(Abstract) Automatic schematization of biological mechanisms via alignment and abstraction of causal relations among domain specific causal pathways

Harrison Pielke-Lombardo

1 Questions and Topics

- What are potential AI applications in which humanlike concept formation, abstraction, and analogy could improve performance and make systems more robust?

- What is known in psychology and neuroscience about the mechanisms by which humans (and non-human animals) develop and use concepts, form abstractions, and make analogies? How can such mechanisms inspire AI research?

- Can gradient-descent-based systems learn to produce analogical reasoning on novel problems? What can the state-of-the-art in inductive program synthesis teach us about abstraction and reasoning?

  - How can abstraction and analogy-making abilities in AI systems be assessed? What can we do to ensure that performance on a test will guarantee generalization?

- Can we discover general computational mechanisms for abstraction and analogy by focusing on idealized microdomains, or could the real challenges lie in interfacing analogical mechanisms with a vast array of commonsense knowledge?

- If we want machines to creatively invent wholly new theories from data, like scientists do, what roles would abstraction, analogy, and strong generalization play?

2 Track

☐ AI for Social Good

☐ Artificial Intelligence in Government and Public Sector

☐ Cognitive Systems for Anticipatory Thinking

☒ Conceptual Abstraction and Analogy in Natural and Artificial Intelligence

☐ Physics-guided AI to Accelerate Scientific Discovery

☐ Trust & Explainability in Artificial Intelligence for Human-Robot Interaction (AI-HRI)

3 Submission preference: poster

4 Keywords

- biological mechanisms

- analogy generalization

- knowledge based reasoning

- semantic ai

5 Abstract

One of the main goals of biomedical research is to explain the biological phenomena involved in human diseases and their interventions, yet computational methods stop short of providing human-like explanations. Presented here is an automatic method based on analogical reasoning for generating causal schema to support the explanation of biological phenomena. The approach used offers the potential for systematic, scalable, and hierarchical mechanism schematization which could be used for mechanism retrieval, identification, evaluation, and inference tasks.

The transmission and transference of ideas among the many domains of biomedical research can be difficult. It is diverse, with many areas of specialty each with there own names, jargon, questions, and interests. Though there is overlap among the processes and pathways which are relevant to each area of specialty, the context in which they are studied can vary by the organism, chemical perterbation, tissue type, genetic variant, or disease. Common among them, however, is the pursuit of mechanistic understanding.

Analogical reasoning can be used to align causal mechanisms to produce human-like explanations in the form of mechanism schema. Analogies (framed in text as metaphors for brevity and listed below as analogy assertions) such as "the heart is a pump", "the lock and key enzyme-substrate hypothesis", and "unzipping DNA by helicase" are often used to communicate biological mechanisms because they relate domain specific knowledge with widely understood phenomena to elucidate the underlying causal pathways.

- Blood is to a heart as water is to a pump.

- Enzymes are to locks as their substrates are to keys.

- Helicase is to a zipper slider as nucleotides are to zipper teeth.

Formal representations of alignments across related biological mechanisms could achieve the same effect. A hypothesis and accompanying potential method for achieving such alignments is as follows.

Domain specific causal knowledge of biological mechanisms, formally represented as subgraphs of a knowledge graph, are structurally aligned to produce formal representations of mechanism schema. The alignment of subject mechanisms begins by using both causal and non-causal knowledge to align and score all possible causal relationships among the mechanisms. Sufficiently high scoring relationship alignments are then used to construct the resulting mechanism schema. Entities and activities which are not shared among the mechanisms are resolved by identifying proxy nodes and their score is weighted by the semantic distances between the proxy and its target. Importantly, chains of causal relations are generalized during the alignment into what could be considered components or sub-mechanisms.

Initial results from this mechanism schematization strategy show that knowledge can be inferred across domain specific mechanisms. For example, the ability for the breast cancer drug tamoxifen to inhibit the growth of estrogen dependent breast cancer can be inferred by comparing a minimal subset of knowledge about breast

cancer to knowledge about LHRH agonist's ability to inhibit the growth of testosterone dependent prostate cancer. Further evaluation will include hierarchical clustering of resulting mechanism schema constructed from RE-ACTOME pathways to compare to the hierarchical characterization of biological processes found in the Gene Ontology's Biological Process hierarchy.

Analogy schematization has a number of potential applications from retrieval and identification, to comparison and transformation, and even for causal inference. For a field such as biology, scalable and systematic methods to schematize mechanisms could improve the transfer of knowledge and improve explanations critical to improving the treatments of complex diseases.

6   Acknowledgments