

玄道 俊 松下 光範 山西 良典 (関西大学)

背景と目的

共同利用可能なデータセットを用いて研究を行うことがトレンド

研究者のメリット

コストがかからず大規模データを使用が可能



データ提供者のメリット

先進技術をいち早く事業に取り入れることが可能



論文を特徴化し、共同利用可能なデータセットの使い方を可視化する

実現可能な未来

研究者の未来

- 自身の研究と目的が同様で、異なるデータセットを用いた研究を探ることが可能
- 研究を始める際、自身の立ち位置を明確化するために使用可能

データ提供者の未来

- 提供したデータセットの使用方法を把握することが可能
- これからデータを提供する提供者は、自身が保有するデータセットと類似しているデータセットを用いた論文から出力結果を予測可能

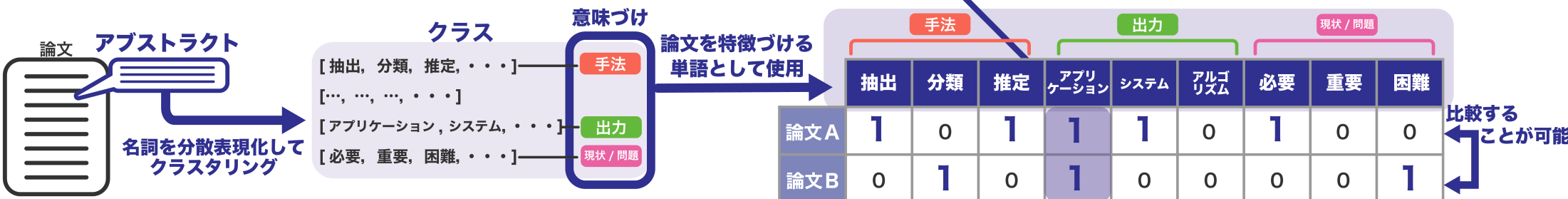
提案手法

論文内で使用したデータセットのスキーマを記載したデータセットを作成

アブストラクト内で出現した単語を基に、バイナリベクトルを用いて論文を特徴化

作成した論文データセット

	アブストラクト	データ1	データ2	データ3	データ4
論文 A	本研究では、...	レシピタイトル	手順の位置	手順の内容	
論文 B	近年、料理が...	レシピタイトル	材料の名前	材料の分量	手順の内容



実施例

提案手法を用いて論文を特徴化した実施例を QR コードのリンク先に掲載しています。



謝辞

この研究は 2021 年度国立情報学研究所公募型共同研究 (21S0501) の助成を受けています。

背景

共同利用可能なデータセットを用いて研究を行うことがトレンド



コストがかからずデータが使える！



先進技術をいち早く取り入れられる！

論文を特徴化し、共同利用可能なデータセットの使われ方を可視化すれば可能性が広がる！

実現可能な未来

例えば・・・



自分の研究と同じ目的だけど、違うデータセットを使っている研究を探することができる！

提供したデータセットの使用方法を把握することができる！



提案手法

論文内で使用したデータセットの**スキーマ**を記載したデータセットを作成

アブストラクト内で出現した単語を基に、**バイナリベクトル**を用いて論文を特徴化

分散表現 + クラスタリング

比較する
ことが可能

	抽出	分類	推定	アプリ ケーション	システム	アルゴ リズム	必要	重要	困難
論文A	1	0	1	1	1	0	1	0	0
論文B	0	1	0	1	0	0	0	0	1