\forall z$ (mushroom(z) $\wedge$ purple(z)) $\Rightarrow$ ((x=z) $\vee$ (y=z))
- **Mary is not tall.**
  - $\neg$tall(Mary)
- **X is above Y iff X is on directly on top of Y or there is a pile of one or more other objects directly on top of one another starting with X and ending with Y.**
  - $\forall x$ $\forall y$ above(x,y) $\leftrightarrow$ (on(x,y) $\vee$ $\exists z$ (on(x,z) $\wedge$ above(z,y)))

81

Translating FOL to English

1. $\forall x$ (bitter(x) $\vee$ sweet(x))
2. $\forall x$ (bitter(x)) $\vee$ $\forall x$ (sweet(x))
3. $\exists x$ $\forall y$ (loves(y,x))
4. $\neg\exists x$ $\neg\exists y$ (loves(y,x))
5. $\exists x$ (noisy(x)) $\Rightarrow$ $\forall y$(annoyed(y))
6. $\forall x$ (frog(x) $\Rightarrow$ green(x))
7. $\forall x$ (frog(x) $\Rightarrow$ ¬green(x))
8. $\neg\exists x$ (frog(x) $\wedge$ green(x))
9. $\exists x$ (frog(x) $\wedge$ $\neg$green(x))
10. $\exists x$ (mech.(x) $\wedge$ likes(x, Bob))
11. $\exists x$ (mech.(x) $\wedge$ likes(x, x))
12. $\forall x$ (mech.(x) $\Rightarrow$ likes(x, Bob))
13. $\exists x$ $\forall y$ (mech(x) $\wedge$ nurse(y) $\Rightarrow$ likes(x, y))
14. $\exists x$(mech(x) $\wedge$ $\forall y$(nurse(y) $\Rightarrow$ likes(y, x))

Exercises: disi.unitn.it/~bernardi/Courses/LSNL/Slides/fl1.pdf

82

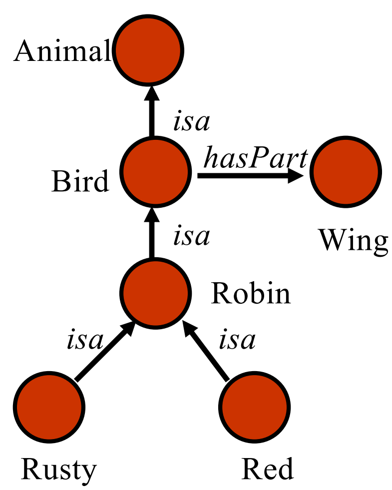
Proving Things: 5 Methods

- Inference by Enumeration
  - List all possible true worlds, check the truth value of a sentence
  - Complete but exponential in time
- Proof by Natural Deduction
  - Writing proofs from laws (e.g., modus ponens)
- Forward Chaining
- Backward Chaining
- Resolution Refutation
  - Show KB ⊨ α by proving that KB ∧ ¬α is unsatisifiable, i.e., deducing False from KB ∧ ¬α

83

Semantic Networks

- The ISA (is-a) or AKO (a-kind-of) relation is often used to link instances to classes, classes to superclasses
- Some links (e.g. hasPart) are inherited along ISA paths.
- The semantics of a semantic net can be informal or very formal
  - often defined at the implementation level



84

Reasoning and Inference

- Given a formally represented world
  - Agents and their behaviors
  - Goals
  - State spaces
- What is **inference**?
- What kinds of inference can you do?
  - Forward Chaining
  - Backward Chaining

86

Forward Chaining

Knowledge Base
1. Allergies lead to sneezing.
   allergies(X) → sneeze(X)
2. Cats cause allergies if allergic to cats.
   cat(Y) ∧ allergic-cats(X) → allergies(X)
3. Felix is a cat.
   cat(Felix)
4. Lise is allergic to cats.
   allergic-cats(Lise)

sneeze(Lise) ← infer truth of (query)

- Find and apply relevant rules

variable binding

cat(Y) ∧ allergic-cats(X) → allergies(X) ∧ cat(Felix)
→
cat(Felix) ∧ allergic-cats(X) → allergies(X) ∧ allergic-cats(Lise)
→
allergies(Lise) ∧ allergies(X) → sneeze(X)
→
sneeze(Lise) ✓

add new sentence to KB

87

Backward Chaining

sneeze(Lise) ← query

- Backward Chaining: apply rules that end with the goal

variable binding

allergies(X) → sneeze(X) + sneeze(Lise)
new query: allergies(Lise)?

