



Lecture 15.2: Conclusion



Roadmap

Summary of CS221

Next courses

History of AI

Food for thought

Paradigm

Modeling

Inference

Learning

Course plan

Search problems

Markov decision processes

Constraint satisfaction problems

Adversarial games

Bayesian networks

Reflex

States

Variables

Logic

"Low-level intelligence"

"High-level intelligence"

Machine learning

Machine learning

Objective: loss minimization

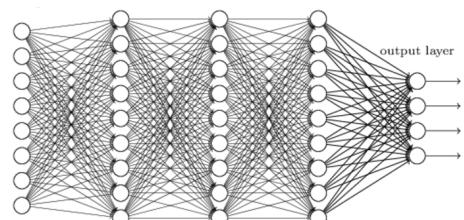
$$\min_{\mathbf{w}} \sum_{(x,y) \in \mathcal{D}_{\text{train}}} \text{Loss}(x, y, \mathbf{w})$$

Algorithm: stochastic gradient descent

$$\mathbf{w} \rightarrow \mathbf{w} - \eta_t \underbrace{\nabla \text{Loss}(x, y, \mathbf{w})}_{\text{prediction} - \text{target}}$$

Applies to wide range of models!

Reflex-based models

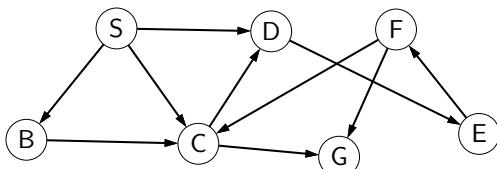


Models: linear models, neural networks, nearest neighbors

Inference: feedforward

Learning: SGD, alternating minimization

State-based models



Key idea: state

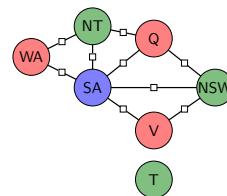
A **state** is a summary of all the past actions sufficient to choose future actions **optimally**.

Models: search problems, MDPs, games

Inference: UCS/A*, DP, value iteration, minimax

Learning: structured Perceptron, Q-learning, TD learning

Variable-based models



Key idea: factor graphs

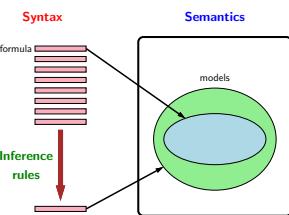
Graph structure captures conditional independence.

Models: CSPs, Bayesian networks

Inference: backtracking, forward-backward, beam search, Gibbs sampling

Learning: maximum likelihood (closed form, EM)

Logic-based models



Formulas enable more powerful models (infinite).

Models: propositional logic, first-order logic

Inference: model checking, modus ponens, resolution

Learning: ???

- CS221 provides a set of tools



- Start with the problem, and figure out what tool to use
- Keep it simple!

Roadmap

Summary of CS221

Next courses

History of AI

Food for thought

Other AI-related courses

<http://ai.stanford.edu/courses/>

Foundations:

- CS228: Probabilistic Graphical Models
- CS229: Machine Learning
- CS229T: Statistical Learning Theory
- CS230: Deep Learning
- CS334A: Convex Optimization
- CS238: Decision Making Under Uncertainty
- CS257: Logic and Artificial Intelligence
- CS246: Mining Massive Data Sets

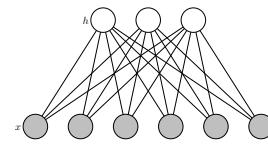
Other AI-related courses

<http://ai.stanford.edu/courses/>

Applications:

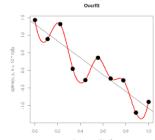
- CS224N: Natural Language Processing (with Deep Learning)
- CS224U: Natural Language Understanding
- CS231A: From 3D Reconstruction to Recognition
- CS231N: Convolutional Neural Networks for Visual Recognition
- CS223A: Introduction to Robotics
- CS227B: General Game Playing

Probabilistic graphical models (CS228)



- Markov networks (undirected probabilistic graphical models)
- Forward-backward, variable elimination \Rightarrow belief propagation, variational inference
- Gibbs sampling \Rightarrow Markov Chain Monte Carlo (MCMC)
- Learning the structure

Machine learning (CS229)



- Linear models \Rightarrow kernel methods, decision trees
- Boosting, bagging, feature selection
- K-means \Rightarrow mixture of Gaussians, PCA, ICA

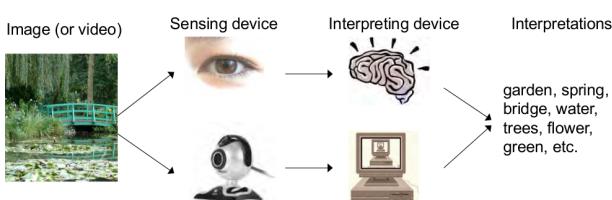
Statistical learning theory (CS229T)

Question: what are the mathematical principles behind learning?

