

氏 名 Md Mostafizur Rahman

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 2187 号

学位授与の日付 2020 年 9 月 28 日

学位授与の要件 複合科学研究科 情報学専攻  
学位規則第6条第1項該当

学位論文題目 Type-Embodied Representation Learning for Knowledge  
Graph Embedding

論文審査委員 主 査 教授 高須 淳宏

教授 武田 英明

准教授 相原 健郎

准教授 市瀬 龍太郎

教授 正田 備也

立教大学 人工知能科学研究科

(Form 3)

## Summary of Doctoral Thesis

Name in full Md Mostafizur Rahman

Title: Type-Embodied Representation Learning for Knowledge Graph Embedding

A Knowledge Graph (KG) is a knowledge base containing facts about real world entities represented as a graph. We have witnessed rapid growth in knowledge graph construction and application in recent years. A large number of KGs, such as Google Knowledge Vault, Wikidata, DBpedia and Freebase have been served several fields of real world applications from semantic parsing and named entity disambiguation to information extraction and question answering. In this thesis we develop new methods for effectively using entity types in KG embedding and ranking entity types for better understand the entities.

We start this thesis by presenting novel techniques for using entity types in knowledge graph embedding for knowledge graph completion. Popular KG embedding methods focus on the structured information of triples and maximize the likelihood of them. However, they completely ignore the semantic information contained in most knowledge graphs and the prior knowledge indicated by the semantic information. To overcome this shortcoming of the embedding methods, we propose an approach that integrates the structured information and entity types in relational context which describe the categories of entities. With the type-based prior distributions, our approach generates multiple embedding representations of each entity in different relational contexts and compute the plausibility for triples using the state-of-the-art methods. Extensive experiments show that entity types exhibit useful semantic information to describe the entities.

Entity type information plays an important role in knowledge graphs (KGs). In a KG, an entity usually holds multiple type properties. In our second contribution, we address the entity type ranking problem by means of knowledge graph embedding models. We try to show how entity type ranking can exhibit the corresponding entity's characteristics. In our work, we show that entity type ranking can be seen as a special case of the KG completion problem. Our proposed approach outperforms the state-of-the-art type ranking models while, at the same time, being more efficient and scalable. Finally, we discuss the problem of assigning relevance scores for triples from type-like relations. We can employ those scores as a fundamental ingredient for ranking results in entity search. We propose a joint semantic relevance learning approach using the text and structured data. The results of the triple scoring work indicate important directions for the future work.

Knowledge is the core power in the age of data, incorporating the knowledge in

graphical representation has been around us for decades now. It is an interesting point to make about how "important" and "strategic" data is and how to extract "knowledge" from raw data and store it. To serve that purpose knowledge graph concept comes to action, which is a way of storing entity-centric data modeled as a graph.

The Knowledge Graph enables us to search anything related to our interest and context. But things were not that smooth before the KG era. For decades search has essentially been about matching keywords to queries that means matching the keywords and searching results were not much satisfactory as nowadays. Surprisingly, queries in search engines have increased awfully only because of user satisfaction. Google receives over 63,000 searches per second and 5.6 billion searches per day.

A KG is a multi-relational graph composed of entities and relations between the entities. Modern KGs have been successfully applied to many real-world applications, from semantic parsing and named entity disambiguation to information extraction and question answering and became an appealing topic in machine learning domain.

This dissertation focuses on the influences of entity types in knowledge graph applications. Representation learning models for knowledge graph embeddings typically encode entities or concepts of a knowledge graph in a continuous and low-dimensional vector space, where the relational inferences of entities are modeled as semantically meaningful vector. Hence, these models provide efficient and versatile methods to incorporate the symbolic knowledge of multi-relational data into machine learning.

Most of the currently proposed KG embedding models perform the embedding task solely on the basis of observed facts. Given a KG, such a model first represents entities and relations in a continuous vector space, and defines a scoring function on each fact to measure its plausibility. Those representation learning based models concentrate on structured information in triples and ignore the rich semantic information of entities and relations, which is contained in the KGs. Semantic information includes entity types, entity's text descriptions/summary, and other textual information. There is no doubt that recently proposed models have significantly improved the embedding representations and increased the prediction accuracy, there is still room for improvement by exploiting semantic information such as entity types. One of the main drawbacks of knowledge graph completion is the polysemy of entities or relations, i.e., each entity or relation may have different semantics in different triples based on the relational context. For example, in the triple (Tom Hanks, starred in, Terminal), the type of entity Tom Hanks can be considered as actor here, while in (Tom Hanks, director of, Larry Crowne), Hanks is a director. If we keep the same vector representation for the entity Tom Hanks for these two relations, it may increase confusion. The current popular models miss the rich semantic type information.

