COGS 200: Introduction to Cognitive Systems

Bob Woodham

Department of Computer Science
University of British Columbia

Lecture Notes 2014/2015 Term 1

14W1: September–December, 2014

Today's Reading

Today's assigned "reading" is twofold:

1. Reprise Peter Norvig's September 23, 2010, talk, "The unreasonable effectiveness of data," UBC Computer Science Distinguished Lecture Series. (video)

2. Peter Norvig's "On Chomsky and the two cultures of statistical learning." (on line essay)

Noam Chomsky was one of six presenters in the "Keynote Panel: The Golden Age – A Look at the Original Roots of Artificial Intelligence, Cognitive Science, and Neuroscience" at the MIT150 Celebration Symposium: Brains, Minds, and Machines, May 3–5, 2011. Norvig's essay is in response to an answer Chomsky gave to a question posed by Steven Pinker, the panel moderator

Today's Learning Goals

1. To extend our notion of language to include knowledge representation (KR) and reasoning
2. To distinguish two aspects of an intelligent system's KR:
   - knowledge of the 3D world (objective, external knowledge)
   - knowledge of itself as a problem solver (subjective, internal knowledge)
3. To introduce the Bayesian interpretation of probability and associated statistical inference
4. To identify three exemplar approaches to KR based, respectively, on structured, semi-structured and unstructured "big data" knowledge bases

Methodological Issues (Across Disciplines)

- What problems do each discipline address?
- What questions do each discipline ask about those problems?
- What tools and techniques do each discipline use to answer these questions?
- What defines “success” in each discipline?

Knowledge Representation (KR): Key Issues

Questions to ask of any KR scheme are:

1. What "language" is being used?
2. What knowledge is represented explicitly by a given set of expressions in the language?
3. What inference mechanism is available to make explicit knowledge which is implicit, but not yet explicit, in the given set of expressions?

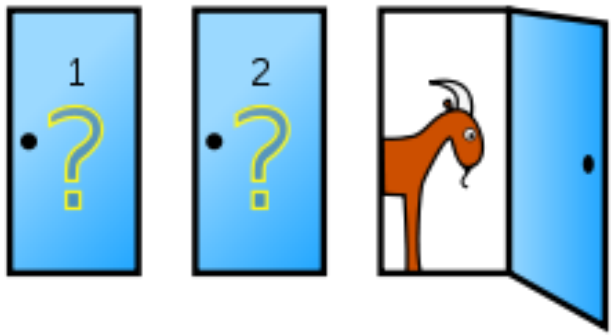
Inference mechanisms can be thought of as methods to reason about (or calculate with) knowledge in order to make explicit that which is only implicit

Example 1: Common Birthday

Number of people	Probability of at least one common birthday
1	0
10	0.117
20	0.411
23	0.507
30	0.706
40	0.891
50	0.970
⋮	⋮
107	0.99999994

Example 2: Monty Hall Problem

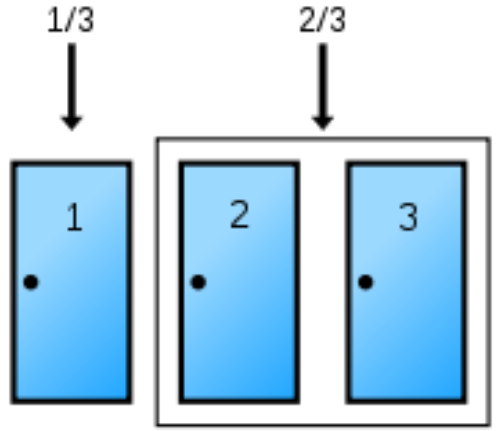
Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?"



