AUTONOMOUS LEARNING: WHAT IT MEANS

Mykola Rabchevskiy

AD HOC LEARNING CAN'T BE REDUCED TO STATISTICS

The term "learning" is one of the most commonly used terms in the AI/AGI field. At the same time, the interpretation of the concept in terms of the result is quite clear: if something was unknown and then became known and can be used if necessary, a person or an AI system has learned this. Naturally, when it comes to AI, the question immediately arises of ***how new knowledge is obtaine***d. Accordingly, there is a particular classification system for ways to acquire new knowledge, and the natural first step is to distinguish ***autonomous learning*** from ***knowledge transfer***. Autonomous learning means that new knowledge is obtained by the learning system ***without any participation of another subject who already has this knowledge*** (and, in one way or another, helps the learning system acquire it). The critical sign of ***knowledge transfer*** is the presence of a specific channel for ***transmitting information from a teacher*** to an AI system. External help can be both extremely explicit - telling the trainable system of rules or algorithms - and less straightforward when, as in the case of neural network training (a person prepares a training dataset containing the required knowledge). The least apparent variants of knowledge transfer are ***reinforcement learning***, when the teacher's participation is minimized, and ***learning by imitating*** the actions of someone with the required knowledge.

Autonomous learning implies the absence of another knowledge carrier when an autonomously learning system finds the required knowledge by observing the situation without any instructor. Here, a division immediately arises into ***passive learning***, when the system being trained does not take any actions ***explicitly aimed at discovering*** the unknown (that is, it does not experiment), and ***active (exploratory) learning***.

AGI engineering is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.

Naturally, the question arises as to ***whether an AI-controlled technical system requires autonomous learning***. Let's analyze this using the example of an autonomous car driver.

For such an autopilot, there is no need to ***modify*** the system's ***behavior*** during operation; moreover, a behavior change may be considered undesirable for safety reasons (if behavior can change, this may mean a more dangerous driving style). Naturally, a hypothesis arises about the uselessness of autonomous learning for autopilots.

However, analysis of the facts does not confirm this hypothesis. No matter how broad the knowledge embedded in one or another way in the AI system, there are often enough situations when the ***current scene includes objects unknown to the autopilot***. It can be an animal:

an airplane:

a robot

and so on. Of course, it is ***reasonable to require the autopilot to be able to recognize the situation and take appropriate action***. But ***if the AI system can distinguish an unknown object from other elements of the current scene, it has learned to identify it!*** That is, the ***adequacy of the response to an unfamiliar situation requires the ability to learn autonomously***. Accordingly, the inability to do this guarantees the periodic occurrence of emergency situations, as was the case in the video about the collision of a Tesla controlled by an autopilot with a jet.

The situations described above convincingly show that ***such learning must be fast*** - the ***period of recognition of unknown objects must correspond to the rate of change in the situation***. In turn, this implies autonomous learning using a ***small amount of data***.

The approach described in AGI: STRUCTURING THE OBSERVABLE**,** based on the search/construction of ***invariants***, can detect unknown objects with a ***small data set***. Required high computing performance is achievable at the current technical level since the approach allows parallelization of the search for the desired invariant.

The requirement of the ability to detect unknown objects, making do with a small amount of data, makes it ***problematic to use any statistical methods*** since all the capabilities they demonstrate require a rather voluminous data set, which cannot be collected.

This, however, is not the only obstacle. For example, ***clustering*** is unsuitable for this purpose because it is a ***classification*** method, that is, a way to ***divide a set of known objects into several subsets***, evaluating the ***similarity of objects***. Recognition of unknown objects requires an entirely different task: to find a ***group of arbitrary distinct objects***, the ***combination*** of which forms the desired new object. The subject of assessment, in this case, is ***not the degree of similarity*** of objects but the ***stability of the system of relations*** between the attributes of the object components, and these ***relations are absent in the input data*** - it is necessary to "***invent***" them. Therefore, the illustrative task

(https://gnosiseng.com/vid/struct.mp4) of detecting two objects, which form the third object, cannot be solved by clustering methods.

SUMMATION

- ***The reliable functioning of technical systems controlled by artificial intelligence requires the ability to quick autonomous (ad hoc) learning during operation.***
- ***Ad hoc learning is not achievable using statistical methods, including cluster analysis.***
- ***An adequate response to the occurrence of an unknown situation can be solved using an approach based on the search/construction of invariants.***

AGI engineering is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.