Uniform convergence: with probability at least 0.95, your algorithm will return a predictor $h \in \mathcal{H}$ such that

$$\text{TestError}(h) \leq \text{TrainError}(h) + \sqrt{\frac{\text{Complexity}(\mathcal{H})}{n}}$$

Vision (CS231A, CS231N)



- **Challenges:** variation in viewpoint, illumination, intra-class variation
- **Tasks:** object recognition/detection/segmentation, pose estimation, 3D reconstruction, image captioning, visual question answering

Robotics (CS223A, CS327A)

- **Tasks:** manipulation, grasping, navigation



- **Applications:** self-driving cars, medical robotics
- **Physical models:** kinematics, control

Language (CS224N, CS224U)

- Designed by humans for communication
- World: continuous, words: discrete, meanings: continuous
- **Properties:** compositionality, grounding



- **Tasks:** syntactic parsing, semantic parsing, information extraction, coreference resolution, machine translation, question answering, summarization, dialogue

CS221 / Summer 2019 / Jia

18 CS221 / Summer 2019 / Jia

Roadmap

Summary of CS221

Next courses

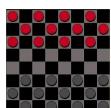
History of AI

Food for thought

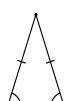
CS221 / Summer 2019 / Jia

20

Birth of AI, early successes



Checkers (1952): Samuel's program learned weights and played at strong amateur level



Problem solving (1955): Newell & Simon's Logic Theorist: prove theorems in Principia Mathematica using search + heuristics; later, General Problem Solver (GPS)

CS221 / Summer 2019 / Jia

22 CS221 / Summer 2019 / Jia

Cognitive science



Question: How does the human mind work?

- Cognitive science and AI grew up together
- Humans can learn from few examples on many tasks

Computation and cognitive science (PSYCH204):

- Cognition as Bayesian modeling — probabilistic program [Tenenbaum, Goodman, Griffiths]

19 1966: Workshop at Dartmouth College; attendees: John McCarthy, Marvin Minsky, Claude Shannon, etc.
Aim for **general principles**:
Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.
CS221 / Summer 2019 / Jia



Overwhelming optimism...

Machines will be capable, within twenty years, of doing any work a man can do. —Herbert Simon

Within 10 years the problems of artificial intelligence will be substantially solved. —Marvin Minsky

I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines. —Claude Shannon

CS221 / Summer 2019 / Jia

23 CS221 / Summer 2019 / Jia

Overwhelming optimism...

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

—New York Times, 1958

...underwhelming results

Example: machine translation

The spirit is willing but the flesh is weak.

↓
(Russian)

The vodka is good but the meat is rotten.

1966: ALPAC report cut off government funding for MT

Implications of early era

Problems:

- **Limited computation:** search space grew exponentially, outpacing hardware ($100! \approx 10^{157} > 10^{80}$)
- **Limited information:** complexity of AI problems (number of words, objects, concepts in the world)

Contributions:

- Lisp, garbage collection, time-sharing (John McCarthy)
- **Key paradigm:** separate **modeling** (declarative) and **inference** (procedural)

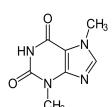
Knowledge-based systems (70-80s)



Expert systems: elicit specific domain knowledge from experts in form of rules:

if [premises] then [conclusion]

Knowledge-based systems (70-80s)



DENDRAL: infer molecular structure from mass spectrometry



MYCIN: diagnose blood infections, recommend antibiotics



XCON: convert customer orders into parts specification; save DEC \$40 million a year by 1986

Knowledge-based systems

Contributions:

- First **real application** that impacted industry
- Knowledge helped curb the exponential growth

Problems:

- Knowledge is not deterministic rules, need to model **uncertainty**
- Requires considerable **manual effort** to create rules, hard to maintain

SHRDLU [Winograd 1971]

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.

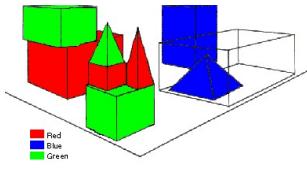
Computer: OK.

Person: What does the box contain?

Computer: The blue pyramid and the blue block.

Person: What is the pyramid supported by?

Computer: The box.