cat(Y) ∧ allergic-cats(X) → allergies(X) + allergies(Lise)
new query: cat(Y) ∧ allergic-cats(Lise)?

cat(Felix) + cat(Y) ∧ allergic-cats(Lise)
new sentence: cat(Felix) ∧ allergic-cats(Lise) ✓

Knowledge Base
1. Allergies lead to sneezing.
allergies(X) → sneeze(X)
2. Cats cause allergies if allergic to cats.
cat(Y) ∧ allergic-cats(X) → allergies(X)
3. Felix is a cat.
cat(Felix)
4. Lise is allergic to cats.
allergic-cats(Lise)

88

Uses of Inference

- Ontologies
  - Conclude new information
  - Sanity check
- Semantic Networks
  - Conclude new information
  - Build out network
  - Maintain probabilities
- Planning

89

Planning

- Classical Planning
- Partial-order planning
- Probabilistic planning

90

Planning Problem

- Find a sequence of actions [operations] that achieves a goal when executed from the initial world state.
- That is, given:
  - A set of operator descriptions (possible primitive actions by the agent)
  - An initial state description
  - A goal state (description or predicate)
- Compute a plan, which is
  - A sequence of operator instances [operations]
  - Executing them in initial state → state satisfying description of goal-state

91

With "Situations"

- **Initial state** and **Goal state** with explicit situations

  $At(Home, S_0) \wedge \neg Have(Milk, S_0) \wedge \neg Have(Bananas, S_0) \wedge \neg Have(Drill, S_0)$

  $(\exists s)\ At(Home,s) \wedge Have(Milk,s) \wedge Have(Bananas,s) \wedge Have(Drill,s)$

- **Operators:**

  $\forall(a,s)\ Have(Milk,Result(a,s)) \Leftrightarrow$
  $((a=Buy(Milk) \wedge At(Grocery,s)) \vee$
  $(Have(Milk, s) \wedge a \neq Drop(Milk)))$

  $\forall(a,s)\ Have(Drill,Result(a,s)) \Leftrightarrow$
  $((a=Buy(Drill) \wedge At(HardwareStore,s)) \vee$
  $(Have(Drill, s) \wedge a \neq Drop(Drill)))$

92

With Implicit Situations

- **Initial state**

  $At(Home) \wedge \neg Have(Milk) \wedge \neg Have(Bananas) \wedge \neg Have(Drill)$

- **Goal state**

  $At(Home) \wedge Have(Milk) \wedge Have(Bananas) \wedge Have(Drill)$

- **Operators:**

  $Have(Milk) \Leftrightarrow$
  $((a=Buy(Milk) \wedge At(Grocery)) \vee (Have(Milk) \wedge a \neq Drop(Milk)))$

  $Have(Drill) \Leftrightarrow$
  $((a=Buy(Drill) \wedge At(HardwareStore)) \vee (Have(Drill) \wedge a \neq Drop(Drill)))$

93

Planning as Inference

At(Home) ∧ ¬Have(Milk) ∧ ¬Have(Drill)
At(Home) ∧ Have(Milk) ∧ Have(Drill)

- Knowledge Base for MilkWorld
  - What do we have? Not have?
  - How does one "have" things? (2 rules recommended)
  - Where are drills sold?
  - Where is milk sold?
  - What actions do we have available?

94

Planning as Inference

At(Home) ∧ ¬Have(Milk) ∧ ¬Have(Drill)
At(Home) ∧ Have(Milk) ∧ Have(Drill)

- Knowledge Base for MilkWorld
  - What do we have? Not have?
  - How does one "have" things? (2 rules
  - Where are drills sold?
  - Where is milk sold?
  - What actions do we have available?

<u>Knowledge Base</u>
1. We're currently home.

2. We don't have anything.

3. One has things when they are bought at *appropriate* places.

4. You have things you already have and haven't dropped.

5. Hardware stores sell drills.

6. Groceries sell milk.

7. Our actions are:

95

Inference

- What two things do we combine first (by number)?
  - How about 1 and 7(a)?
  - action 1 = Go(GS)
  - action 2 = Buy(Drill)
- What then changes in the knowledge base?
  - ¬At(X)
  - At(GS)