In the today's world search queries are entity centric data oriented. One of the important piece of information associated with the entity is the type/category information.

However, an entity is usually not associated to a single generic type but rather to a set of more specific types, which may be relevant or not given the document context. Ranking between those types may help Search Engine Result Pages (SERPs) to display meaningful search results of user initiated queries. So efficient entity type ranking method is required to tackle such problems.

In this dissertation, we aim at investigating multi-relational representation learning methods leveraging entity type for knowledge graph completion, which capture complex properties of the real-world facts. We also address the problem of entity type ranking and propose some techniques to tackle that task.

Knowledge graphs are becoming an increasingly popular way of thinking about and organising data within major business firms. As with all data management and governance projects, we can define the use of KG and achieve the expected growth in managing the data. The way KG is growing it may become the new data management system.

## 博士論文審査結果

Name in Full  
氏名 Md Mostafizur Rahman

Title  
論文題目 Type-Embodied Representation Learning for Knowledge Graph Embedding

本学位論文は、知識グラフにおいてエンティティのタイプを考慮した埋め込み表現法を提案している。まず、知識グラフの主要な問題の1つであるリンク補完問題においてエンティティのタイプに応じて埋め込み表現を変換する方法を導入した。既存の埋め込み表現と組み合わせてリンク補完を行う方法を提案し既存手法との比較実験を通してその有効性を示した。次に、エンティティ検索においてエンティティに関連の深いタイプを求めるための知識グラフの構成法とその埋め込み表現を提案しその有効性を実験的に示した。

本学位論文は7章で構成され英語で書かれている。

第1章では、本研究が扱う知識グラフの概要と知識グラフ構築の課題について論じ、本学位論文の主要な貢献を述べている。

第2章では、知識グラフに関する基本的な概念や用語を説明した後、代表的な知識グラフの特徴について概説している。

第3章では、関連研究をサーベイしている。まず、知識グラフの構築や利用において重要な役割を果たす埋め込み表現について、その種類や特徴を示した後に、本学位論文が扱う知識グラフのリンク補完、知識グラフを用いたエンティティタイプのランキングについての関連研究をまとめている。続く3つの章で本論文の主たる貢献を述べている。

第4章ではエンティティのタイプを利用した知識グラフのリンク補完法を提案している。知識グラフはエンティティとその間の関係をグラフで表現したものであるが、エンティティ間の関係をすべて列挙することは難しい。このような不完全な知識グラフにおいて明示的に記述されていない関係を補完する問題は、知識グラフの重要な研究課題として研究が進められている。従来の手法は、主にグラフ構造に基づいてエンティティやその関係の埋め込み表現を求め、その表現の類似度に基づいて知識グラフのリンクの補完を行う。本学位論文では、知識グラフの構造に加えて、タイプに基づいてエンティティの埋め込み表現を求める方法を提案している。具体的にはリンクの種類とエンティティタイプの関係に基づいて埋め込み表現を変換する行列を導入している。リンク補完の研究においてその性能評価によく用いられる4種類の知識グラフを用いて提案手法の性能評価を行い、translation based model と呼ばれる種類の知識グラフの埋め込み表現においてタイプを使うことで補完性能が向上することを示している。また、関連するタイプ数の多いエンティティほど大きな性能向上が得られることを示している。

第5章では、知識グラフの埋め込み表現を用いて各エンティティに関連づけられたタイプのランキング法を提案している。通常、各エンティティには複数のタイプが関連づけられており、その重要度や関連度に基づいてエンティティの内容を示すことで検索結果をわ

かりやすく提示することができる。本研究ではタイプをエンティティと見做して知識グラフを構成し、その埋め込み表現に基づいてタイプを関連度に基づいてランキングする方法を提案している。先行研究で構築された評価データを用いて比較実験を行い、提案手法の有効性を示している。

第6章は、エンティティの説明文および知識グラフにおけるエンティティとタイプの間  
の経路を組み合わせてエンティティとそのタイプの関連度を求める方法を提案している。

第7章では、以上の結果をまとめるとともに今後の課題を示している。

公開論文発表会において、出願者はおよそ45分で博士論文の内容を説明し、その後、15分程度の質疑応答が行われた。続いて審査委員による質疑が行われ、提案手法の特徴、評価実験の詳細について質問とコメントが寄せられ、出願者は適切に回答した。

質疑応答後に審査委員会を開催し、審査委員で議論を行った。博士論文審査の結果、出願者は情報学分野の十分な知識と研究能力を持つと認められ、また研究内容は学位論文として十分なレベルの新規性、有効性があると認められた。本論文の内容に関し、電子情報通信学会論文誌に1編、査読付き国際会議に3編の主著論文が採択されている。以上より、審査委員会全員一致で、博士論文として十分な水準の研究であると認め、学位の授与に値すると判断した。