Foundations of AI

18. IJCAI or What is the Chinese Room?

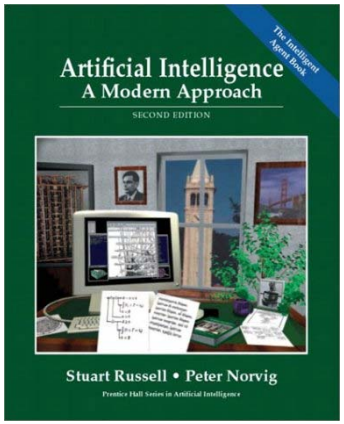
Wolfram Burgard, Andreas Karwath, Bernhard Nebel, and Martin Riedmiller

Contents

- The Publication Food Chain
- IJCAI and other outlets
- IJCAI 2009
  - How hard is it to manipulate an Election?
  - How convincing is Searl's *Chinese Room* argument?

Where do text books come from?

- Text book such as “AI: A Modern Approach” are not the product of the ingenuity of the authors alone
- They compile and structure a lot of individual research results



The publication food chain

- *Before*: Idea & solution & results
- Pre-Publication: Technical Report
  - no review
- First discussion: Workshop
  - review for plausibility (acceptance rate 95%)
- Presentation to peers: Scientific Conferences
  - strict but fast review (acc. 15-30%)
- Archival publication: Scientific Journal
  - strict review with multiple rounds (acc. 30%)

Note: not all stages necessary

Publication Outlets: AI Conferences

- International Joint Conference on Artificial Intelligence *IJCAI* (bi-annual, odd years)
- European Conference on Artificial Intelligence *ECAI* (bi-annual, even years)
- American National AI Conference *AAAI* (annual, except when IJCAI is in the US)
- German AI Conference

- … other conferences (e.g. application oriented)
- … specialized conferences (planning, learning, robotics, etc)

Publication Outlets: AI Journals

- *Artificial Intelligence* Journal
  - The most prestigious AI journal (focusing on formal approaches)
- Journal of Artificial Intelligence Research
  - Free online journal with high reputation and short turn-around times
- AI Communication
  - Journal by ECCAI
- ... other (usually) specialized AI journals

International Joint Conference on Artificial Intelligence

- Takes place in different locations (e.g., 2009: Pasadena, 2011: Barcelona, 2013: Bejing)
- Approx. 1000 attendees
- Approx. 1200 submitted papers, 300 accepted
- Proceedings as hardcopy, CD, and online (back to 1969)
- 6 day conference
- including workshops (20-30) and tutorials (10-20)
- costs around 600-700k US-$ each time
- 100k US-$ spent on travel grants for students

IJCAI 2009 - Talks

- 4 invited talks, 1 keynote
- 3 award talks (Computer & Thought, Research Excellence)
- Technical papers (332):
  - Agent-based & multiagent systems 55
  - Constraints, satisfiability, search 43
  - Knowledge representation, reasoning, logic 51
  - Machine learning 66
  - Multidisciplinary & applications 20
  - Natural language processing 20
  - Planning & Scheduling 30
  - Robotics & Vision 11
  - Uncertainty in AI 18
  - Web & knowledge-based information systems 16

IJCAI 2009 – Freiburg

- 5 technical papers (1.5%)
  - *Qualitative CSP, Finite CSP, and SAT: Comparing Methods for Qualitative Constraint-based Reasoning* (Matthias Westphal, Stefan Wölfl)
  - *On Combinations of Binary Qualitative Constraint Calculi* (Stefan Wölfl, Matthias Westphal)
  - *A Fixed-Parameter Tractable Algorithm for Spatio-Temporal Calendar Management* (Bernhard Nebel, Jochen Renz)
  - *Eliciting Honest Reputation Feedback in a Markov Setting* (Jens Witkowski)
  - *Learning Kinematic Models for Articulated Objects* (Jürgen Sturm, Vijay Pradeep, Cyrill Stachniss, Christian Plagemann, Kurt Konolige, Wolfram Burgard)
- 1 Award
  - IJCAI/JAIR Best Paper / Honorable Mention: Malte Helmert

2 selected papers

- *Where Are the Really Hard Manipulation Problems? The Phase Transition in Manipulating the Veto Rule* (Toby Walsh)
  - Analyzing the claim that NP-hardness is a tool to prevent strategic manipulation in elections from an empirical point of view.
- *Is It Enough to Get the Behavior Right?* (Hector J. Levesque)
  - The Chinese Room argument, which says that *strong AI* is impossible because AI systems can only fake intelligent behavior, is challenged. The only paper with a philosophical touch at IJCAI 2009.

Elections and Social Choice

- Social Choice Theory:
  - Given a set of candidates, and a set of voters with preferences over the candidates, a social choice function (election rule) should return the most preferred candidate
- Subarea of Game Theory
- Interesting for preference aggregation (e.g. in CSPs), in coordination (e.g. in MAS), and in electronic communities and markets

Example: Choosing a lecturer for next semester

- Voting:
  - 10 students: Karwarth > Nebel > Burgard
  - 7 students: Nebel > Burgard > Karwarth
  - 15 students: Burgard > Nebel > Karwarth
  - 6 students: Nebel > Karwarth > Burgard
- Which one should do it?
- Many possibilities (sometimes ignoring parts of the preferences):
  - Plurality
  - Veto
  - Borda count
  - ...

