

中央大学特定課題研究費 ー研究報告書ー

所属	理工学部	身分	教授
氏名	難波 英嗣		
NAME	Hidetsugu Nanba		

中央大学特定課題研究費による研究期間終了に伴い、中央大学学内研究費助成規程第15条に基づき、下記のとおりご報告いたします。

1. 研究課題

格文法と言語モデルを用いた特許オントロジーの自動構築

2. 研究期間

2021・2022年度

3. 費目別収支決算表

掲載省略

4. 研究の概要（背景・目的・研究計画・内容および成果 和文 600 字程度、英文 50word 程度）

（和文）

【背景】 自然言語処理において、上位下位語関係は知識体系の中核であり、多くの下流タスクに役立つ。特許を対象とした技術動向調査や審査等はその一例である。しかし、このような用語間の関係を人手で整備するには非常にコストがかかる。

【目的】 自然言語処理技術を用いて特許データベースから専門用語の用語間の関係を解析し、特許や論文等の専門文書を執筆する際に有用な情報源として活用される特許オントロジーを自動構築する。

【研究計画】 文構造に関する言語理論のひとつである格文法と、近年着目を集めている言語モデル BERT を用いることで、正確に上位下位関係を抽出し、同時にファセット(属性)も抽出する手法を提案する。

【内容】 本研究では、以下の2つの手法でオントロジーを自動構築した。

(手法 1) まず、GAN(Generative Adversarial Network)を用いて上位下位関係にある用語を識別する。次に、前の手順で構築された上位下位関係のグラフに対し、ConvE と GraphSAGE を組み合わせたリンク予測を行い、本来であれば上位下位関係にあるべき欠落したエッジを予測する。

(手法 2) 定型表現「A 等の B」を用いて、上位下位関係の用語対を抽出する。抽出した用語対に対して、特許文による事前学習済み BERT を用いて文脈を考慮したベクトル表現を獲得する。次に、得られたベクトル表現に対して GAN によるデータ拡張を行い、学習後 Discriminator を分類器として上位下位関係の有無を判定する。

【成果】

(手法 1) GAN を用いた上位下位関係の識別および ConvE と GraphSAGE を組み合わせたリンク予測の両方において、提案手法が従来手法よりも優れていることがわかった。

(手法 2) GAN を用いた提案手法は、SVM やロジスティック回帰を用いた従来手法よりも再現率を大幅に向上させる効果があることがわかった。

（英文）

In this paper, we propose the following two methods to extract hypernym-hyponym relations from patent text data:

(Method 1) Construct a graph of hypernym-hyponym relations using a linguistic pattern and estimate missing edges in the graph.

(Method 2) Identify hyponymy between terms using linguistic patterns and BERT.

We experimentally confirmed the effectiveness of our methods.