

S17 GANを用いた特許からの上位下位関係の抽出

岩熊耕平, 難波英嗣 (中央大学)

研究背景・研究目的

■上位下位関係は用語間の情報として有用

例) 文献の検索, オントロジー構築

■テキストからの自動抽出が提案されている

- 「等の」のような定型表現に注目



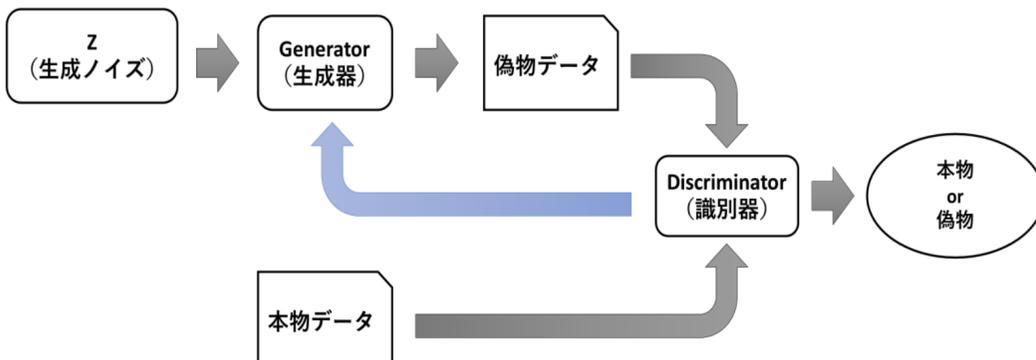
- 誤った関係性まで抽出してしまう



目的

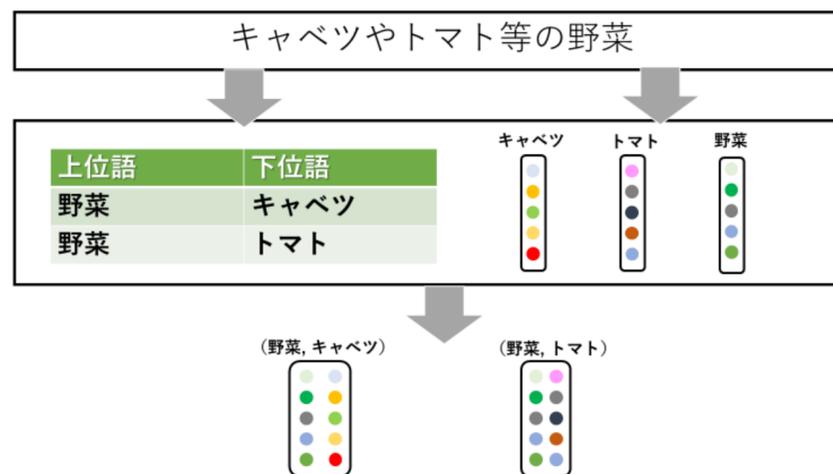
GAN及びBERTを用いることで、日本語特許から正しい上位下位関係を自動で抽出する

GAN(敵対性生成ネットワーク)

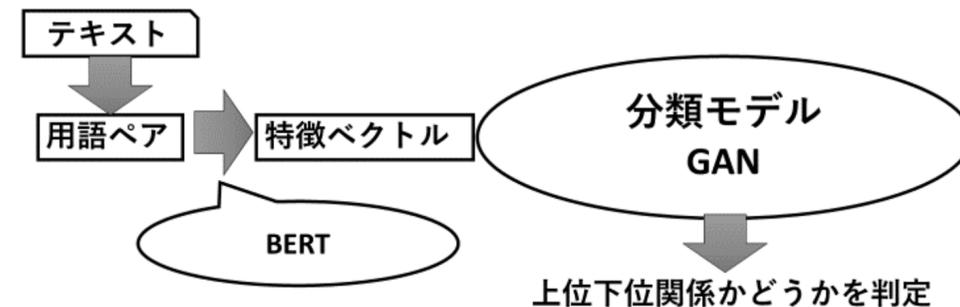


提案手法

- テキスト内の上位語・下位語候補を抽出
 - ✓ 日本語特許中の「などの」「等の」を含む文から上位語・下位語のペアを獲得する
 - ✓ 各ペアは自然言語処理ライブラリspaCyを用いて形態素解析を行い抽出
- 特徴ベクトルの生成
 - ✓ 獲得した用語対の特徴ベクトルをBERTを用いて生成
 - ✓ 各用語の特徴ベクトルは768次元で出力



- GANによる上位下位関係の識別
 - ✓ 獲得した特徴ベクトルを基にGANを用いた分類モデルを生成
 - 学習後GANにおけるDiscriminator (識別器) を分類器として用いる
 - ✓ 学習したモデルを用いてテキストから抽出された用語ペアが上位下位関係かどうかを予測



今後の課題

分類モデルの実装及び評価
その他関係性を考慮した用語抽出

参考文献

- [1] Marti A. Hearst, "Automatic Acquisition of Hyponyms from Large Text Corpora," Proceedings of the 14th International Conference on Computational Linguistics, p. 539-545, (1992)
- [2] Thomas Kober, Julie Weeds, Lorenzo Bertolini, and David Weir, "Data Augmentation for Hypernymy Detection", Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, p.1034-1048, (2021)