Ductive cognition - a model of cognitive computation for Artificial General Intelligence

Aaron Turner
aaron.turner@bigmother.ai https://bigmother.ai

This is not a paper. It's a collection of thoughts that may one day find their way into a paper via a blog.

Cognitive computation

Whereas the instantaneous state of a classical computation is maintained by a set of variables to which classical operations are applied, thereby yielding a continuously evolving set of variables:

variables ← *initial values*
while (¬*finished*) {
  variables ← *operation*(*variables*) // decomposing to classical primitives such as +, −, ×, ÷
}

the instantaneous state of a **cognitive computation** is maintained by a **belief system** (set of beliefs) to which **cognitive operations** are applied, thereby yielding a continuously evolving belief system:

belief_system ← *initial beliefs* // such as the axioms of NBG set theory, for example
while (¬*finished*) {
  belief_system ← *cognitive_operation*(*belief_system*) // decomposing to cognitive primitives
}

Just as the set of all possible classical computations is determined by the set of classical primitives, so the set of all possible cognitive computations is determined by the set of cognitive primitives.

Beliefs

The central concept of **belief** may be defined with varying degrees of precision; for example:

1. A belief is a declarative statement (a statement that is deemed to be either true or false)
2. A belief is a wff (well-formed formula) of some suitable logical system, such as propositional logic, first-order logic, Church's type theory, lambda calculus, or some extension thereof
3. A belief is a theorem of first-order logic with equality plus descriptions (FOLEQ)
4. A belief is a theorem of FOLEQ further extended with named definitions (UL)
5. A belief is a theorem of UL further extended with the axioms of NBG set theory (UL-0-NBG)
6. A belief is a theorem of UL-0-NBG further extended with a toolkit of definitions encompassing sets, relations, functions, orders, numbers, sequences, linear algebra, calculus, probability, etc
7. A belief is a theorem of some logical system defined using the notation of 6 as metalanguage

Following Rule G, beliefs of type 7 are sufficiently expressive to capture all of mathematics.

Cognitive primitives

We propose the following set of "IDA" cognitive primitives:

1. induction – synthesises new beliefs derived from and justified by observations of the universe
2. deduction – synthesises new beliefs each of which is a logical consequence of prior beliefs
3. abduction – synthesises new beliefs of the form p(x) [essentially, "find term x such that p(x)"]

Intuitively:

- induction ≈ learning, recognising patterns/regularity, gaining intuition through experience
- deduction ≈ revealing implicit/latent knowledge implied by explicit knowledge, prediction
- abduction ≈ creativity, problem-solving, hypothesis generation ["find h such that h ⇒ o"]

Some observations:

- these are **low-level** primitives – the real magic happens when they are intricately combined
- learning in AGI ≠ learning in ANI (e.g. contemporary ML) – these are very different things
- GOFAI, which still gets bad press, was just deduction on it's own – that's insufficient for AGI
- similarly, the AI field's current obsession with ML/neural nets is also insufficient on its own
- even neuro-symbolic AI (effectively induction + deduction) is insufficient without abduction

Following Rule G, each of the above cognitive primitives is maximally general in nature. Consequently, we believe the set of IDA primitives to be **cognitively complete**, meaning that it is sufficient for the purposes of implementing any meaningful cognition. (In particular, any **rational** cognition, which is the only kind we consider worthwhile. "We believe that AGI should be irrational" said no one ever.)

Ductions

The IDA primitives shall be called **ductions**. A belief derived via a finite sequence of ductions shall be called **ductive** (non-initial beliefs are **properly ductive**). A cognitive computation all of whose beliefs are ductive is ductive. "Ductive cognitive computation" may be abbreviated to "**ductive cognition**".

Provable rationality

Following Rule R, if a ductive cognition's initial beliefs are justified by rational argument (e.g. "after extensive use across all of mathematics, modern set theory appears, empirically, to be consistent"), and each IDA cognitive primitive is provably rational, then **the overall cognition is provably rational**.

Provably rational oracles and agents

A provably rational ductive cognition may provide the foundation for a provably rational AGI oracle. A provably rational AGI oracle may then provide the foundation for a provably rational AGI agent.