Manipulation

- A social choice function (or election scheme) can be manipulated if by stating preferences insincerely, one can get a more favorable outcome (as an individual or group)
- Example:
  - For plurality, it can make more sense to state the second choice as the most preferably one, if one owns candidate would not get enough votes
- If a social choice function is immune to manipulation, one calls it “incentive compatible”

The Gibbard-Satterthwaite impossibility result

- Gibbard and Satterthwaite proved that any social choice function that
  - handles more than 2 candidates,
  - is surjective (allows all candidates to win), and
  - is incentive compatible

will also be

- a dictatorial choice function (only one voter decides)!

NP-hardness as a tool against manipulation

- All social choice function (election schemes) can be manipulated (Gibbard/Satterthwaite)
- However, it might be computationally hard to decide *whether* and *how* this could be done!
- For some election schemes, it can be proven that manipulation is NP-hard (for some, winner determination is actually NP-hard!)
- So here, NP-hardness is a GOOD thing!

- Since it is a worst-case notion, the question is, whether it appears in practice

Manipulating elections according to the veto rule is NP-hard

- Destructive manipulation (avoiding a candidate) is actually easy (polynomial time)
- Constructive manipulation is NP-hard
- However, as shown in the paper, only for very few cases one gets a computationally hard *phase transition*
- Throwing in another random voter makes everything easy again
- For veto voting, the theoretical worst-case result seems to mostly irrelevant.
- What about other election schemes?

Intelligence, Behavior, Philosophy …

- Most papers at AI conference are about technical results (methods, algorithms, empirical results …)

- This paper takes up an issue from the 80‘s voiced by the philosopher Searl, who states that strong AI is impossible

What is Intelligence?

- Turing:
  - Hard to tell
  - Let's call a machine intelligent if it behaves intelligently
  - **Turing test**: If the (linguistic) behavior is indistinguishable from the human behavior over a long time, then a machine passes the test
  - Be careful with partial satisfaction of the test, which can very easily achieved by trickery!

What is Intelligence?

- Searl:
  - Whatever intelligence is, it cannot be achieved by a machine!
  - Machines might be able to simulate (fake) intelligent behavior, but it is not acting because of (real) intelligence
  - So, AI is doomed to failure – if AI is understood in the *strong sense*, namely, if we want to make machines intelligent (as humans are)
  - In AI research we do not care much about Searl's argument … nevertheless …

The *Chinese Room* argument

- Let's assume, AI has succeeded in creating a system that perfectly understands and generates Chinese sentences: `chinese.py`
- Instead of running this program, we could put Searl and `chinese.py` in a room, and Searl could process the inputs and generates outputs according to the rules of `chinese.py`

- It is obvious that Searl does not understand Chinese at all, while an outside observer would think the system understands Chinese (according to the Turing test)

Chinese Room: The System Reply

- Of course, Searl does not understand Chinese
- But the system consisting of Searl and the book `chinese.py` (CPU+program) understands Chinese!

- Searl's reply:
  - Assume I read and memorize the book `chinese.py` and then throw it away.
  - After that, I process the inputs and generate outputs as before

  ➢ I still do not understand Chinese!

Type I and Type II books

- Implicit in Searl's reply is that there two types of books or programs:
  - **Type I**: You can memorize, but you do not understand Chinese afterwards
  - **Type II**: After you have memorized them, you understand C[illegible]e (e.g., as a second language)

Can there be Type I books?

- While understanding Chinese as a second language (using a Type II book) is not interesting from an AI point of view, there are probably also Type II books using programming languages
- The question is, if there can be Type I books for the Chinese room at all
- Hard to tell
- Let's simplify this and consider the Summation Room

The Summation Room

- An input is a list of 20 ten-digit numbers
- The required output is the sum
- Assume a book/program `sum20.py`
- Could be a lookup table
  - Type I book
- But a lookup table is too large: 10^{200}
- There are only 10^{100} atoms in the universe

Other books for the Summation Room

- One could write a program performing addition based on a 10x10 single digit addition table
  - This would be a Type II book!
  - Having memorized it, one really does summation and knows what one does (even when the name for the operation might be unknown)
- Even all other “small” books would implement *addition* as such (e.g. base 100 addition or parallel addition)
- There is no Type I book for the Summation Room

Summary

- Searl's *Chinese Room* argument suggest that AI can only simulate intelligent behavior
- This is based on a thought experiment, where a human memorizes a rule body and executing it, without understanding it
- Difficult to make precise for Chinese language processing
- More obvious for the *Summation Room*
- However, here it is impossible to memorize a (small) rule set without doing (real) summation when executing the rules
- So Searl's answer to the *System reply* is not convincing

Conclusion

- The interesting stuff is happening at scientific conferences (not in the text book)
- Try to read such papers (e.g. go to ijcai.org)
- For a Bachelor thesis in AI, you may want to aim to publish it at the German AI conference
- For a Master thesis, you may want to go for AAAI, ECAI or IJCAI
- But for now, you may want to relax (in the next few weeks)