Watson

“An analytical computing system that specializes in natural human language and provides specific answers to complex questions at rapid speeds”

– I.B.M.

OHJ-2556 Artificial Intelligence
Guest lecturing Teemu J. Heinimäki
March 31, 2011
Outline

What is Watson?
History
Implementation and properties
How Does Watson Work?
Strengths and Weaknesses
Jeopardy!
Tactics of Watson
The IBM Challenge and its Significance
Future Views
Computer vs. Human Brain
The Essence of "Real AI"
What is Watson?

• An open domain QA machine by I.B.M.
  • "understands" questions in natural language
  • finds information in relevant sources
  • determines the confidence
  • responds with factual answers
• An instance of I.B.M.'s DeepQA technology
• Application of several fields, like machine learning, natural language processing, information retrieval, knowledge presentation, hypothesis generation...
• Combination of hardware and software
• Tackled Jeopardy!
• Near the Holy Grail of the AI?
History of Watson (1)

Achievements of latest decades:
- linguistic tools: synonym finders, rhyming dictionaries, classifiers for recognizing parts of the speech...
- QA systems for simple tasks

Latest 10 years:
- material became available online
  (blogs, wikis, newspapers, academic papers...)
- Computing resources → statistical document analysis became practical
1997: Deep Blue vs. Garry Kasparov
- Much publicity for I.B.M.
- However, not a marketable product → little direct income

Another high-profile project, the "grand challenge" (that would be applicable in the real world) was sought after. An advanced QA system seemed to be a good choice, while QAs had become important for different firms.
Possibility of competing with humans in Jeopardy! was suggested. First thought to be simply too tough a task (human language is hard, unlike chess=)

- Complex questions to be answered in seconds. The scope is very wide.
- Natural language: intended meanings, nuances, connotations, allusions, riddles, irony, ambiguities...
- Occasionally the answer cannot be found from any single source, but it must be synthesized based on several sources
2006: the I.B.M.'s most advanced QA system was tested with poor results.

2007: Dr. Ferrucci got 3–5 years and his team size was increased.

2008: Watson could theoretically beat some lesser Jeopardy! champions. Possibility of having a real Jeopardy! show aired in the national TV was inquired. Practice matches were organized (error analysis).


February 14–16, 2011: Two special Jeopardy! exhibition matches on TV. Watson against former champions Ken Jennings and Brad Rutter for $1M.
Implementation, properties

• Distinguishes itself by speed and memory capacity: 90 IBM Power 750 Linux servers, 2880 Power7 processor cores (3.55 GHz), running at 80 teraflops, 15–16 TB of RAM. (Deep Blue: about 1 teraflops)
• Code written in Java, C++ and Prolog. Components deployed and integrated using UIMA-AS (a standard asynchronous messaging middleware).
• Approximately 6 million logic rules to determine the best answer
• Knowledge base: about 200 million pages of documents of different types (4 TB)
• Besides natural language texts, also some structural and semistructural data (formal KBs, structured text)
How Does Watson Work? (1)

- Uses several well-known statistical methods
- More than 100 algorithms run simultaneously to analyze a question
- Another algorithm set rank the answer candidates e.g. by
  - inserting possible solutions into the original phrase and running the search again for support
  - cross-checking against time and space
- Machine learning framework used for weighting different algorithms and their results
- Also dynamic learning within categories: Verifying category interpretation: knowledge about knowledge
How Does Watson Work? (2)

Data (unstructured)

Pipeline of text analysis processes

Structure & meaning

Primary searches

Parallel analysis, evidence gathering, scoring

Question

NL processing techniques

What is looked for

Search results

300–500 answer candidates

Candidates with confidence values

ML

Merging scores
Strengts and weaknesses

• Knowledge acquired by the machine itself in reasonable time
• Good at brute force searches
• Applicability

• Cannot prove things (as opposed to knowledge-based systems)
• Not so good with short clues
• Must calculate confidences, not "instantaneous" knowledge about knowledge
Jeopardy!

