Rational Curiosity

by Sven Nilsen, 2020

In this paper I provide a simple model for rational curiosity.

Assume that there is a computer system which:

1. Asks the user questions
2. Uses the answers of the questions to build its knowledge base
3. Infers from its knowledge base new useful questions to ask
4. Goes to 1)

A such system can be thought of as “rationally curios”, if it minimizes the number of questions asked, such that it can build the greatest knowledge base possible from what it learned, when needed.

One counter-intuitive property of a such system is that it is not necessary to expand its knowledge base to maximum while asking the user questions. It would be sufficient to *know* that it can figure stuff on its own later. Hence, the truth values of questions are not as important to reason about as the relationship between answers of questions.

For example, if a bit is `0` and `1`, then the following two questions:

- Is the bit `0`?
- Is the bit `1`?

Are exclusive when minimizing the number of questions asked.
This holds no matter what the answer of the questions are.

I showed in the paper “Symmetric Avatar Paths” that “answered” and “truth” predicates of questions are related by the `?[truth => answer]f` symmetric avatar path of `f` transformations of questions.

This means that in order to be rationally curious, the system should reason over the symmetric avatar paths of its models of the environment in the knowledge base and construct only questions that can not be answered from its existing knowledge.

Rational curiosity also requires an efficient meta-optimization of its ability to reason over symmetric avatar paths. This cost must be balanced with constraints of time and computing capacity.

If a such system is rationally curios, then it might still fail to be generally intelligent.

For example, when something takes nearly an infinite amount of time to figure out on its own:

It would choose to not ask questions that could save a lot of time later,
but instead only ask questions which answers it can not possibly figure out on its own.

This means that the definition of rational curiosity deviates from the definition of general intelligence.