From Neural Networks to Reasoning Machines

Behdad Forghani

One of the most exciting developments in the last decade has been the advances in neural networks and intelligent machines. The promise of neural networks has been machines that learn without being explicitly programmed and therefore solve the messy process of software development. Moreover, for some tasks like image recognition, there are no better algorithms that can perform better than neural networks. The current state of neural networks can be summarized as the following.

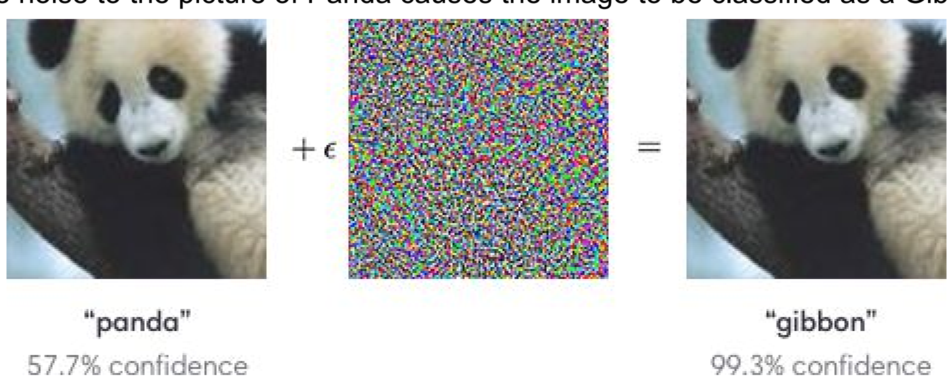
State of Machine Learning

The good

Deep neural networks are the best solution available for image, speech and language recognition.
The most labor intensive part of training supervised machine learning algorithms is labeling the training set. This task of labeling is much cheaper and more predictable than software development and the quality of the labels can be verified more easily. There are also some use cases that unsupervised models can be used and therefore no labelling is necessary.

The bad

Neural networks are very fragile. This can be demonstrated by Panda/Gibbon example, where adding some noise to the picture of Panda causes the image to be classified as a Gibbon



instead of a Panda.
Training neural networks takes a lot of iterations and samples. This makes the process very time consuming and resource intensive. A child can pick up the concept of an elephant with just

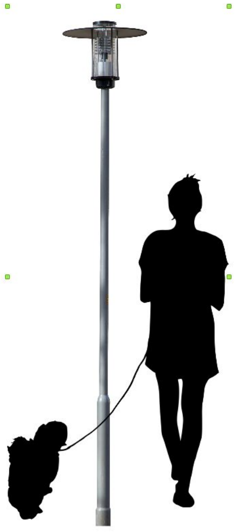
one example. Current machine learning techniques are nowhere that efficient for the reasons that will be explained.

The ugly

Current machine learning implementations perform very poorly in novel situations and they must explicitly be trained for every setting and context. This makes them very unreliable when they encounter situations that they were not trained for. By contrast, once humans learn how to drive a car, even if they have only driven in the cities, they can drive in mountain roads and beaches.

Machinery of Reasoning

Comparison between the reasoning of dogs and humans provides a good insight into the machinery involved in reasoning.
Dogs are very smart and they are nature's statisticians. They do not run into solid objects, even if they have not encountered the object and the settings before. By not colliding in objects, we can conclude that dogs create mental models of the object in their environment and they identify solid objects. However, dogs cannot project the position of external objects like human mental machinery does. As a result, if you are walking with a dog on a leash, and a lamp post is between you and the dog, the dog cannot predict that the leash will be caught in the lamp post and this will prevent the two of you from moving further.



Humans are not naturally as good a statisticians as dogs, on the other hand, their mental models are more sophisticated. They have the model of the leash, the dog, the lamp post, the connection of the leash and the direction all of the above are travelling. They can automatically and effortlessly project the future position of the leash and foresee the problem of leash getting tangled with the lamp post. No amount of training can teach dogs about leash not getting caught in obstacles, because dogs are missing the machinery to projects the objects modeled by their consciousness into the future.

Perception and Reasoning

Perception

The first element of reasoning and consciousness is the creation of mental models of the world using the input to our senses. This is very similar to the concept of a holodeck.

When we perceive an object we create a duplicate of it in our mind. This is done in many dimensions. For example, in addition to the geometry and the position of the objects, our mental model can contain attributes like solid/liquid, flexible/rigid, movement/speed and temperature.

