Bootstrapping Technique for Semantics of Natural Languages

by Sven Nilsen, 2019

In the field of path semantics, the most important issue is to define or formalize some procedure or create some tool to help understanding mathematical languages. It is not that important that the procedure or tool is practical, because grounding of language could mean counter-factual reasoning, such as “if we could perform experiment X, then Y would happen”.

Natural languages are much harder to analyze than e.g. logic, since the identity of words in natural language is more complex than the identity of words in logic. This corresponds to different interpretations of the path semantical axiom, or “ways of talking about symbols”.

If two words are the same in natural language, they are used the same way. However, since usage of words can be very complex, it is difficult to construct an environment where experiments can be controlled.

To overcome this limitation, I suggest creating a very simple environment which is used to “bootstrap” the semantics of natural languages. From this very simple environment, more complex environments can be constructed that handle sub-sets of predictions about the simpler environment.

This simple environment consists of the following:

- A program that outputs tasks randomly in text form
- A human that clicks a button on a standard keyboard (just one key press)
- The program pauses 1 second when a button is clicked, then outputs a new task
- The only data collected is the text and which button was clicked (identified by a number)
- The human is not receiving feedback or any rewards for which buttons are clicked, nor prevented from clicking buttons that are not described in the text

The motivation of this environment is that the simplicity of it makes it easy to set up without requiring lots of resources. By ignoring all relevant variables that might influence the human decision maker, one can study the ambiguity that arises from not knowing these variables.

Since complex environments, that handle sub-sets of predictions about the simpler environment, are deeper, they can generate more data and explain what is happening in the simpler environment.

Unlike in perfect information systems, this way of bootstrapping means that there are hidden states. It might be possible to predict a button press by using the sequence of previous actions. Even if the tasks are chosen randomly, the response of the human might contain information about what it will do next. However, overall this environment has not enough information to distinguish the meaning of most sentences from each other. It is only sufficient to cluster semantics into groups based on their probabilistic characteristics. Imperfections and biases must also be dealt with through careful analysis and comparison of data between various environments.