Question: Is it to your advantage to switch your choice?

http://en.wikipedia.org/wiki/Monty_Hall_problem

Example 2 (cont'd): Monty Hall Problem

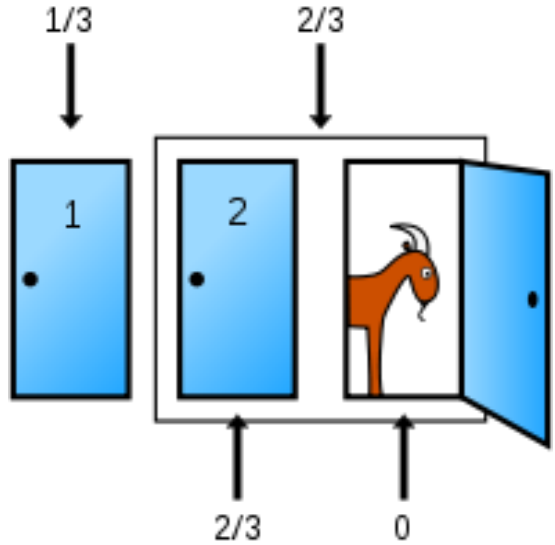


The probability that the car is behind door 1 is $\frac{1}{3}$

The probability that the car is behind one of the two other doors is $\frac{2}{3}$

http://en.wikipedia.org/wiki/Monty_Hall_problem

Example 2 (cont'd): Monty Hall Problem



The host opens door 3 revealing a goat. The probabilities for the two cases, door 1 (alone) and doors 2 and 3 (together), don't change. But, the probability that the car is behind door 3 now is 0

Therefore, the probability that the car is behind door 2 now is $\frac{2}{3}$

http://en.wikipedia.org/wiki/Monty_Hall_problem

Example 3: Meadow's Law

"One sudden infant death is a tragedy, two is suspicious and three is murder until proved otherwise."

Sir Roy Meadow
ABC of Child Abuse, 1989

http://en.wikipedia.org/wiki/Meadow%27s_law

Interpretations of Probability (in Statistics)

1. Frequentist (objective): Probability is the observed frequency of an event, based on repeated sampling of the world. It is obtained from empirical data (i.e., measurements) of the external world

   Example: In Vancouver, it rains 80% of the days in November

2. Bayesian (subjective): Probability is the degree of confidence an agent has regarding an uncertain event. It represents the internal belief an agent has about the external world

   Example: The patient has a 70% chance of having cancer

Interpretations of Probability (in Statistics)

1. Frequentist (objective): Probability is the observed frequency of an event, based on repeated sampling of the world. It is obtained from empirical data (i.e., measurements) of the external world

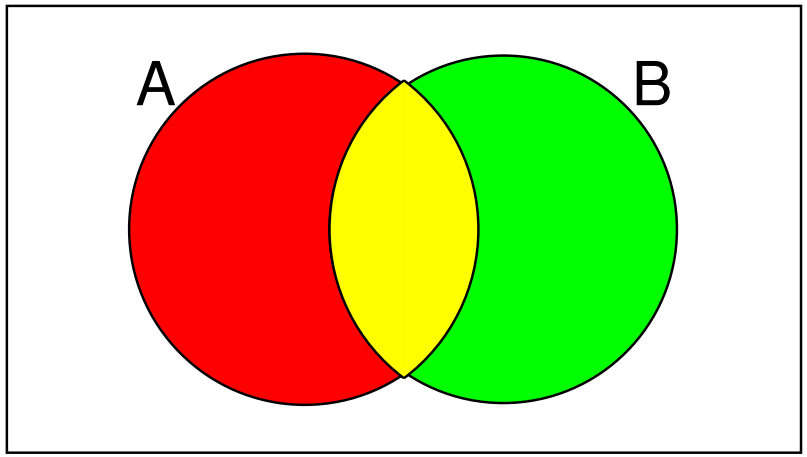
   Example: In Vancouver, it rains 80% of the days in November

2. Bayesian (subjective): Probability is the degree of confidence an agent has regarding an uncertain event. It represents the internal belief an agent has about the external world

   Example: The patient has a 70% chance of having cancer

Conditional Probability

A Venn diagram shows the intersection, $A \cap B$, of two sets A and B



$A \cap B$

We can write $P(A \cap B)$ in two ways

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ P(A \cap B) &= P(B|A)P(A) \end{aligned}$$

so that

$$P(A|B)P(B) = P(B|A)P(A)$$

or

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' Theorem (aka Bayes' Rule)

