

氏 名 蛭子 琢磨

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 2159 号

学位授与の日付 2020 年 3 月 24 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 Integration of Rules and Embeddings for Knowledge Graph
Completion

論文審査委員 主 査 准教授 市瀬 龍太郎

教授 相澤 彰子

教授 武田 英明

助教 小林 亮太

准教授 村田 剛志

東京工業大学 情報理工学院 情報工学系

(様式3)

博士論文の要旨

氏 名 蛭子 琢磨

論文題目 Integration of Rules and Embeddings for Knowledge Graph Completion

Knowledge graphs are useful for many artificial intelligence tasks but often have missing data. However, to manually construct a knowledge graph, a lot of human power is required. Hence, a method for constructed knowledge graphs has been developed. The approaches for knowledge graph construction can roughly be categorized into two groups: The external approach and the internal approach.

The external approach makes use of external resources which include human knowledge, semi-constructed data, and natural language texts. Those resources are converted into triples to construct a knowledge graph. A lot of knowledge graphs, including Wikidata and Freebase, have been constructed and we can use them for our purpose now. However, knowledge graphs often have lots of missing facts. To solve this problem, we needed another approach.

The internal approach predicts missing fact making use of an existing knowledge graph, which includes embedding models, the Path Ranking Algorithm, and rule evaluation models. We think there are two necessary things that a model for a knowledge graph completion model should have: Reliability of Prediction and Interpretability. However, each method has each limitation. For example, rule-based models do not have the reliability of prediction as much as state-of-the-art embedding models. embedding models are generally like "black-box", i.e., lack interpretability. To achieve models with the reliability of prediction and interpretability, we proposed novel methods dealing with issues above models.

A rule-based model is interpretable. Hence, we attempt to develop it for more efficiency. In this dissertation, we propose GRank, which can effectively useful rules underlying in a knowledge graph. We show GRank is as effective as embedding models, which means we get the efficient interpretable model for link prediction. In other words, the explanation of prediction by rules is credible.

Translation-based models were relatively clear with its mechanism among embedding models. However, TransE, the original translation-based model, suffers from problems coming from its low expressiveness. We deal with the low expressiveness in two steps by generalizing TransE. To solve part of the problem, we propose the knowledge graph embedding model on a Lie group (KGLG), which is a generalized translation-based embedding model where entities and relations are represented by a point of a Lie group. KGLG relaxes the problem coming from the characteristics of a Euclidian space for the embedding space. Attention KGLG (AKGLG) is a more generalized version of KGLG.

We propose AKGLG to make KGLG more expressive.

In AKGLG, entities and relations are assigned an attention vector in addition to an embedding on a Lie group.

The attention vector is used to separate the Lie group to separately store information about each entity and relation. We also show AKGLG is unified embedding models that contain many of state-of-the-art embedding models. As a result, we can obtain efficient translation-based models that have a clear mechanism. The rule evaluation method on embeddings (REE) is a method to evaluate association rules from embeddings of AKGLG. REE is proposed based on the clear mechanism of AKGLG. We can use REE to understand what instances of AKGLG learned, in other words, REE provide interpretability for AKGLG. We can also use REE for faster evaluation of rules than GRank.

Finally, we propose the path-based framework (PBF), which is a framework to integrate several approaches for link prediction. In this framework, embeddings of entities and relations are firstly learned from an instance of AKGLG. Then, extract useful rules from the embeddings to construct softmax regression models. Then the softmax regression models and the embeddings are simultaneously used to score entities for link prediction. In this framework, each approach that has been separately developed so far is collaboratively working and shortcomings of them are compensated by other approaches. As a result, PBF outperforms existing models in terms of link prediction and additionally have interpretability because its component has it.

In this dissertation, it became possible to make more reliable triple prediction by higher accuracy. Also, by focusing on interpretability, unlike existing studies, it became possible to provide information to help humans verify the predicted knowledge. In the future, it is expected that this study helps computers to assist human decision-making and to build knowledge in cooperation with human experts.

博士論文審査結果

氏名 蛭子 琢磨

論文題目 Integration of Rules and Embeddings for Knowledge Graph Completion

知識グラフは、エンティティと呼ばれる記述対象とその関係を有向グラフで表現したものであり、質疑応答や検索など様々な AI システムで利用されている。その有用性から、様々なデータリソースを用いて、巨大な知識グラフが作成されてきた。しかし、作成された知識グラフは不完全であり、それを補うための新しい知識グラフ補完手法が必要となる。本博士論文では、主に既存知識グラフから新しい知識を予測する内的な手法について研究を行った。内的な手法は、ルールを用いて知識を補完するルールベース手法と、知識グラフの構造をモデル化して新しい知識を予測する埋め込み手法がある。既存のルールベース手法は予測精度が低い問題点があり、埋め込み手法は、予測理由を知ることができない問題点がある。本博士論文では、それぞれの手法を統合することで、これらの欠点の克服を試みた。

本論文は、全7章からなる。第1章「Introduction」では、知識グラフにおいて知識を補完する問題に対して、研究の背景を述べると共に、本論文の貢献について説明している。

第2章「Fundamentals and Related Work」では、本論文の前提となる基礎的な事項と関連研究についてまとめている。最初に、知識グラフとそれを補完するエンティティ予測タスクを定義し、その後、知識グラフ補完手法の関連研究を述べ、これまでの研究で残された課題について議論し、本論文で取り組む課題の明確化を行っている。

第3章「Refinement of Rule-based Method」では、ルールベース手法における新しいルールの評価関数を用いたモデル GRank を提案し、エンティティ予測タスクに対して、埋め込み手法と同程度の精度を達成できることを実験的に示した。

第4章「Knowledge Graph Embedding on Lie Group with Attention」では、埋め込み手法の1つであるトランスレーションモデルは、ルールを学習するモデルであるという仮説に基づき、ルールを効率的に学習できるように埋め込み手法を発展させた。従来のモデルでルール学習を妨げる因子を取り除くと共に、モデルの抽象化を行い、任意のリー群を埋め込み空間とする KGLG を提案した。さらに、埋め込み表現に、重みを追加し、多種多様なエンティティ、関係を学習できる AKGLG を提案し、AKGLG が多くの既存モデルを包含するモデルであることを示した。

第5章「Integration of Approaches for Link Prediction」では、ルールベース手法と埋め込み手法を統合した。まず、前述の仮説に基づき、AKGLG が学習したルールを復元する手法を提案し、エンティティの予測理由を説明しながら、従来のルールベース手法より高速にルールを抽出できる手法を提案した。そして、ルールベース手法と埋め込み手法を組み合わせて予測可能なフレームワーク PBF を提案した。

第6章「Discussion」では、提案したモデルとフレームワークである GRank, AKGLG,

PBF について総合的な考察を行い、この論文の到達点、限界点などを議論した。

第7章「Conclusions」では、博士論文の総括を行うと共に、展望を述べ、結論をまとめた。

上記のように、本博士論文は、既存の知識グラフ上で、新しい知識を自動的に予測するため課題を提示し、その解決手法を示した点で、この研究分野の発展に貢献するものである。また、本論文で示した考え方により、これまで個別に研究をされていたルールベース手法と埋め込み手法の関係を明らかにし、結果として、説明性を担保しながら精度高く知識グラフを補完できる手法を示しており、知識処理のための基盤技術開発という観点からも意義が認められる。さらに、博士論文の内容は、1本の査読付きジャーナル論文、2本の査読付き国際会議論文で発表されており、社会からも評価されている。以上より、本論文は博士論文として、十分な水準であると審査委員全員一致で認められた。