AI Intro

Alex S.[*]

1 AI

AI is intelligence exhibited by non-living things. A more precise definition is problematic, since intelligence itself is poorly defined. What we can say about intelligence is that *we will recognize it when we see it*.

What aspects of intelligence will we recognize? Intelligent beings must act or respond rationally. That means recognizing patterns in their inputs, correctly interpreting those patterns, learning from them, and producing output that maximizes some utility. A mechanism would not be considered intelligent if it failed to learn from experience or acted in ways detrimental to itself.

Is pattern recognition intelligence? No, but it is a vital part of it. The same goes for data manipulation, learning, reasoning, planning, knowledge representation, etc.

2 Finding a model to mimic observations.

Learning can be viewed as mimicry. Every process generates data. The learner makes observations, which are then used to model the process that generated those observations. If the learner is successful, the model will *mimic* some aspects of those observations. There is hope (not requirement) that this model also mimics the functionality of the generating process—this model is our mind's eye view of how the generating process works, and is used in place of the process to make predictions.

For example, Newton observed objects falling down, and measured their speed. He then constructed a mathematical model that provides a good approximation of nature and produces results that appear to agree with everyday observations. With this model, Newton was able to predict how objects will fall on the moon, or how planets go around the sun. The model (Newtonian Physics) is obviously not representative of how things *really* work—it just makes useful predictions for most situations humans encounter. This is what learning is all about—to observe things, and then imagine a generalized model that is useful in some way.

Automation of science is one of AI's goals.

There are essentially three ways to learn: *memorization*, *deduction*, and *induction*. With memorization, the learner simply memorizes (stores in a database or file) all facts and observations. These facts can then be recalled, or matched to input. A rote memorizer cannot

[*] alex@theparticle.com

cope with inputs that were not previously observed, nor can any predictions be drawn from the data.

With deduction, the learner gets predictive power; starting with knowledge that the learner knows to be *true* (facts, observations, or previously proved items), the learner attempts to derive other truths by applying laws of logic and math. A classic example, given the facts: *"All men are mortal"* and *"Socrates is a man"*, the learner may *deduce* that *"Socrates is mortal"*. One generally ignores that axioms of logic and math are not themselves rigorously proven.

Induction is what we are interested in for intelligent beings; it is deduction with assumptions. The inductive learner makes observations (usually input/output pairs) and assumes (guesses) that they were generated by some *process*, e.g., a normally distributed random variable. Then working backwards, the learner uses these observations to construct a model of this process that mimics the observations in some way, e.g., the learner may construct a stochastic process that has the same *mean* and *variance* as the observations. This model is the 'learned knowledge', and can be used in place of the process to make predictions.

The amorphous concept of *model* can be anything at all: our brain is a model, our computer is a model, etc. Most general computational models we can imagine are not (yet) practical for systematic learning. If our model is a Turing machine, then having it systematically 'learn' non-trivial things would be equivalent to having the computer program itself! Needless to say, that is beyond the current state of the art.

Over the years, practitioners in the field have come up with quite a few models suitable for systematic usage. Many practical models are parametric: we assume some mathematical construct, such as a function or distribution; learning then involves adjusting the model parameters.

For example, imagine we observe a few dozen two-dimensional points. Let us then *assume* that these observations were generated by a *linear process*, with normally distributed error. We can then use a *linear model*, such as a line, to mimic the generating *linear process*.

Our inductive assumption in this case is that the training set roughly fits a line, such as $f(x, y) = Ax + By + D$. The model parameters in this case are A, B, and D. Inputs are x and y. The goal is to *optimize* (by adjusting A, B, and D) this model to better fit all the training samples.

Once we find the best fitting A, B, and D, (and by extension $f(x, y) = Ax + By + D$) we can make predictions about the input. We can plug in any (x, y) into that function and find whether it is on the line or not (whether it matches the original inputs). We can even solve for y given any x, and in cases where x represents time, we can make predictions about the future.

The big philosophical question is whether inductive reasoning is correct. What rational basis do we have to generalize a few observed points into a line, or assume that future observations will follow the same pattern as before? For all we know, the points could be random, and the seemingly linear pattern is a result of chance. This is unavoidable. We must always suspect that our models may be wrong, are based on past observations, and that the very relevant future may turn out to be something completely unexpected.