Used (historically) to address practical problems:[1]

- assessing evidence in legal proceedings
- setting insurance rates
- decoding the German Enigma cipher during WWII (Alan Turing)
- search and rescue (e.g., recovering lost nuclear bombs, submarines, etc)
- assessing likelihood of nuclear accidents
- verifying authorship of the US Federalist Papers

[1] McGrayne, S. B. (2011). *The theory that would not die*. Yale University Press

Example 4: US Federalist Papers

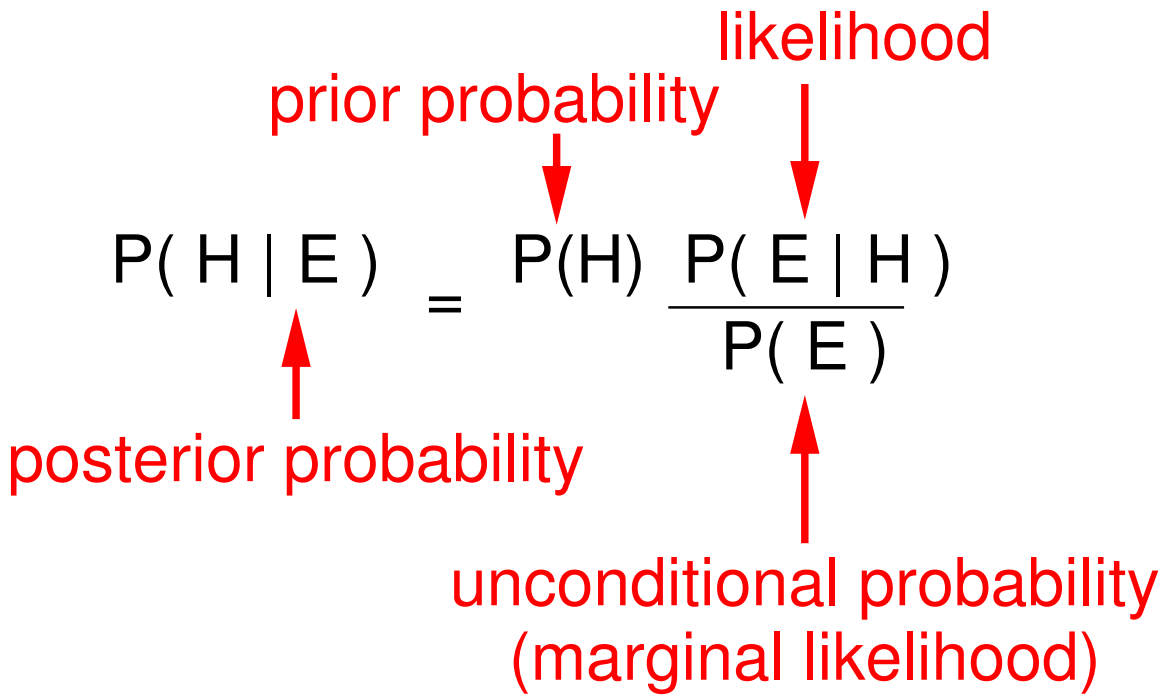
"The 1964 publication of Inference and Disputed Authorship *made the cover of Time magazine and the attention of academics and the public alike for its use of statistical methodology to solve one of American history's most notorious questions: the disputed authorship of the Federalist Papers.*

Back in print for a new generation of readers, this classic volume applies mathematics, including the once-controversial Bayesian analysis, into the heart of a literary and historical problem by studying frequently used words in the texts. The reissue of this landmark book will be welcomed by anyone interested in the juncture of history, political science, and authorship."

Book description (for the 2008 re-issue) at amazon.com.

