Kuncheng Feng
CSC 466

Chapter 6: A Closer Look at Machines That Learn

→ Reading/Mining/Discussion Assignment

1). TRUE/FALSE - The learning-from-data approach of deep neural networks has generally proved to be more successful than the “good old-fashioned AI” strategy, in which human programmers construct explicit rules for intelligent behavior. However, contrary to what some media have reported, the learning process of ConvNets is not very human-like.

True

2). Why does your professor like the previous question?

The more in depth answer to the previous question shows the overhead for implementing a machine that will learn on its own, as the setups for a learning from data approach is still quite similar to symbolic AI, where lots of human efforts are needed to define what is to be learned, and sort out the data that AI learns from.

3). TRUE/FALSE - As we’ve seen, the most successful ConvNets learn via a supervised-learning procedure: they gradually change their weights as they process the examples in the training set again and again, over many epochs (that is, many passes through the training set), learning to classify each input as one of a fixed set of possible output categories.

True

4). List some significant differences between the way that humans learn about objects and the way that ConvNets learn about objects.

Machines can only classify inputs into a fixed set of possible outputs, and it needs lots of iterations to assign the correct weights to get it right. In contrast humans can learn an open-ended set of questions and give out an answer, quickly recognize a new category after only seeing a few examples. The biggest difference is that humans can learn actively by asking questions.

5). Why is it inaccurate to say that today's successful ConvNets "learn on their own?"

A huge amount of human effort is required to collect and classify the data that the machine is going to use to learn, the machine's learning architecture and methods also requires a lot of setup by humans. (Tuning the hyperparameters).

6). In answer to the rhetorical question "Where does all of the data come from to fuel big data applications?," MM answers "You - and probably everyone you know." Please elaborate on the answer.

The data come from the content that the users uploaded, be it image, video, voice, comments. They all have been fueled into the data set that trains those AIs.

7). How do car companies acquire the big data (labeled images of pedestrians, cyclists and other obstacles) needed to train robo-cars?

Some test drivers mount a camera on their car while driving, or in the case of Tesla every car is sharing their driving videos by default. And then they employ many workers to label things in the video, often frame by frame.

8). What is the "long tail" phenomenon, and how does it relate to machines that learn (ConvNets)?

For an AI, there is a statistical distribution of the events that it is likely to encounter when performing a task, like seeing a red light or encountering a pedestrian while driving. However at the end of the distribution there are countless events with every low possibility to occur (long tail), so these situations are not common in the training set, as a result the AI will have a difficult time recognizing them.

9). TRUE/FALSE - A commonly proposed solution to the long tail problem in AI systems is to complement supervised learning with unsupervised learning.

True

10). What is "unsupervised learning?"

Learning categories or actions without labeled data.

11). What colorful remark did Yann LeCun make about unsupervised learning?

"Unsupervised learning is the dark matter of AI".

12). TRUE/FALSE - For general AI, almost all learning will have to be unsupervised, but no one has yet come up with the kinds of algorithms needed to perform successful unsupervised learning.

TRUE

13). TRUE/FALSE - Humans have a fundamental competence lacking in current AI systems: common sense. We have vast background knowledge of the world, both its physical and social aspects. We have a good sense of how objects - both animate and living - are likely to behave, and we use this knowledge extensively in making decisions about how to act in any given situation.

TRUE

14). TRUE/FALSE - Many people believe that until AI systems have common sense as humans do, we won't be able to trust them to be fully autonomous in complex real-world situations.

TRUE

15). TRUE/FALSE - Superficial changes to images, such as slightly blurring or speckling an image, changing some colors, or rotating objects in the scene, can cause ConvNets to make significant errors even when these perturbations don't affect humans' recognition of objects. This unexpected fragility of ConvNets – even those that have been said to "surpass humans at object recognition'" – indicates that they are overfitting their training data and learning something different from what we are trying to teach them, a phenomenon that results in various manifestations of unreliability.

True

16). The unreliability of ConvNets can result in embarrassing – and potentially damaging – errors. Select a particularly embarrassing/damaging example of unreliability in ConvNets, and describe it in just a sentence or two.

Google's AI tagged two African Americans as Gorillas, camera software not detecting dark skinned faces or labeling Asians as blinking.

17). At the end of the section on biased AI, MM observes that the problem of bias in applications of AI has been getting a lot of attention recently, with many articles, workshops, and even academic research institutes devoted to this topic. What questions does she raise in conjunction with this observation? What do you think are the appropriate answers to these questions?

Question was - Should the data set being used to train AI accurately mirror our own biased society, or should they be tinkered with specifically to achieve social reform?

I think it's hard enough to develop AI as it is, let's first get it working then we adjust it to fit our societal norms.

18). TRUE/FALSE - You can often trust that people know what they are doing if they can explain to you how they arrived at an answer or a decision. However, "showing their work" is something that deep neural networks – the bedrock of AI systems – cannot easily do.

True

19). TRUE/FALSE - Recall that a convolutional neural network decides what object is contained in an input image by performing a sequence of mathematical operations (convolutions) propagated through many layers. For a reasonably sized network, these can amount to billions of arithmetic operations. While it would be easy to program the computer to print out a list of all the additions and multiplications performed by a network for a given input, such a list would give us humans zero insight into how the network arrived at its answer. A list of a billion operations is not an explanation that a human can understand.

True

20). What, according to MIT's Technology Review, is the dark secret at the heart of AI?

Even human AI experts cannot tell how they arrive at their answers.

21). What does the phrase "theory of mind" refer to, and how is it related to our interactions with AI systems such as deep networks?

We humans cannot look into the mind of others either, but we trust those who have mastered basic cognitive tasks. Like object recognition and language comprehension.

22). One of the hottest new areas of AI is variously called "explainable AI," "transparent AI," or "interpretable machine learning." To what do these terms refer?

AI that can explain how they arrived at their answers.

23). The field of “adversarial learning” has emerged in response to the fact that AI systems can readily be fooled in dramatic fashion, like mixing up a guy in glasses with Milla Jovovich, or misclassifying a stop sign for a speed-limit sign. Briefly describe the field of adversarial learning.

Developing strategies that defend against potential (human) adversaries who could attack machine-learning systems.

24). Jeff Clune, an AI researcher at the University of Wyoming, made a very provocative analogy when he noted that there is “a lot of interest in whether Deep Learning is ’real intelligence’ or a ’Clever Hans.’” Explain the essential question that underlies this analogy, being sure to incorporate a few words on the actual Clever Hans.

Clever Hans was a horse in Germany that would use the cues from the questionnaire to arrive at a correct answer, maybe the AI is also using unintentional cues to arrive at the right answer, instead of actually understanding and knowing the question.