The Lonely Robot, or Why AI Needs to Get Language Now

By James Lewis

Imagine a spaceship that sails on light. Light Sail 2 is sailing around the Earth right now, and we may soon be able to use light from our sun to send a probe to the farthest reaches of our solar system and even to other stars, and at fantastic speeds (18,600 miles per second) - no gravity assist required. The spaceship must be small, no humans onboard. A robot will have to pilot the probe, using an onboard gyroscope. This robot will essentially be software running on an onboard computer (hardened against the effects of radiation and changes in temperature).

If we outfit the probe with a guidance computer, sensors, cameras and so on, the robot will be able to navigate space, see where its going and stay out of trouble. Its sole directive will be to add to its store of memories and tell us about its experiences. It would be quite a journey for the little robot, and one day we might follow it. Wouldn't we be a little bit curious to find out how it feels to be out there in space so far away, all alone?

We can only guess at what the robot will encounter, so it will have to decide for itself where to go and how to get there. And, if it does manage to arrive at another solar system, it needs to be smart enough to know what to do when it gets there, and not fly into another sun, or someone else's planet, releasing all its momentum in a massive explosion. To decide well it needs an idea of possible outcomes of decisions it makes. Knowing its own history would help, and the histories of other space travellers. The robot must also interpret these histories in such a way that it can extrapolate (predict values for points outside the range of available data) future outcomes from decisions that it, and other space travellers, have NOT made. In other words, it must apply knowledge of historical situations to situations that are completely dissimilar.

Some would say the robot needs an artificial intelligence (AI) powered by machine learning (ML). ML or deep learning (DL) has accomplished amazing feats, such as solving a Rubik's cube with a robotic hand (OpenAI, October 2019). But there's a problem. Solving the Rubik's cube with ML consumed 2.8 gigawatts hours of electricity, or the output of three nuclear power plants for one hour. Our little space probe cannot take three nuclear power plants along on its journey. There's another problem. “So it turns out that we can make a gigantic progress whenever we have access to data... And ultimately that was holding us back in terms of robotics,” said OpenAI co-founder Wojciech Zaremba, when talking about OpenAI’s decision to disband its robotics team (in podcast hosted by startup Weights & Biases). ML can interpolate well if it has lots of data. But we don't have a lot of data on what's out there, and since a lack of data is part of problem, ML is not a good fit.

Symbolic AI is an older form of AI that's been largely superseded by ML, but some suggest a hybrid form of AI combining Symbolic AI and ML is the future. Symbolic AI has rules that work only in very specific cases. For example, if we have a symbolic AI chatbot taking pizza orders over the phone, the chatbot does fine if a caller says, "I want a pepperoni pizza." But it doesn't know what to do if another caller says, "Yeah, I'd like a pizza and I want some pepperoni on it, yeah, that's it, just pepperoni. Please. And extra cheese." With ML and enough training the chatbot might be able to take the pizza orders for both these callers. But even with ML the chatbot still can only take pizza orders; it can't handle anything unexpected. If someone injures themselves, by falling down stairs for example, and calls for pizza because the number's on speed dial, the chatbot wouldn't know what to do. Handling the unexpected is exactly what our robot needs to do as it travels through space.

How does our robot extrapolate from the past to make decisions about the future?

It begins with very old technology, older that even symbolic AI, a log. The robot logs events as they happen, and to make the log more descriptive everything that can possibly be logged gets a label in natural language. Left side, right side, port and starboard if you prefer, all the equipment on board, sail, power source, navigator, everything. Even the robot needs a label. The robot needs to have an identity and be able to refer to its own age, weight, size, temperature, etc. We also need labels for everything the probe's cameras may view. A deep learning model for image classification could classify shapes and identify stars, planets, comets, asteroids, gas clouds, and planetary features and weather. The model would need its own server with a powerful graphics processor (GPU) onboard the probe.

The robot will also have lots of histories, or narratives, stories if you will, onboard, and not only technical procedures for space travel, but fictional stories as well, novels will be particularly helpful, since they are sequential in nature. And, of course, the robot needs a vocabulary, to interpret these narratives and events in its own history.

It's going to need something else, a model. A kind of manifold (Kantian) one might say, a way for the robot to react to the many, varied, and unpredictable events that will occur far from the earth, far out into deep space. This model will be an emotional model. An example: You're rushing to an event one evening, and you clumsily knock your dinner off the kitchen table, the plate smashing into pieces, spaghetti and meat sauce spread across the floor. After the event, while pulling out of a particularly cramped parking spot in an underground garage, you smash your drivers side mirror on a concrete post. The outcome of both these completely different events? You get annoyed, probably angry. What's worse is you know both events are your own fault because you weren't paying attention. How do you avoid that unpleasant emotion in future? You pay attention. So it is with our robot.

Now let's say as the probe approaches the fringes of our solar system its cameras register a flashing light ahead.

The image of the flashing light in the darkness of space would mean nothing to the robot, just a bunch of pixels. DL can help with that by labeling the image as a flashing point of light in space. Writing a program instructing the robot to record details of the image such as the intensity of the light, the periodicity of its flashing, and so on, and then send that data back to Earth would be straightforward. But what about acting, approaching the light for a closer view, for instance? Or perhaps the robot should steer clear of the light, or just ignore it. How does the robot decide what to do? How does the robot know that it needs to decide? Why would the robot decide at all? Machines do not take initiative in our experience. Why would it in this case?

The answer gets back to the robot's sole directive, to add to its store of memories. When I say "directive" what I really mean is purpose, or function. The robot is software running on a computer and all software has a function. This software's function is to add to its store of memories, and tell us about them.

Before the robot records the above event in its log, it checks to see if it has already recorded the event. Finding no record of a flashing light, it records the event. If it attempts to record the event again, or a very similar event a few minutes later, it will find it already has a memory of the flashing light, and so it doesn't record it twice, there's no reason for that. It's store of memories remains the same. Then it must do something, take some action, to increase its store of memories with respect to the light. Now it must decide on what action it should take.

In part 2, we'll explore how emotions help the robot decide on a course of action, and approaches to deductive reasoning.

September 24, 2021