Language and Learning

Introduction to
Artificial Intelligence

COS302

Michael L. Littman

Fall 2001
Administration

Break ok?
Search and AI

Powerful techniques. Do they solve the whole AI problem?
Let’s do a thought experiment.

一百三十九是奇数吗?
Chinese Room Argument

Searle: There is a fundamental difference between symbol manipulation and understanding meaning.

Syntax vs. semantics
Was Searle Right?

Yes/no: The richness of corpora.

What is judgment? Can it be automated?
Famous Quotes

“The difference between chess and crossword puzzles is that, in chess, you know when you’ve won.” --- Michael L. Littman

“Trying is the first stem toward failure.” --- Homer Simpson via my cryptogram program
Cryptogram Example

Is this wrong?
Can you write a program that would agree with you on this?
What would your program be like?
Language Resources

Three major resources:

- Dictionaries
- Labeled corpora
- Unlabeled corpora

Each useful for different purposes.

Examples to follow...
Google

Word matching on large corpus
Hubs and authorities (unlabeled corpus, statistical processing)
Hand tuned ranking function

http://www.google.com
Also machine translation...
Ionaut

Question answering:  
www.ionaut.com

Hand-built question categorization

Named-entity tagger trained from tagged corpus

Large unlabeled text corpus

Hand-tuned ranking rules
Ask Jeeves

Hand-selected web pages and corresponding questions

Proprietary mapping from query to question in database

www.ask.com
NL for DB

Hand constructed rules turn sentences into DB queries

START

Eliza

Chatterbots very popular. Some believe they can replace “customer care specialists”.

Generally a large collection of rules and example text.

http://www.uwec.edu/Academic/Curric/jerzdg/if/WebHelp/eliza.htm
Wordnet

Hand built

Rich interconnections

Showing up as a resource in many systems.

http://www.cogsci.princeton.edu/cgi-bin/webwn
Spelling Correction

Semi-automated selection of confusuable pairs.

System trained on large corpus, giving positive and negative examples (WSJ)

http://l2r.cs.uiuc.edu/~cogcomp/eo/h/spelldemo.html
OneAcross

Large corpus of crossword answers: www.oneacross.com

IR-style techniques to find relevant clues

Ranking function trained from held-out clues

Learns from users
Essay Grading

Unsupervised learning to discover word representations
Labeled graded essays

http://www.knowledge-technologies.com/IEAdemo.html
More Applications

- Word-sense disambiguation
- Part of speech tagging
- Parsing
- Reading comprehension
- Summarization
- Cobot
- Cross-language IR
- Text categorization
Synonyms

Carp = quit, argue, painful, scratch, complain

Latent Semantic Indexing
• Corpus, deep statistical analysis

Pointwise Mutual Information
• Huge corpus, shallow analysis

WordNet...
Analogies

Overcoat:warmth::
- Glove:hand
- Jewelry:wealth
- Slicker:moisture
- Disguise:identification
- Helmet:protection

Dictionary not sufficient
Labeled corpus probably wouldn’t help
Unlabeled corpus, not obvious...
What to Learn

Difference between straight search problems and language
Why learning might help
Three types of resources (hand-created, labeled, unlabeled)
A Rule

Follow this carefully.

___ 是奇数吗?

Rule: If the last character filling in the blank is 一, 三, 五, 七 and 九, say “是”. Otherwise, say “不是”.
Explanation

Note:
一，三，五，七 and 九 mean one, three, five, seven and nine, respectively.

Example:
1) 二十五 is an odd number?
   Q: Is 25 an odd number?
   Yes
2) 五万 is an odd number?
   Q: Is 50,000 an odd number?
   No
3) 一百三十九 is an odd number?
   Q: Is 139 an odd number?
   Yes
1. The value iteration algorithm from the *Games of Chance* lecture can be applied to deterministic games with loops. Argue that it produces the same answer as the “Loopy” algorithm from the *Game Tree* lecture.

2. Write the matrix form of the game tree below.
Game Tree

X-1

Y-2

X-4 +2

Y-3

+5 +2

-1 +4
3. How many times (on average) do you need to flip a coin before you flip 3 heads in a row?  (a) Set this up as a Markov chain, and (b) solve it.
Homework 6 (due 11/14)

1. Use the web to find sentences to support the analogy traffic:street::water:riverbed. Give the sentences and their sources.

2. More soon...