Bayes' Theorem (aka Bayes' Rule)

Let H be the hypothesis (i.e., belief) and let E be the evidence



Bayesian Inference

Bayesian inference is an iterative process in which new evidence repeatedly updates an initial probability distribution

Before any evidence is taken into account, one starts with some belief about H, expressed as an initial prior probability

At each iteration step, the prior probability is taken to be the posterior probability from the previous iteration

See http://en.wikipedia.org/wiki/Bayesian_inference

“Big Data” (Wikipedia Definition)

Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis and visualization. . .

. . . The world’s technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s. As of 2012, every day 2.5 exabytes (2.5×10^{18}) of data were created

http://en.wikipedia.org/wiki/Big_data

“Big Data” (Vancouver Sun)

“Move over, firefighters and Internet magnates. The sexiest new jobs belong to statisticians.”

“Deluge of data creates advantage for tech-savvy companies”
Scott Simpson, Vancouver Sun, September 30, 2011

“Big Data” Knowledge Bases:

Approach	Example
Structured	Cyc
Semi-structured	Wikipedia
Unstructured	Google (the entire WWW)

Cyc

The objective of Cyc is to codify, in a formalized representation, sufficient “pieces of knowledge” to comprise human common sense. In 1986, Lenat estimated the effort to complete Cyc would be 250,000 rules and 350 man-years of effort

In 1994, Cycorp, Inc. was founded “to further develop, commercialize, and apply the Cyc technology”

Cyc uses a proprietary knowledge representation language and associated inference engine based on first-order predicate calculus (FOPC) with some extensions

Versions of Cyc have been made available under an open source licence

Cyc

The objective of Cyc is to codify, in a formalized representation, sufficient “pieces of knowledge” to comprise human common sense. In 1986, Lenat estimated the effort to complete Cyc would be 250,000 rules and 350 man-years of effort

In 1994, Cycorp, Inc. was founded “to further develop, commercialize, and apply the Cyc technology”

Cyc uses a proprietary knowledge representation language and associated inference engine based on first-order predicate calculus (FOPC) with some extensions

Versions of Cyc have been made available under an open source licence

Cyc Update

An article dated July 3, 2014, in Business Insider, India, titled "The Most Ambitious Artificial Intelligence Project In The World Has Been Operating In Near-Secrecy For 30 Years" quotes Doug Lenat,

> *"We've been keeping a very low profile, mostly intentionally. No outside investments, no debts. We don't write very many articles or go to conferences, but for the first time, we're close to having this be applicable enough that we want to talk to you."*

(Selected) Wikipedia Statistics

Page statistics	
(English) Content pages	4,622,986
All pages (including talk pages, redirects, etc.)	34,003,182
Edit statistics	
Page edits since Wikipedia was set up	739,028,174
Average edits per page	21.73
User statistics	
Registered users	22,798,785
Active registered users (performed an action in the last 30 days)	129,522

http://en.wikipedia.org/wiki/Special:Statistics

October 14, 2014

Wikipedia Other Languages

- More than 1,000,000 articles: Deutsch · español · français · italiano · Nederlands · polski · русский · svenska
- More than 400,000 articles: català · 日本語 · norsk bokmål · português · Tiếng Việt · українська · 中文
- More than 200,000 articles: العربية · Bahasa Indonesia · Bahasa Melayu · čeština · српски / srpski · فارسی · 한국어 · magyar · română · suomi · Türkçe
- More than 50,000 articles: български · dansk · eesti · Ελληνικά · English (simple) · Esperanto · euskara · galego · עברית · hrvatski · latviešu · lietuvių · norsk nynorsk · slovenčina · slovenščina · srpskohrvatski / српскохрватски · ไทย

http://en.wikipedia.org/wiki/Main_Page

February 3, 2014

WikipediaMiner

Quoting from http://wdm.cs.waikato.ac.nz:8080/

"WikipediaMiner is a toolkit for tapping the rich semantics encoded within Wikipedia

It makes it easy to integrate Wikipedia's knowledge into your own applications, by:

- providing simplified, object-oriented access to Wikipedia's structure and content
- measuring how terms and concepts in Wikipedia are connected to each other
- detecting and disambiguating Wikipedia topics when they are mentioned in documents

Try the demos and services to see what we mean"