Ph.D. Thesis

Knowledge Base Construction Methods for Neuro-Symbolic IA Exploitation

Diego Rincon-Yanez

Abstract

Knowledge representation has long been considered a key task in the information sciences. Nowadays, with AI's broad scope in all aspects of daily life, understanding how to represent, display, and analyze human knowledge has become increasingly vital. This thesis relies on computational systems' inherent ability to deliver comprehensive and accurate responses using a combination of inductive learning, commonly known as Deep Learning, and logic learning; the topic is Neuro-Symbolic AI (NeSyAI). To fully use the NeSyAI system-like characteristics, Knowledge Bases must be built, which are a collection of statements ordered to form a Graph type network structure, also known as a Knowledge Graph (KGs) in the semantic community. Knowledge Graphs can offer background information on one or more domains, adding extra context to the task at hand. This thesis offers numerous methods for integrating contextual knowledge into Neuro-Symbolic AI systems by producing semantically enhanced knowledge bases for later exploitation.

Initially, the Internet of Things (IoT) real was explored, where sensor network characteristics are mapped into a traditional OWL-type ontology using and building a knowledge base of the sensor, events and observations with the aim of predicting events and places. For this, a state-of-the-art ontology was created for IoT event recognition purposes, followed by the implementation of a convolutional Neuron Network, which is able to take the KG and discover new wave information. These two tasks combined in a single flow produce and validate a new architectural pattern for Neuro-Symbolic AI. To validate the proposed approach, two use cases were implemented: a volcano discovering seismic events and a cleanroom monitoring environment for discovering pollution events.

In the second part, knowledge graph embeddings (KGE) were exploited to push the boundaries of the knowledge base construction task in several alleys, such as knowledge validation, question answering, and link prediction from the Neuro-symbolic AI perspective. These tasks were used along with well-known metrics to test the construction approaches from prosaic text leveraged by the external knowledge of one of the biggest Open KG available, such as DBpedia and transformers-type models, arriving at a compelling

results compared to the current SOTA, allowing to test also new and improved NeSy AI architecture patterns. On the other hand, the need to explore synthetic data generation was raised considering the need for continuous experimentation and automation to test the experimentation scenarios. Finally, an initial explainability approach for KGE was conceptualized and applied to the commodities tradeflow prediction use case.