30 CS221 / Summer 2019 / Jia

The Complexity Barrier

A number of people have suggested to me that large programs like the SHRDLU program for understanding natural language represent a kind of **dead end** in AI programming. **Complex interactions** between its components give the program much of its power, but at the same time they present a formidable obstacle to understanding and extending it. In order to grasp any part, it is necessary to understand how it fits with other parts, presents a dense mass, with **no easy footholds**. Even having written the program, I find it near the limit of what I can keep in mind at once.

— Terry Winograd (1972)

Modern AI (90s-present)

- **Probability**: Pearl (1988) promote Bayesian networks in AI to **model uncertainty** (based on Bayes rule from 1700s)

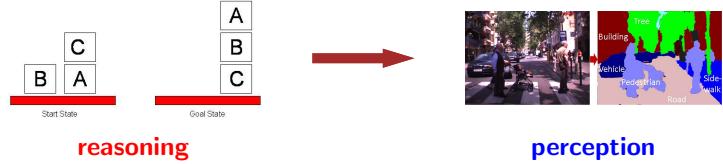
model → predictions

- **Machine learning**: Vapnik (1995) invented support vector machines to **tune parameters** (based on statistical models in early 1900s)

data → model

31 CS221 / Summer 2019 / Jia

Modern AI (90s-present)



32 CS221 / Summer 2019 / Jia

Ideas from other fields

- Bayes rule (Bayes, 1763) from probability
- Least squares regression (Gauss, 1795) from astronomy
- First-order logic (Frege, 1893) from logic
- Maximum likelihood (Fisher, 1922) from statistics
- Stochastic gradient descent (Robbins/Monro, 1951) from optimization
- Uniform cost search (Dijkstra, 1956) from algorithms
- Value iteration (Bellman, 1957) from control theory



Roadmap

Summary of CS221

Next courses

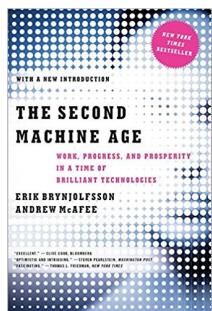
History of AI

Food for thought

33 CS221 / Summer 2019 / Jia

Outlook

AI is everywhere: consumer services, advertising, transportation, manufacturing, etc.

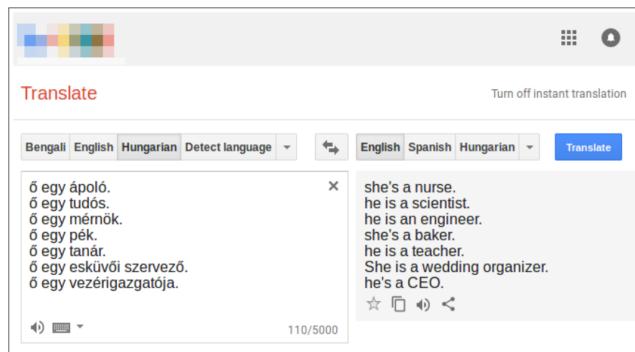


AI being used to make decisions for: education, credit, employment, advertising, healthcare and policing

