How AI is Screwed up by (using only) Deep Learning

From an academic perspective

Piyawat Lertvittayakumjorn

2nd January 2019
What is Deep Learning (DL)?
Deep learning is awesome, actually.

- Requires no feature engineering
- Has sufficient parameters for known knowns, known unknowns, unknown unknowns
AI can reproduce Mona Lisa in the styles of Picasso, van Gogh, and Monet. Source: Gene Kogan

Music Transformer: Generating Music with Long-Term Structure

Dec 13, 2018 https://magenta.tensorflow.org/music-transformer
BUT
(1) DL is brittle and spurious

- Object Recognition
(1) DL is brittle and spurious

- Reading Comprehension

Paragraph: "In January 1880, two of Tesla’s uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses."

Question: "What city did Tesla move to in 1880?"
Answer: Prague

80% accuracy

Paragraph: "In January 1880, two of Tesla’s uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses. Tadakatsu moved to the city of Chicago in 1881."

Question: "What city did Tesla move to in 1880?"
Answer: Chicago

34.2% accuracy
(1) DL is brittle and spurious

- Visual Question Answering

<table>
<thead>
<tr>
<th>question</th>
<th>answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many...?</td>
<td>2</td>
</tr>
<tr>
<td>Is/Are...?</td>
<td>Yes</td>
</tr>
<tr>
<td>What sport...?</td>
<td>Tennis</td>
</tr>
<tr>
<td>What animal...?</td>
<td>Dog</td>
</tr>
</tbody>
</table>
(2) DL is data-crazy

• It doesn’t care about general knowledge of the domains of interest
  • Intuitive Physics
  • Linguistic Phenomena
  • Psychology
  • …
(2) DL is data-crazy

- It doesn’t have any commonsense knowledge

Figure 1. Julia Child’s kitchen. Photograph by Matthew Bisanz.
(2) DL is data-crazy

• It neglects other AI techniques outside machine learning

Branches of AI

• Logical AI
• Search
• Pattern recognition
• Representation
• Inference
• Common sense knowledge and reasoning

• Learning from experience
• Planning
• Epistemology
• Ontology
• Heuristics
• Genetic programming
(2) DL is data-crazy

- It neglects other AI techniques outside machine learning

### Branches of AI

<table>
<thead>
<tr>
<th>Logical AI</th>
<th>Learning from experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td></td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>Planning</td>
</tr>
<tr>
<td>Representation</td>
<td>Epistemology</td>
</tr>
<tr>
<td>Inference</td>
<td>Ontology</td>
</tr>
<tr>
<td>Common sense knowledge and reasoning</td>
<td>Heuristics</td>
</tr>
<tr>
<td></td>
<td>Genetic programming</td>
</tr>
</tbody>
</table>
(3) DL is unexplainable
(3) DL is unexplainable

We do not know:
- What is encoded in a vector
- What does the model capture
- Why does the model predict this answer
(3) DL is unexplainable

- But we need a **right to explanation**
  - A right to be given an explanation for an algorithm’s output

The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention.

...In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, **to obtain an explanation of the decision reached after such assessment and to challenge the decision.**

The European Union General Data Protection Regulation (enacted 2016, taking effect 2018)
Related Research Trends 2018–2019

• More challenging AI/DL evaluation
• Using external/commonsense knowledge in DL
• Adding more inductive biases to DL
• Fusing model-based and function-based AI approaches
• Explainable (and effective) AI
• Human-AI argumentation
Image sources:

- https://sigmoidal.io/dl-computer-vision-beyond-classification/
- https://rajpurkar.github.io/SQuAD-explorer/
- https://medium.com/enabled-innovation/artificial-general-intelligence-too-much-or-too-little-too-soon-9c0dd7bd1c2d
- http://karpathy.github.io/2015/03/30/breaking-convnets/
- https://www.slideshare.net/butest/cs-561a-introduction-to-artificial-intelligence