And so on…

Knowledge Base

1. We're currently home.
   At(Home)
2. We don't have anything.
   ¬Have(Drill)
   ¬Have(Milk)
3. One has things when they are bought at *appropriate* places.
   Have(X) ⇔
   (At(Y) ∧ (Sells(X,Y) ∧ (a=Buy(X))
4. You have things you already have and haven't dropped.
   (Have(X) ∧ a ≠ Drop(X)))
5. Hardware stores sell drills.
   (Sells(Drill,HWS)
6. Groceries sell milk.
   (Sells(Milk,GS)
7. Our actions are:
   At(X) ∧ Go(Y) => At(Y) ∧ ¬At(X)
   Drop(X) => ¬Have(X)
   Buy(X) [defined above]

96

Partial-Order Planning

- A **linear planner** builds a plan as a **totally ordered sequence** of plan steps
- A non-linear planner (aka **partial-order planner**) builds up a plan as a set of steps with some temporal constraints
  - E.g., S1<S2 (step S1 must come before S2)
- Partially ordered plan (POP) refined by either:
  - adding a new plan step, or
  - adding a new constraint to the steps already in the plan.
- A POP can be linearized (converted to a totally ordered plan) by topological sorting*

97

Non-Linear Plan: Steps

- A non-linear plan consists of

 (1) A set of **steps** $\{S_1, S_2, S_3, S_4 \ldots\}$

 Each step has an **operator description**, **preconditions** and **post-conditions**

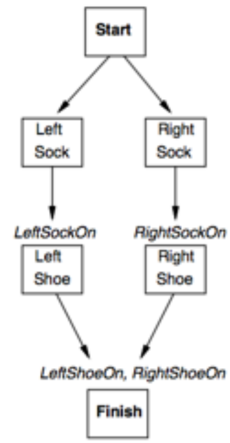
 (2) A set of **causal links** $\{ \ldots (S_i,C,S_j) \ldots\}$

 (One) goal of step S_i is to achieve precondition C of step S_j

 (3) A set of **ordering constraints** $\{ \ldots S_i<S_j \ldots \}$

 if step S_i must come before step S_j

Partial Order Plan:



98

Back to Milk World…

- Actions:

1. Go(GS)
2. Buy(Milk)
3. Go(HWS)
4. Buy(Drill)
5. Go(Home)

- Does ordering matter?

<u>Knowledge Base</u>

1. We're currently home.
 At(Home) ← this was not true throughout!
2. We don't have anything.
 ¬Have(Drill)
 ¬Have(Milk)
3. One has things when they are bought at *appropriate* places.
 Have(X) ⇔
 (At(Y) ∧ (Sells(X,Y) ∧ (a=Buy(X))
4. You have things you already have and haven't dropped.
 (Have(X) ∧ a ≠ Drop(X)))
5. Hardware stores sell drills.
 (Sells(Drill,HWS)
6. Groceries sell milk.
 (Sells(Milk,GS)
7. Our actions are:
 At(X) ∧ Go(Y) => At(Y) ∧ ¬At(X)
 Drop(X) => ¬Have(X)
 Buy(X) [defined above]

99

Specifying Steps and Constraints

- Go(X)
  - Preconditions: ¬At(X)
  - Postconditions: At(X)
- Buy(T)
  - Preconditions: At(Z) ^ Sells(T, Z)
  - Postconditions: Have(T)
- Causal Links: Go(X) → At(X)
- Ordering Constraints: Go(X) < At(X)

100

POP Constraints and Search Heuristics

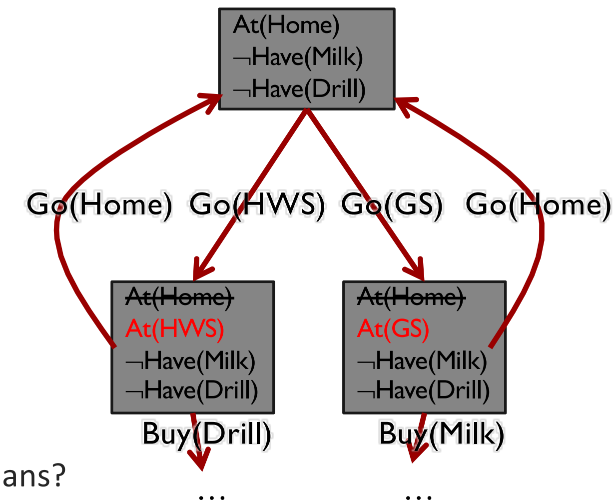
- Only add steps that reach a not-yet-achieved precondition
- Use a least-commitment approach:
  - Don't order steps unless they need to be ordered
- Honor causal links $S_1 \rightarrow S_2$ that **protect** a condition c:
  - Never add an intervening step S_3 that violates c
  - If a parallel action **threatens** c (i.e., has the effect of negating or clobbering c), resolve that threat by adding ordering links:
    - Order S_3 before S_1 (**demotion**)
    - Order S_3 after S_2 (**promotion**)

101

Eventually…

1. Go(GS)
2. Buy(Milk)
3. Go(HWS)
4. Buy(Drill)
5. Go(Home)

- Ordering is not strict.
- Go(HWS) preconditions:
  - ¬At(HWS) ^ ¬Have(Drill)
- So, 1<2, 3<4
- How many non-loopy paths – i.e., plans?



