Investigating Ontology Use in Artificial Intelligence and Machine Learning for Biomedical Research

- A Preliminary Report from A Comprehensive Literature Review

Asiyah Yu Lin[1], Andrey Ibrahim Seleznev[2], Tianming "Danny" Ning[3], Paulene Grier[4,5], Lalisa Mariam Lin[6], Christopher Travieso[7], Jaleal Sanjak[8], Ansu Chatterjee[9]

[1] National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
[2] Walter Johnson High School, Bethesda, MD 20817, USA
[3] Winston Churchill High school, Potomac, MD 20854, USA
[4] Thomas Stone High School, Waldorf, MD 20601, USA
[5] College of Southern Maryland, La Plata, MD 20646, USA
[6] Walt Whitman High School, Bethesda, MD 20817, USA
[7] Our Lady Of Good Counsel High School, Onley, MD 20832, USA
[8] National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD 20850,USA
[9] Office of Director, National Institutes of Health, Bethesda, MD 20852, USA

Abstract

As a form of knowledge representation, ontologies organize the knowledge and data hierarchically ("tree-like") and horizontally ("network-like" or "graph-like") using semantic relations, such as "is-a" or "part-of". Artificial Intelligence (AI) and/or Machine Learning (ML) often apply mathematical models that require numeric data as input. The fast growing and big volume of biomedical data has benefited the fast-advancing AI/ML algorithms and frameworks. However, leveraging the non-numerical, semantic, and hierarchical relations from an ontology remains a challenge in AI/ML. In this report, the authors conducted a comprehensive literature review to answer a question: how ontologies are being used in the AI/ML approaches to solve biomedical research problems?

On the date of Sep.4, 2022, a total of 503 papers were retrieved from PubMed Central® (PMC) archive using keywords appeared in title and abstract: *ontology, artificial intelligence, machine learning, deep learning, neural network and embedding*. After reviewing the top 250 highly relevant papers, the authors identified the four major categories of using ontology in AI/ML: 1. Use ontology itself or ontology terms as data labels to be the training datasets; 2. Transform the ontological representation as numerous data to be used in the downstream AI/ML, which includes calculate term's semantic similarities, construct concepts association matrix, and use word embedding algorithms, and etc.; 3. The ontology as a graph structure or network structure is a part of neural network architecture; 4. The ontology classification is the target of the AI/ML classifier. A lot of studies are genomics based. The most popular ontologies used is Gene Ontology (GO), followed by UMLS. There are many specifically built ontologies for the specific tasks.

The future plan of this literature review include completing the reviewing of the rest of 253 retrieved papers, utilize ontological terms to better categorize the AI/ML approaches especially in above-mentioned #3, analyze the trend of ontology use, trend of AI/ML methodologies, analyze the result reports, parameters used for the performance evaluation, funding mechanism analysis, and the biomedical domain analysis.

In conclusion, ontology provides contextually rich data to help the AI/ML to achieve a higher performance compared to the same methods without ontologies. Utilizing the graph-structural and semantics within an ontology requires more complex neural network architecture along with many other components. Explainable AI is an emerging field where the explanatory techniques can explicitly show why a recommendation or a prediction is made.

Keywords

Ontology, Artificial Intelligence, Machine Learning, literature review