

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺 太郎 (教授))		
学籍番号	1911224	提出日	令和 3年 1月 25日
学生氏名	廣瀬 雄士		
論文題目	Study of integrating embedding and relation path for knowledge graph completion 知識グラフ補完のためのエンベディングとリレーションパスの統合に関する研究		
要旨			
<p>実世界の知識をデータ化した知識グラフは一般に不完全で、欠損したトリプレットを予測するタスクは知識グラフ補完と呼ばれる。主要な予測モデルとしてエンティティやリレーションのエンベディングを用いるエンベディングモデル、エンティティ間のリレーション系列を用いるリレーションパスモデルが存在する。両者には長所と短所があり、エンベディングモデルは、知識グラフ全体から特徴を学習できるが、知識グラフの経路情報のモデル化が明示的に行われていない。リレーションパスモデルは、偽陽性が少ないが、エンティティ間にリレーション系列が存在しなければそのトリプレットを予測することはできない。本研究ではエンベディングとリレーションパスを統合するような二つの手法を提案する。一つ目の手法ではエンベディングモデルとリレーションパスモデルの両者のスコアを用いるもので、実験の結果、予測精度はあまり向上しないことがわかった。二つ目の手法ではトランスダクティブ学習設定でリレーションパスのルールを用いたデータ拡張を行う。拡張データを用いてエンベディングモデルの学習・予測評価を行ったところ、精度が向上した。分析の結果、リレーションパスルールによって、真のデータを予測する、または真のデータに類似したデータを予測することができたことが精度向上に寄与したことがわかった。今後の課題として、手法1はアンサンブル学習などの他の学習法、手法2は拡張データを負例から取り除くことが考えられる。</p>			