Using Reasoning to Evaluate Scientific Discoveries

Jaimie Murdock and Ed Carroll, Sandia National Labs

There are many approaches to the creation and evaluation of a "novel" idea. LLM agents provide a simple way to generate new ideas via prompt engineering. These new ideas are a recombination of their training data and represented as textual data. To evaluate these ideas, we propose the use of large-scale, curated knowledge graphs (KGs) as a ground truth. These KGs are credible because they have been verified by human editors (such as Wikidata and OpenCyc). The LLM-generated concepts can be translated to a KG via traditional graph mining algorithms. Comparing the LLM-derived KG to these reference KGs allows us to establish credibility of the LLM output.

To measure the novelty of a "new" idea, we measure the number of assertions that would need to change in a ground truth KG to the novel idea. This is a "belief revision" approach with more substantial concepts requiring a larger amount of belief revision. This is a form of validation for a new idea.

Novel concepts are often novel not in their apparition out of nowhere, but rather their recombination of existing concepts. Reasoning over a KG offers an alternative method of proposing new ideas. For this category of problems, we have revealed the unknown-known, which are facts that may be inferred from other facts, but which are themselves not consciously known to a system. Rather than looking at statistical relations, these discoveries are more readily derived from symbolic relations within a KG.

Unfortunately, there are many things which are known, but are not stated due to triviality. This necessitates the creation of a ranking mechanism to determine which non-asserted inferences are most important or "interesting". The structure of the KG can be used to weight the "interestingness" of new inferences by leveraging the concept hierarchy in the KG. At this symposium, we would like to lead discussion on what makes a novel concept "interesting" and ways to leverage KGs to evaluate this problem.

At Sandia National Labs, we have been examining this within the domain of systems engineering. For example, each system requirement should specify at least one component, and each component should perform at least one function. These are the easy questions to reason about. The more interesting ones are whether the requirement that was the basis of a function is the same requirement that specified the component which performed that function. Some requirements may be deliberately underspecified, such as the screws used in an assembly, because they involve an industry standard part or because the component is the same one incorporated in a previous version of the system. Any inference method used to propose new axioms should be robust to filter these trivialities from the result set.

In this presentation, we will explore differences between discoveries of inference vs. discoveries of belief revision, showing how knowledge graphs can give great generative power for the former category, while LLMs approach the latter. We believe the combination of graph reasoning and LLM modeling can be very powerful in revealing new scientific discoveries by evaluating the "interestingness" of a new concept.

References

Jaimie Murdock, Edward R. Carroll, and Kathryn Gray. The Model Assessment Wizard (MAW): A Visualization System for Ontology Constraint Violations. SAND2023-14860. Albuquerque, NM: Sandia National Laboratories, 2023. url: https://prod-ng.sandia.gov/sand_doc/2023/2314860.pdf .

Jaimie Murdock and Edward R. Carroll. Simplifying and Visualizing the Ontology of Systems Engineering Models. SAND2021-7079. Albuquerque, NM: Sandia National Laboratories, 2021. url: https://prod-ng.sandia.gov/sand_doc/2021/217079.pdf.

Jaimie Murdock, Cameron Buckner, and Colin Allen. “Evaluating Dynamic Ontologies”. In: Knowledge Discovery, Knowledge Engineering, and Knowledge Management: Second International Joint Conference, IC3K 2010, Valencia, Spain, October 25-28, 2010, Revised Selected Papers. Ed. by Ana Fred, Jan L. G. Dietz, Kecheng Liu, and Joaquim Filipe. Vol. 272. Communications in Computer and Information Science. Springer-Verlag, 2013, pp. 258–275. doi: https://dx.doi.org/10.1007/978-3-642-29764-9_18