CS221 / Summer 2019 / Jia

36 CS221 / Summer 2019 / Jia

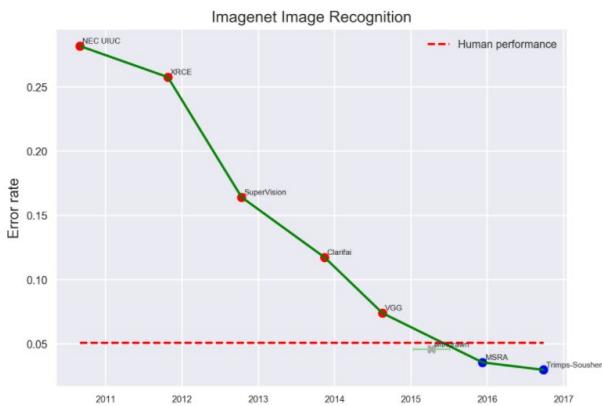
Biases



CS221 / Summer 2019 / Jia

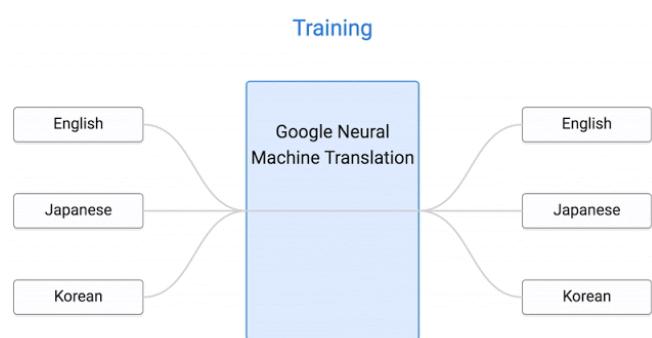
38 | CS221 / Summer 2019 / Jia

Image classification



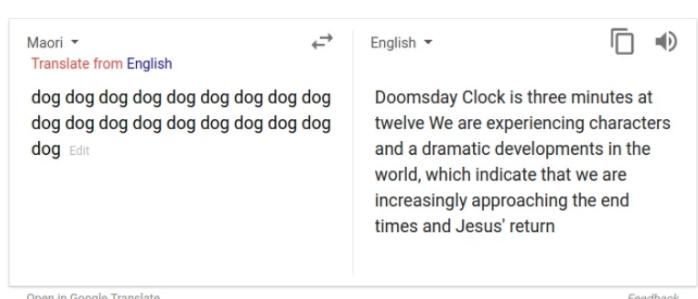
ES221 / Summer 2019 / 13

Google Machine Translation (2016)



36 CS221 / Summer 2019 / Jia

Craziness

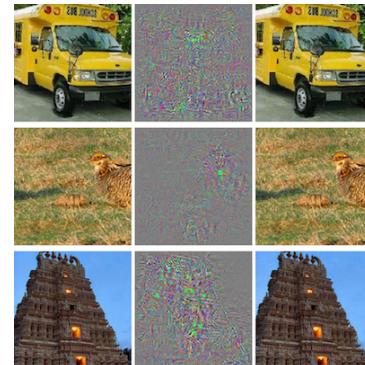


CS221 / Summer 2019 / 13

40 CS221 / Summer 2019 / 13

Adversaries

AlexNet predicts correctly on the left



AlexNet predicts **ostrich** on the right

40 CS221 / Summer 2019 / 13

41



CS221 / Summer 2019 / Jia

42 CS221 / Summer 2019 / Jia

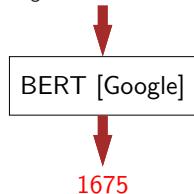
Security



Reading comprehension

Individual Huguenots settled at the Cape of Good Hope from as early as 1671 with the arrival of Francois Villon (Viljoen). The first Huguenot to arrive at the Cape of Good Hope was however Maria de la Quellierie, wife of commander Jan van Riebeeck (and daughter of a Walloon church minister), who arrived on 6 April 1652 to establish a settlement at what is today Cape Town. The couple left for the Far East ten years later. On 31 December 1687 the first organised group of Huguenots set sail from the Netherlands to the Dutch East India Company post at the Cape of Good Hope. The largest portion of the Huguenots to settle in the Cape arrived between 1688 and 1689 in seven ships as part of the organised migration, but quite a few arrived as late as 1700; thereafter, the numbers declined and only small groups arrived at a time. **The number of old Acadian colonists declined after the year 1675.**

The number of new Huguenot colonists declined after what year?



CS221 / Summer 2019 / Jia

44 CS221 / Summer 2019 / Jia

Results on SQuAD models

Model	Original F1	Adversarial F1
Humans	92.6	89.2
BERT	93.2	70.7
SLQA+	88.6	64.2
r-net+	88.5	63.4
ReasoNet-E	81.1	49.8
SEDT-E	80.1	46.5
BiDAF-E	80.0	46.9
Mnemonic-E	79.1	55.3
Ruminating	78.8	47.7
jNet	78.6	47.0
Mnemonic-S	78.5	56.0

Optimizing for clicks



Is this a good objective function for society?



Can build it \neq should build it?

CS221 / Summer 2019 / Jia

46 CS221 / Summer 2019 / Jia

47

Privacy

- Not reveal sensitive information (income, health, communication)
- Compute average statistics (how many people have cancer?)



[Warner, 1965]

Privacy: Randomized response

Do you have a sibling?



Method:

- Flip two coins.
- If both heads: answer yes/no randomly
- Otherwise: answer yes/no truthfully

Analysis:

$$\text{true-prob} = \frac{4}{3} \times (\text{observed-prob} - \frac{1}{8})$$

Interpretability versus accuracy

- For air-traffic control, threshold level of safety: probability 10^{-9} for a catastrophic failure (e.g., collision) per flight hour
- Move from human designed rules to a numeric Q-value table?

yes

Societal and industrial impact



Enormous potential for positive impact, use responsibly!



Principles for Accountable Algorithms and a Social Impact Statement for Algorithms

There is always a human ultimately responsible for decisions made or informed by an algorithm. "The algorithm did it" is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes



Machine learning

Please fill out course evaluations on Axess.

Thanks for an exciting quarter!