

氏 名 Hung Nghiep Tran

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 2189 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 Multi-Relational Embedding for Knowledge Graph
Representation and Analysis

論文審査委員 主 査 教授 高須 淳宏
教授 相澤 彰子
教授 山田 誠二
准教授 相原 健郎
教授 宮尾 祐介
東京大学 大学院情報理工学系研究科

(Form 3)

Summary of Doctoral Thesis

Name in full: Hung Nghiep Tran

Title: Multi-Relational Embedding for Knowledge Graph Representation and Analysis

Multi-relational data, such as knowledge graphs, bibliographic data, and information networks are prevalent in real-world datasets. Managing, exploring, and utilizing these large and complex datasets effectively are challenging. In recent years, multi-relational embedding methods have emerged as a new effective approach to model multi-relational data by representing both the entities and the relations as embedding vectors in semantic space. On knowledge graphs, multi-relational embedding methods aim to model the interactions between these embedding vectors to predict the relational link between entities. These knowledge graph embedding methods solve the important inherent task of link prediction for knowledge graph completion, but also provide the embedding representations that have various potential applications. The goal of this thesis is first to study multi-relational embedding on knowledge graphs to propose a new embedding model that explains and improves previous methods, then to study the applications of multi-relational embedding in representation and analysis of knowledge graphs.

For the first part of the thesis, we study the theoretical framework of knowledge graph embedding methods to explain and improve them. We review and analyze the popular class of semantic matching knowledge graph embedding methods, with a focus on the state-of-the-art trilinear-product-based models such as ComplEx. Based on our analysis, we identify two fundamental complementary aspects that a knowledge graph embedding model needs to address, that is, computational efficiency and model expressiveness. Previous trilinear-product-based models use specially designed interaction mechanisms to manually provide a trade-off between the two aspects. However, their interaction mechanisms are specially designed and fixed, potentially causing them to be suboptimal or difficult to extend. In this thesis, we propose the multi-partition embedding interaction (MEI) model with block term format to systematically address this problem. MEI divides each embedding into a multi-partition vector to efficiently restrict the interactions. Each local interaction is modeled with the Tucker tensor format and the full interaction is modeled with the block term tensor format, enabling MEI to control the trade-off between expressiveness and computational cost, learn the interaction mechanisms from

(Form 3)

data automatically. The model combines advanced tensor representation formats and modern deep learning techniques to achieve state-of-the-art performance on the link prediction task. The theoretical framework of the MEI model is then used as a general mechanism of knowledge graph embedding to analyze, explain, and generalize previous models. We also draw the connections to word embeddings and language modeling to provide some new insights and generalizations.

For the second part of the thesis, we study how to apply multi-relational embedding in representation and analysis of knowledge graphs. Unlike word embedding, the semantic structures such as similarity and analogy structures in knowledge graph embedding space are not well-studied, and thus not usually utilized for data representation and analysis. To demonstrate the application of multi-relational embedding, we formalize a framework for data representation and analysis by semantic queries on the multi-relational embedding space. We build a knowledge graph from scholarly data and show how various tasks on the original datasets can be approximated by appropriate semantic queries, which are multi-linear algebraic operations on the multi-relational embedding spaces. We also theoretically study the entity analogy reasoning task in multi-relational embedding space, which can be formulated as an open-relational query by examples task, doing relational query on unseen relations. Using the above mathematical connections between knowledge graph embeddings and word embeddings, we analyze the semantic structures in the knowledge graph embedding space and propose potential solution to the above entity analogy reasoning task. The goal of this endeavor is to explore potential applications of recent advancements in multi-relational embedding to data representation and analysis, especially to improve its effectiveness on scholarly data.

博士論文審査結果

Name in Full
氏名 Hung Nghiep Tran

Title
論文題目 Multi-Relational Embedding for Knowledge Graph Representation and Analysis

本学位論文は、多重グラフの埋め込み表現 Multi-Partition Embedding Interaction (MEI)を提案している。多重グラフは、知識グラフ、テキスト中の語の相互関係、学术论文の各エンティティの関係など、様々なデータを表現できる。MEIは、グラフのノードの特徴を表すベクトルを複数の独立な部分に分割し部分ベクトル同士の類似度の和によってノード間の類似度を求めるもので、ノード間の複数の関係に基づいた類似度を効率的に求めることを可能にしている。本論文は、MEIが記述力の高い多重グラフの埋め込み表現であり、いくつかの知識グラフ埋め込み表現がMEIの特殊例となっていること、MEIのパラメタを調整することで対象とする多重グラフに適した埋め込み表現が求められることを示している。また、MEIを学術情報のエンティティ検索に応用し、エンティティ間の多様な関係に基づいた検索を統一的に扱えることを示している。

博士論文は7章で構成され英語で書かれている。

第1章では、本研究が扱う多重グラフについて本研究で取り組んだ課題および主な貢献を示している。

第2章では、多重グラフの例、多重グラフの埋め込み表現、本研究で用いるテンソルの概要などの本研究の背景を述べている。

第3章では、本博士論文の関連研究である知識グラフ、テキスト、グラフの埋め込み表現についてまとめている。続く3つの章で本論文の主たる貢献を述べている。

第4章では多重グラフの埋め込み表現法 MEI を提案している。MEI は多重グラフのノードの特徴ベクトルを複数の部分ベクトルに分割し、ノード間の類似度を部分ベクトル間の類似度の和と定義する。既存の4つの知識グラフの埋め込み表現は MEI の特殊ケースであり、分割数や部分ベクトル間の類似度の計算に一定の制約を課すことで表現できることを示している。

第5章では、知識グラフのリンク補完問題に MEI を適用し、その性能評価を行っている。評価用データセットとしてこの分野でよく用いられる2種類の知識グラフを使用し、11の既存手法と比較実験を行い、MEIが既存手法と比較し高い補完性能を有していることを示している。

第6章は、MEIを学術情報の検索に応用している。まず、学術情報として、Microsoft Academic Graph (MAG)を用い、MAGより著者や論文といった学術情報のエンティティをノードとする多重グラフを構築している。次に、MEIによりこの多重グラフの埋め込み表現を求め、著者や論文といったエンティティをそれらの間の関係に基づいて検索する問題を解いている。また、学術エンティティの検索性能の評価実験を行ない、MEIが他の埋め

込み表現と比較して高い検索性能を有することを示している。

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。続いて審査員による質疑応答を行い、MEIにおける分割数と性能の関係やその最適分割数の求め方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、トップ国際会議に1編(ECAI2020)、査読付き国際会議に2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。