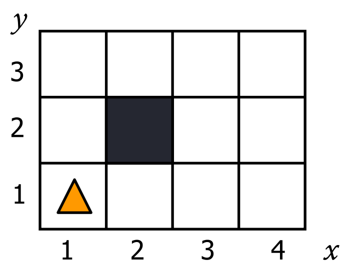
102

Probabilistic Planning

- Core idea: instead of actions having single effects:
  - a1: A → B   a2: B → C
- Actions have possible effect**s**, requiring a table:
  - a1: A → B: 80%   a2: B → C: 80%
  - a1: A → A: 20%   a2: B → B: 20%
- At each plan step, propagate probabilities forward
  - Where am I now, **with what probability?**

103

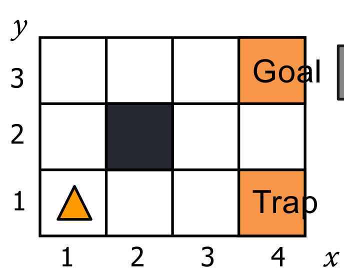
Transition Model in Practice



- In each state, the possible actions are U, D, R, and L
- The effect of U is as follows (transition model):
  - With probability 0.8, the robot moves up one square (if the robot is already in the top row, then it does not move)
  - With probability 0.1, the robot moves right one square (if the robot is already in the rightmost row, then it does not move)
  - With probability 0.1, the robot moves left one square (if the robot is already in the leftmost row, then it does not move)
- D, R, and L have similar probabilistic effects

104

Transition Model in Practice



- In each state, possible actions are U, D, R, and L
- The transition model) of U is:
  - up: 0.8
  - left: 0.1
  - right: 0.1
- D, R, and L have similar probabilistic effects

Plan: U, U, R, R, R

- Where am I?
  - Step 1: (1,2): 0.8 (1,1): 0.1 (2,1): 0.1
  - Step 2: (1,2) → (1,3): 0.8
    - (1,2) → (1,2): 0.1
    - (1,2) → (1,2): 0.1
    - (1,1) → (1,1): 0.1
    - (1,1) → (1,2): 0.8
    - (1,1) → (2,1): 0.1
    - …
- Now: What are the odds I'm at 1,3? 1,2?

105

What does that mean?

- We must evaluate each sequence of actions
  - "Utility"
- Based on what we believe about events
  - But we can replan throughout
- In practice, we define (or learn) a *policy*.
  - I'm at X. What's best at X?
    - And does it matter how I got there? No – this is a Markovian problem.
- Value Iteration?
  - 17.13, 17.17

106

Reinforcement Learning

- **Reinforcement learning systems**
  - Learn **series** of actions or decisions, rather than a single decision
  - Based on feedback given at the end of the series
- A reinforcement learner has
  - A goal
  - Carries out trial-and-error search
  - Finds the best paths toward that goal

107

Reinforcement Learning

- A typical reinforcement learning system is an active agent, interacting with its environment.
- It must balance
  - Exploration: trying different actions and sequences of actions to discover which ones work best
  - Exploitation (achievement): using sequences which have worked well so far
- Must learn **successful sequences of actions** in an uncertain environment

108

Learning States and Actions

- A typical approach is:
- At state S choose, some action A ← How?
- Taking us to new State S_1
  - If S_1 has a positive value: increase value of A at S.
  - If S_1 has a negative value: decrease value of A at S.
  - If S_1 is new, initial value is unknown: value of A unchanged.
- One complete learning pass or **trial** eventually gets to a terminal, deterministic state. (E.g., “win” or “lose”)
- Repeat until? Convergence? Some performance level?

