Review on "The Symbol Grounding Problem" by Stevan Harnad

Irdi Balla

18/Jan/2108

1 The Symbol Grounding Problem

Harnad In this paper discusses about the "Symbol Grounding Problem", which he describes as making the semantic interpretation of a formal symbol system, intrinsic to that system. He compares this task to that of learning Chinese from a Chinese/Chinese dictionary.

2 Modeling the Mind

2.1 Symbolism

According to Harnad a symbol system is:

1. a set of arbitrary "physical tokens" scratches on paper, holes on a tape, events in a digital computer, etc. that are

2. manipulated on the basis of "explicit rules" that are

3. likewise physical tokens and strings of tokens. The rule-governed symbol-token manipulation is based

4. purely on the shape of the symbol tokens (not their "meaning"), i.e., it is purely syntactic, and consists of

5. "rulefully combining" and recombining symbol tokens. There are

6. primitive atomic symbol tokens and

7. composite symbol-token strings. The entire system and all its parts – the atomic tokens, the composite tokens, the syntactic manipulations both actual and possible and the rules – are all

8. "semantically interpretable:" The syntax can be systematically assigned a meaning e.g., as standing for objects, as describing states of affairs.

Symbolists believe that there is a symbolic level(or mental level) with specific rules independent of the physical realizations.This level differentiates the physical from the cognitive phenomena. Recent success in AI conforms with this level as well. All eight properties are necessary for a system to be considered symbolic and symbolism is a systematic property, meaning only a system can be symbolic (not just a chunk of a system).

2.2 Connectionist System(Nets)

On the other side of the Symbolic system, theres the Connectionism which has re-appeared lately under names like "Neural Networks" or "Parallel Distributed Processing". Connectionism, according to Harnad is based on the principle that "cognition is dynamic patterns of activity in a multi-layered network of nodes or units with weighed positive and negative interconnections". The patterns change according to internal constraints and are adjusted based on the input. This system seems to be based on statistical structures.

2.3 Scope and Limits of Symbols and Nets

The limitations and capabilities of symbolic and net systems are still unclear. Symbolic systems seem better at language-like tasks while net systems do better at sensory and motor task but both approaches go as far as carrying out "toy" tasks. There has also been talks that nets are symbolic which were refused by Fodor and Pylyshyn.

3 The Symbol Grounding Problem

3.1 The Chinese Room

In this example the core assumption of Symbolic AI is challenged, which states that if an AI could give the same answers as a human to the same questions(even when being tested for a very long time) must have a mind. Searle counters this assumption with and imaginary example of himself. If he were to manipulate the Chinese symbols(while not knowing any chinese) so that the output would be correct and close to that of a Chinese person he passes the test but yet he does not understand Chinese.

3.2 The Chinese/Chinese Dictionary-Go-Round

Harnads own example of the symbol grounding problem consists of 2 parts; one hard and one impossible. The first scenario is to learn Chinese as a second language by using a Chinese/Chinese dictionary. This could also be considered similar to that of translating old languages. Harnad would red through the dictionary going from symbol to symbol never understanding what each symbol means.

The second scenario is to learn Chines as a first language from a Chinese/Chinese dictionary.

3.3 Connecting to the World

Harnad proposes a hybrid symbolic/non-symbolic system. In this system the symbols are grounded in 2 kinds of non-symbolic representations that pick out the distal object categories of the symbol.

4 Human Behavioral Capacity

Human beings can :

1. discriminate
2. manipulate
3. identify
4. describe the objects, events and states of affairs in them
5. and then produce descriptions
6. and respond to descriptions

4.1 Discrimination and Identification

Discriminations is a relative judgment; our ability to tell things apart(telling horses apart, and judging which ones are similar). Identification is an absolute judgment to tell whether an input belongs to a category(When seeing a horse being able to call it a horse).

4.2 Iconic and Categorical representations

Iconic Representations are defined as "internal analog transforms of the projection of distal objects on our sensory surfaces". Same/different judgments are based on the sameness or difference of the representations. Discrimination and identification are independent. We can tell things apart without knowing what they are. Invariant features of icons, that can distinguish members of different categories, are needed for identification. These outputs are called "categorical representations" by Harnad. The categorical representations are either innate or learned. Iconic and categorical representations are non-symbolic.

5 A complementary role for Connectionism

Harnad's system does not account for the way the categorical representations are formed. This is where connectionism comes to play. With its ability to learn patterns it holds a strong case as a candidate. Like a Neural Network it can dynamically adjust the weights of the features, after taking as input the icons and names of the objects, and thus reducing the icons to the invariant features. This hybrid model addresses the main weaknesses of the independent models. Here there is no symbolic level but rather a dedicated symbol system, whose symbols are connected to non-symbolic representations to pick the objects they refer to using a network that extracts invariant features.