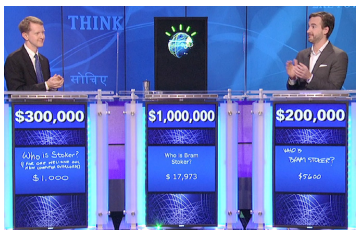
Language Understanding Systems



IBM Watson

Can we create a computer system to compete against the best humans at a task thought to require high levels of intelligence?

NELL

Never-Ending Language Learning System

Tom Mitchell, CMU

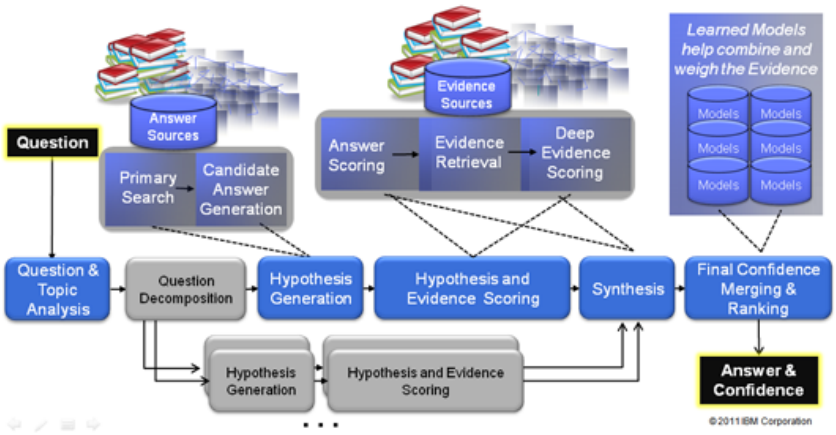
Learns continuously from unstructured information on the Web

Natural language understanding is hard for computers

- intended meanings, nuanced allusions, connotations

Jeopardy is especially hard:

- puzzling clues, breadth of trivia
- not like search engines – need answers!



DeepQA Architecture

Made possible by advances in *computer speed, memory*

- Early implementation ran on a single processor, took 2 hours to answer a single question
- Now scaled up to over 2,500 computer cores, IBM servers, reduced the time to about 3 seconds

Massive parallelism:

- considers ***many*** interpretations/hypotheses simultaneously

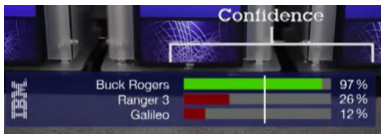
Many experts:

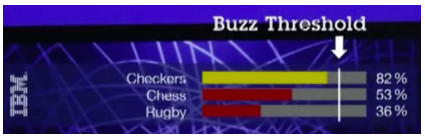
- facilitate integration, application, contextual evaluation of wide range loosely coupled probabilistic question and content "experts"

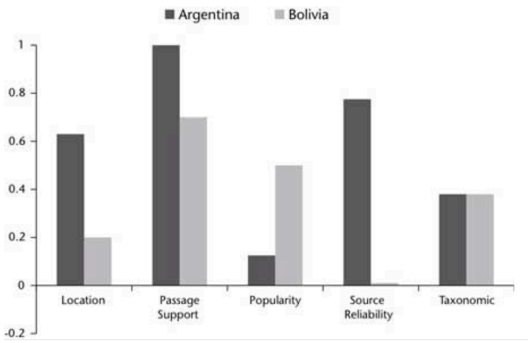
Pervasive confidence estimation:

- no one component commits an answer, all produce features and associated confidences, scoring different question and content interpretations

Confidence Profile from Many Factors







The Outtakes...

Category: "Letters"

Clue: "In the late 40s a mother wrote to this artist that his picture *Number Nine* looked like her son's finger painting"

Correct answer: "Jackson Pollock"

Watson's answer: "Rembrandt"

Reason: Watson failed to recognize that "late 40s" referred to the 1940s

Category: "U.S. City"

Clue: "Its largest airport is named for a World War II hero, its second largest for a World War II battle"

Correct answer: "Chicago"

Watson's answer: "Toronto"

Reasons:

(1) by studying previous competitions, Watson "learned" to pay less attention to the category part of the rule

(2) Watson knew that a Toronto team is in the U.S. baseball league, and one of Toronto's airports is named for a WWI hero

Question: *Can we create a computer system to compete against the best humans at a task thought to require high levels of intelligence?*

Some of Watson's legacy:

- great piece of engineering
- remarkable performance (not thought possible at outset)
- re-ignited public interest in Artificial Intelligence
- new technology with broad applications

But ...

- IBM has not created a machine that thinks like us
- Watson's success does not bring us closer to understanding human intelligence
- Watson's occasional blunders should remind everyone that this problem is still not solved

NELL: Never-Ending Language Learner

"We will never truly understand machine or human learning until we can build computer programs that, *like people,*

- learn many different types of knowledge or functions,
- from years of diverse, mostly self-supervised experience,
- in a staged curricular fashion, where previously learned knowledge enables learning further types of knowledge,
- where self-reflection and the ability to formulate new representations and new learning tasks enable the learner to avoid stagnation and performance plateaus."

