Response Paper
Juan Pablo Rodriguez
September 10th, 2021

Parallels between Machine Learning and Theories of Concepts

I found parallels between the characterization of some theories of concepts and Artificial Intelligence (AI) Paradigms. For instance, the Classical Theory (CT) with Symbolic AI, Prototype Theory with Subsymbolic AI and Dual Theories with Neural-Symbolic AI.

1. Symbolic AI uses symbols as primitives and relies heavily on classical logic, which resonates with the CT's postulate that concepts are structured mental representations that encode necessary and sufficient conditions for their applications.
2. Subsymbolic AI derives implicit representations without the use of symbols. I can't help but notice the parallel with the Prototype Theory which suggests that the features of a concept express statistically prominent properties.
3. Neural-symbolic AI combines the strength of Symbolic AI's to reason over symbols with the Subsymbolic's AI flexibility. Neural-symbolic approaches emerged, much like the motivation to come up with dual theories, to harness the power of two apparently opposing approaches that complement each other.

The use of phenomena a theory of concepts is responsible for to evaluate language models

Generative language models build an internal representation for language that is encoded in the weights of a neural net. We should draw inspiration from the phenomena a Theory of Concepts is responsible for explaining to inspect the quality of language models. For example, one could:

- Inspect the way it applies and extends categories.
- Inspect its analyticity.
- Inspect how it composes complex concepts.
- Inspect its stability to variation in surrounding beliefs.
- Inspect the ignorance and error problem.

I believe that looking at models through this lens and drawing inspiration from experimental design by psychologists, one could create a suite of evaluations that would allow us to gauge language model's shortcomings and strengths. In addition, it will provide us with a framework to think about their safe deployment in various domains.