

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	2311194	提出日	令和 7年 1月 21日
学生氏名	中谷 響		
論文題目	POIジオコーディングのための対照学習を利用したテキスト埋め込みモデル		
要旨			
<p>ジオコーディングは、位置に関する言及をその地理的な位置に結びつける基本的な技術であり、記述されている出来事がどこで起こったのかを理解する上で重要な役割を果たす。従来のジオコーディング研究が位置の大まかな特定に重点を置いていたのに対し、本稿はより細粒度なpoints-of-interests (POI) レベルでのジオコーディングに着目し、類似したPOI名が多数存在する中で適切な地理データベースエントリを見つける課題に取り組んだ。この課題を解決するため、(1) エントリのエンコード表現、(2) モデルの曖昧性解消能力を強化するためのhard negativeマイニング手法に焦点を当てたテキスト埋め込みに基づくモデルを開発した。評価実験により、二つ目のアプローチがモデルのジオコーディング精度向上に大きく寄与することを示した。また、モデルの学習を支えるデータセットとして、Wikipediaの記事に出現するメンションとそれに紐付けられたWikidataエントリの収集を行い、大規模なデータセットを構築した。このデータセットを活用することで、異なるドメインのデータに対してもモデルが高精度で解析可能であることを確認した。</p>			