AI in Healthcare

Limits and potentials

Paul Attie
Department of Computer Science
American University of Beirut

History of AI and Machine Learning

- Early definitions of AI
  - “Programming computers to do things that require intelligence in humans” (AI – Elaine Rich)
  - “Make computers more useful and understand the principles of intelligence” (AI – Patrick Winston)
- Some early research areas
  - Play games (chequers, chess)
  - Recognize images (blocks, etc)
  - Understand natural language
  - Planning: devise a sequence of actions that achieves a goal, subject to constraints

Expert systems

- MYCIN:
  - Rules for inexact reasoning
  - Each rule requires several observations, and contributes either a degree of *belief* or a degree of *disbelief*
  - Final verdict is the cumulative result of all applicable rules

Limits of hardcoded approach to AI

- “Hardcoded” AI hit limitations
- Good at solving easily formalized problems, e.g., play chess
- Not so good at solving informal problems that require “human intuition”, e.g., conduct a conversation

Machine learning

- Acquire knowledge, by extracting patterns from data
  - Learn the mapping from representation to output
- Performance depends on representation
  - Arithmetic in roman numerals is hard, in Arabic numerals easy
  - Learn the representation itself
- Deep learning
  - Introduce representations in terms of simpler representations
  - Build complex concepts out of simpler concepts

Deep learning

- Feedforward deep network
  - A function mapping inputs to outputs
  - Formed by composing many simpler functions
  - Sequence of one input layer, several hidden layers, and one output layer
  - Each layer takes input from previous layer, and gives output to next layer
- Very good at “pattern matching”
- But, can be easily fooled…..

Deep learning weaknesses

- Can be fooled by intentionally inserted “noise”
- Fooled two classifiers into misreading STOP signs [1]
  - Attack was to place stickers on the stop sign!
- Fooled face recognition software by altering images in ways imperceptible to humans, and by using “inconspicuous” accessories [2]
- None of these attacks are remotely close to fooling a human
- Conclusion:
  - *What deep learning does and what the human brain does are very different!!!*

Risks

- “No one knows exactly how neural networks work” [3]
- Developed software that observes neural networks “in reverse”
- Two neural nets that recognize horse photos
  - One recognizes horses bodies
  - The other recognizes copyright symbols!!!
    - Works since copyright symbols correlated with horse forums

So what's ML in Medicine good for?

- Intelligent query processing
  - Doctor asks for info/guidance, based on ongoing conversation with patient
  - Pro: can search through huge data sets quickly
- Intelligent real-time interactive assistant
  - ML avatar listens to doctor-patient conversation
  - Makes suggestions autonomously
  - Narrative medicine provides input to the avatar
- ML must justify its decisions
  - Harder as application and classifiers get more complex
  - Existing work on explainable AI
  - Not specific to medicine

Justification of decisions

- Want a short justification
  - Cf. short *certificates* in complexity theory
- Can also be achieved by a short interaction with the ML
  - Cf. *interactive proofs* in complexity theory
    - Interaction between “prover” and “verifier”

Conclusion

- Saves much time for physician if:
  - Interaction with the ML takes much less time than solving the problem manually in the first place
- Still not reliable enough to be used without human checking: you just don’t know what the ML classifier is matching against!

References

- [1] Robust Physical-World Attacks on Deep Learning Visual Classification, Eykholt et. al., *Vision and Pattern Recognition*, June 2018, Salt Lake City, Utah
- [2] Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition, Sharif et. al., *23rd ACM Conference on Computer and Communications Security*, October 2016, Vienna, Austria
- [3] Fraunhofer-Gesellschaft, Press release, 1 February 2017, https://www.fraunhofer.de/en/press/research-news/2017/february/watching-computers-think.html