Mitchell et al., AAAI, 2015

Natural language *understanding* requires a *belief system*

- I understand, and already knew that
- I understand, and didn't know, but accept it
- I understand, and disagree because ... "

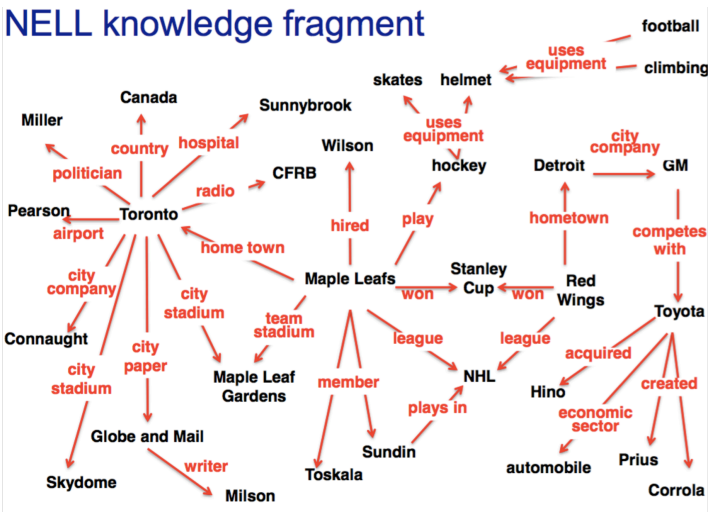
NELL: Never-Ending Language Learner

Inputs:

- initial *ontology* with hundreds of categories and relations to "read about" on the web
  - *categories*, e.g. `person, sportsTeam, fruit, emotion`
  - *relations*, e.g.
    ```
    playsOnTeam(athlete,sportsTeam)
    playsInstrument(musician,instrument)
    ```
- 10-15 examples of each category and relation
- the web (~ 500 million webpages + access to search engines)

The task:

- run continuously, forever
- each day:
  1) extract new instances of categories and relations (noun phrases)
  2) learn to read (perform step (1)) better than yesterday



http://rtw.ml.cmu.edu/rtw

Early Work – Simple "Bootstrap" Learning

Learn which noun phrases are cities:

Too unconstrained!

Paris
Pittsburgh
Seattle
Montpelier

(known)

San Francisco
Berlin
denial

anxiety
selfishness
London

mayor of <arg>
live in <arg>

<arg> is home of
traits such as <arg>

Learn based on multiple cues *simultaneously*, e.g.:

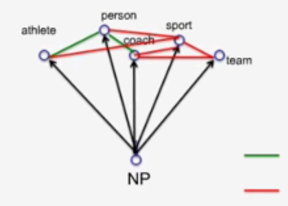
(1) distribution of text contexts (appear in phrases with the same words)

(2) same features of character string (e.g. capitalized, ends with "... burgh")

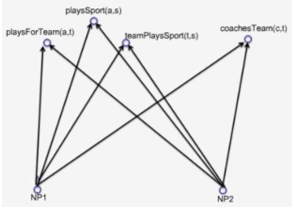
Coupling of Categories & Relations

NELL is *simultaneously* trying to learn:
(1) which noun phrases refer to which categories
(2) which noun phrases participate in which relations

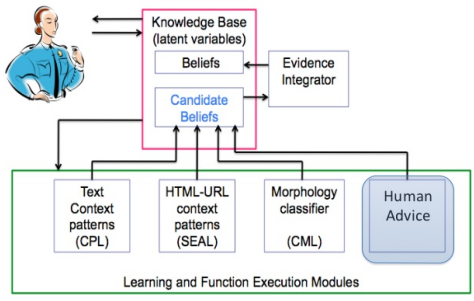
NELL knows:
(1) constraints between categories
e.g. athletes must be people, sports cannot be people
(2) constraints on categories for relations
e.g. `playsSport(athlete,sport)`



athlete(NP) → person(NP)
athlete(NP) → NOT sport(NP)
sport(NP) → NOT athlete(NP)



NELL Architecture



Use expanded knowledge base to "re-train" classifiers

Read web pages to find new candidate beliefs (classify instances of categories and relations)

Further learning by NELL:

(1) new constraints between relations:

```
if (athlete X plays for team Z) and (team Z plays sport Y)
    then (athlete X plays sport Y)
```

(2) new relations:

```
<musical instrument> master <musician>
<mammals> eat <plant>
```

(3) new subcategories, e.g. pets, predators (subcategories of animal)

Learns new relations from frequent co-occurrences of instances of two categories ...

Category Pair	Frequent Instance Pairs	Text Contexts	Suggested Name
MusicInstrument Musician	sitar, George Harrison tenor sax, Stan Getz trombone, Tommy Dorsey vibes, Lionel Hampton	ARG1 master ARG2 ARG1 virtuoso ARG2 ARG1 legend ARG2 ARG2 plays ARG1	Master
Disease Disease	pinched nerve, herniated disk tennis elbow, tendonitis blepharospasm, dystonia	ARG1 is due to ARG2 ARG1 is caused by ARG2	IsDueTo
CellType Chemical	epithelial cells, surfactant neurons, serotonin mast cells, histomine	ARG1 that release ARG2 ARG2 releasing ARG1	ThatRelease
Mammals Plant	koala bears, eucalyptus sheep, grasses goats, saplings	ARG1 eat ARG2 ARG2 eating ARG1	Eat
River City	Seine, Paris Nile, Cairo Tiber river, Rome	ARG1 in heart of ARG2 ARG1 which flows through ARG2	InHeartOf

- athleteWonAward
- animalEatsFood
- languageTaughtInCity
- clothingMadeFromPlant
- beverageServedWithFood
- fishServedWithFood
- athleteBeatAthlete
- athleteInjuredBodyPart
- arthropodFeedsOnInsect
- animalEatsVegetable
- plantRepresentsEmotion
- foodDecreasesRiskOfDisease
- clothingGoesWithClothing
- bacteriaCausesPhysCondition
- buildingMadeOfMaterial
- emotionAssociatedWithDisease
- foodCanCauseDisease
- agriculturalProductAttractsInsect
- arteryArisesFromArtery
- countryHasSportsFans
- bakedGoodServedWithBeverage
- beverageContainsProtein
- animalCanDevelopDisease
- beverageMadeFromBeverage

... but humans have veto power!