109

Exploration vs. Exploitation

- Problem with naïve reinforcement learning:
  - What action to take?
  - **Best apparent action, based on learning to date** } Exploitation
    - Greedy strategy
    - Often prematurely converges to a suboptimal policy!
  - **Random (or unknown) action** } Exploration
    - Will cover entire state space
    - Very expensive and slow to learn!
    - When to stop being random?
- Balance exploration (try random actions) with exploitation (use best action so far)

110

Clustering

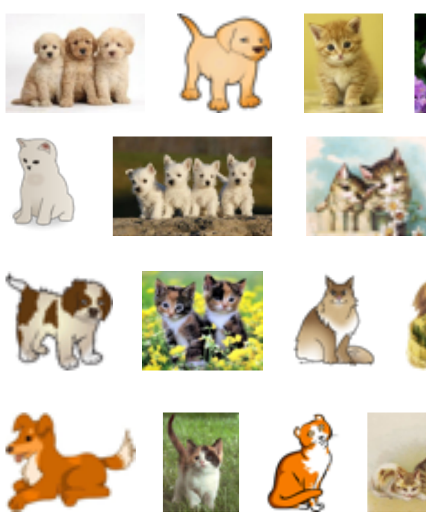
- Given some instances with examples
  - But no labels!
  - Unsupervised learning — the instances do not include a "class"
- Group instances such that:
  - Examples within a group (cluster) are similar
  - Examples in different groups (cluster) are different
- According to some *measure of similarity*, or **distance metric**.
  - Finding the right **features** and **distance metric** are important!

111

Example

112

Example



- What are some two-way clusters we might get? Three way?
  - cats/dogs
  - photos/drawings
  - tan/white/striped
- What are some good features for cats/dogs?
  - Ear pointiness, tail length, …
  - Distance metric for tail length?
- What about the others?

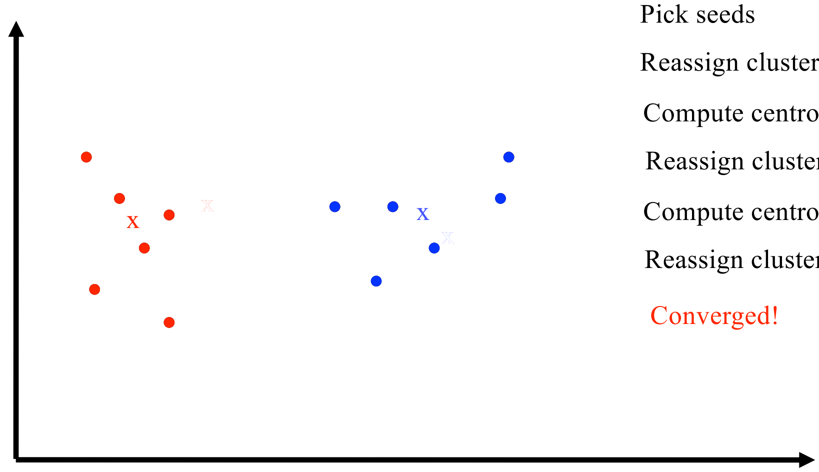
113

K-Means Clustering

- Provide number of desired clusters, k.
- Randomly choose k instances as seeds.
- Form initial clusters based on these seeds.
- Calculate the centroid of each cluster.
- Iterate, repeatedly reallocating instances to closest centroids and calculating the new centroids
- Stop when clustering converges or after a fixed number of iterations.

114

K Means Example (K=2)

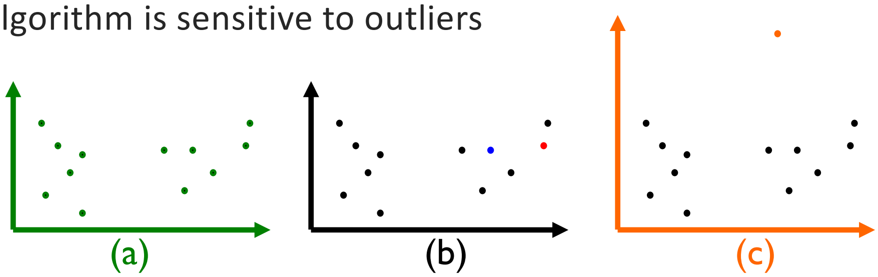


Pick seeds
Reassign clusters
Compute centroids
Reassign clusters
Compute centroids
Reassign clusters
Converged!

115

K-Means

- Tradeoff: more clusters (better focused clusters) and too many clusters (overfitting)
  - What would we likely get for 3 clusters? 4?
- Results can vary based on random seed selection
  - What if these were our starting points?
- The algorithm is sensitive to outliers



116

EM Summary

- Basically a probabilistic K-Means.
- Has many of same advantages and disadvantages
  - Results are easy to understand
  - Have to choose k ahead of time
- Useful in domains where we would prefer the likelihood that an instance can belong to more than one cluster
  - Natural language processing for instance

117