Explaining Explainable AI

S Panda & ST Roy, IIT Kanpur

Explainable AI (XAI)

- The idea of elucidating, clarifying, interpreting AI systems
- AI systems, though very effective, are essentially black boxes
- The need to explain arises for transparency, accountability, more effectiveness among others
- Explanation requires tools, frameworks or mechanisms

Visualization for XAI

- Visualization is one such tool for explanation in XAI
- Visualization is the act of building a visual representation.
- Visualization can be used to,
  - Visually explain & interpret the algorithm
  - Display the live-performance of AI systems
  - Clarify the results

On Visualization

- From visualization research it seems that visualization in and of itself contains many complexities, ambiguities and pitfalls.
  - Cairo's *How Charts Lie*
  - Munzner's *Visualization Analysis and Design*
- Thus there is a need to explain visualization itself.
- Can potentially distort the explanation & cause a recursion.
- This argument can be extended to other tools.

Resolution I

- Using tools, devices that have existing or a minimal explanation.
- For instance, the use of visualization techniques that have been studied intensively in the literature.
- Can explain rudimentary AI systems but not conducive for sophisticated or domain-intense systems

Resolution II

- Inculcate uncertainty into the XAI devices or mechanisms
- The uncertainty is the result of the device explanation; the ambiguities, complexities then form a part of the explanation.
- Can be a qualitative/quantitative metric; as an analogy - Heisenberg's Uncertainty Principle.
- Can be used for sophisticated AI systems.

Resolution III

- Make the user AI-literate
- About finding the balance between the AI system and the user.
- XAI system includes the AI, the explanation & the user.
- User literacy also brings in accountability & ethics.
- Useful in high-stakes, high-risk domains (eg. Aerospace, Medicine)

Conclusion

- XAI requires tools, mechanism for explanation
- Visualization is one such tool.
- Visualization has its own pitfalls which can potentially distort the explanation.
- Three ways to resolve this conundrum.

Thanks