- A popular American quiz show with answer-question format, adapted also internationally
- Three competitors
- Three rounds
  - Jeopardy!
    - Six categories, five clues per category, clues valued $200–$1000
    - One daily double ($5–highest dollar value / true daily double)
  - Double Jeopardy!
    - $400–$2000
    - Two daily doubles
  - Final Jeopardy!
    - Participated by the players with money
    - Single question, bet $0-all
- Normally the winner keeps the money earned and continues to the next game, second and third get consolation prizes
Playing Tactics of Watson (1)

• Hunts for the daily doubles, when opportunity to select the category and the clue
  • No need to worry about competitors being faster to buzz
  • Good chances to boost score for everyone
  • Uses statistical knowledge

• When no daily doubles anymore, selects the lowest clue value in a category with significant number of high valued ones
  • Aims to checks its understanding about the category with low risk before high stakes
Playing Tactics of Watson (2)

- Daily double wagering
  - In-category DD confidence
  - Game State Evaluator – a regression model estimating the winning chances at any stage
  - Computing the expected chance to win for every legal bet
  - Risk analytics also involved
Final Jeopardy! wagering

- Wagering in game one of two-game match resembles the DD wagering, except likely accuracy must be predicted based on the category title only.
- Wagering in single games and second games in matches
  - Positions of the players: predicting the opponents' bets. A library of known strategy rules, special additions for different situations.
  - Either uses a suitable rule or simulates different bets.
  - Considers also prizes for second and third places
The IBM Challenge Contest (1)

- February 14–16, 2011: A match of two special Jeopardy! exhibition games on TV. Watson against former champions Ken Jennings and Brad Rutter for $1M

- Watson represented by an avatar, synthesized voice and answer panel. Buzzing using a robot finger (delay about 8 ms).
First game:
- Jeopardy! round:
  - Watson and Rutter $5000, Jennings $2000
  - Watson repeated a wrong answer of Jennings
- Double Jeopardy! round:
  - Watson buzzed and answered 7 times correctly in a row ($21035)
  - Misunderstood an art period question, but dominated the game again soon
- Final Jeopardy
  - before: W $36681, R $5400, J $2400
  - Humans answered correctly, but W answered "What is Toronto??????" However, bet only $947
  - The game ended with W $35734, R $10400 & J $4800
The IBM Challenge Contest (3)

Second game:

- Jeopardy! round:
  - Humans successful. After this, Watson placed second. (W $4800, J $8600, R $2400)

- Double Jeopardy! round:
  - Wrong answer to DD
  - Humans still quite successful, but Watson took the lead anyway. (W $23440, J $18200, R $5600)

- Final Jeopardy!
  - All answered correctly. The second game ended with W $41413, J $19200, and R $11200.

The final result of the whole match: W $77147, J $24000, and R $21600
The Significance

- A good commercial
- A significant milestone in AI (?)
  - nothing new..?
  - ...but at least a nice compilation of existing techniques
  - a great demonstration of the capability of state-of-art AI
  - ”Human way” of processing language?
  - one step further towards the Star Trek computer
  - one step further towards passing the Turing test
- Possibly important applications
Future Views (1)

• Goal to start selling Watson versions to companies in 1–2 year
• Health diagnosis application I.B.M.'s primary focus.
  • Rapid growth of knowledge
  • Hard diagnoses
  • Medical Watson being built by I.B.M., Nuance Healthcare, Columbia University and University of Maryland.
• Goal to install Watson to every medical center's computational cloud (in U.S.A.)
Future Views (2)

• Virtual call centers
• Help desks
• Helping with bureaucracy
• Web self-service
• Contradiction engines
• Police, CIA & Co.
• etc.

• Price several M$ (needs to be run on at least $1 million I.B.M. server). The situation in ten years? In fifteen?
Future Views (3)

• Concerns:
  • Orders from Watson M.D.?
  • Unemployment?
  • Human resistance, ego
  • Computers finally taking over the world =)
Computer vs. Human Brain (1) (in Jeopardy! etc.)

- Computer–brain analogy in cognitive science
  - High parallelism – Watson simulating some parts of human language processing?
- Power efficiency: several kWs vs. 20 W
- Emotions – a lack of them an advantage
- Reliability of the memory
- Data analysis methods
  - Connecting relevant pieces of text logically
  - Understanding meanings and nuances of the language
- Reaction speed
- Buzzing behavior and confidence
- Nonverbal communication
Computer vs. Human Brain (2)
in Jeopardy! etc.

• Human knowledge representation not only in words – encoding text not enough
• Ability to answer questions with objective facts vs. answering to questions requiring judgment
The Essence of "Real AI"?

- Thinking/acting humanly/rationally
- Consciousness?
- "Understanding"?
- Fluid vs. crystallized intelligence
- Turing test
Thank you!