Reasoning

Once our consciousness has created mental models of the objects in our surroundings, it reasons by applying rules like the laws of physics on these mental models.
For example, we have a rule that two solid objects cannot get past each other and they collide, but solid objects can get into liquids. Similarly a dog knows that it cannot get past a solid object like a wall and will avoid colliding with stationary objects, like walls or any stationary solid object like a statue even if it has never encountered that solid object before.
Humans also understand the concept of momentum and the fact that objects in motion will continue to travel in a straight line unless another force is acted on those objects and they are much better equipped in avoiding collision with moving solid objects like cars.
Human drivers project the position of their cars and moving solid objects like other cars, bicycles and pedestrians and they can detect if their car is in a collision course with other objects. Then they try to solve avoiding collisions by using different alternatives in their skillset like changing the speed of their car or swerving and changing the direction of their movement.
A reasoning machine that identifies all solid objects, determines their momentum can detect collisions reliably without the need to classify the objects in its surroundings. It can also use rules to avoid collisions. We have seen self driving cars failing to notice the bed of the truck because it was higher off the ground than it was trained for.

The important points here are that there is an explicit machinery in the brain of humans that allows for this projection to the future. This is important since the two brains are made of similar neurons, but there is an additional feature in the brain of humans to make this projection and dogs cannot acquire this skill with any amount of training. The second important point is that this manipulation of mental models is automatic and effortless for humans. The effortlessness of this ability points out to its neural network and not algorithmic nature. The algorithmic nature is what allows humans to do tasks like multiplication of multiple digit numbers, where, the task is learnt and not automatic.

The role of reasoning in language processing

In addition to the applications like self driving cars, reasoning using laws and mental models can help with language processing. Consider the two sentences:
The pen is in the box.

And the box is in the pen.

Humans can derive the correct meaning of the pen using their mental models and the rule that a solid object can be inside an empty space.

In the case of the first sentence “The pen is in the box”, the reasoning is done using the mental images of pen and box and some simple laws that pass the tests:
- Writing pen is a solid object
- Box has empty space
- Pen can get into the box, because solid objects fit in empty spaces.

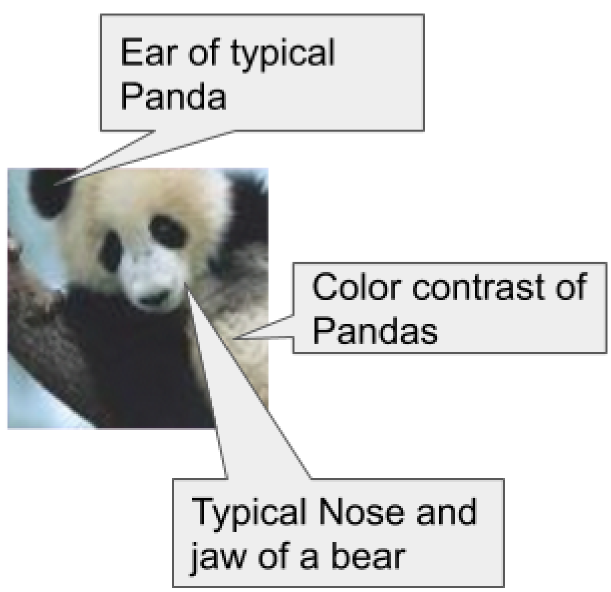
In the case of the second sentence, the above tests do not pass, i.e., a box cannot get inside a writing pen, therefore, the pen must be an animal coup. And these tests pass:

- Box is inside a space
- Animal pen has space
- Box can get into an animal pen

To demonstrate the current state of machine learning has not reached this level of reasoning, try doing a search for “box in pen”. You will not get the correct example in any search engine.

The Role of Reasoning in Object Classification

Now let's go back to the example where traditional machine learning for object classification got fooled when we added some noise to the picture of a Panda. The reason humans do not make mistakes so easily is that we break the objects that we see into its parts. Then we try to match these objects with the models that we have made in our minds to represent those objects.
Our mental models for classes of objects are created by grouping objects in classes using their similarity this quickly narrows down our search. Next we use the salient features of each separate object to classify them. For example, we have models for groups like apes, mammals, birds, etc. We also know that the salient features of bears are they paws, and stalkiness. A child can effortlessly extract the salient features of the elephant trunk and fanned ears and with this identify elephants with one sample. Now, when we see a picture of a Panda bear, we break the picture into parts and classify it as a mammal. Next we look at the parts of the pictures and we see that it matches the salient features of a Panda bear.



Summary

Approaching AI by modeling the objects in the world and applying the laws of physics will result in better AI that can function in novel situations. This is complementary to current simple neural network approaches. For example, a traditional neural network approach is much better in generating correct grammar. But, a reasoning approach where the machine tries to create a coherent and reasonable model of the sentence can resolve sentences like “the box is in the pen” better. This approach handles sentences that it has not encountered and can translate it to a model representing the meaning of the sentence.

A robot that can model flexibility of cables and the rigidness of plugs and the space and clearances of sockets can much more easily be trained to plugin a USB plug into a USB port. Similarly, a self driving car modeling the movement of solid objects without the need to classify them will be able to avoid collisions with a raised bed of a truck with high clearance from the road much more reliably even if it has not successfully classified the trailer or bed of the truck.

The above examples were taken from situations where the current state of machine learning is struggling and it is not well suited for and the reasoning method suggested will be a good complement to resolve the current shortcomings of neural networks.