

新しい社会システムデザインに向けた情報基盤技術の創出  
2018 年度採択研究者

2020 年度 年次報告書
------------------

加藤 誠

筑波大学 図書館情報メディア系  
准教授

オープンデータ利活用のためのデータ検索エンジンの構築

## § 1. 研究成果の概要

2020年度には、主にデータランカの課題8(事実検証型クエリ), ならびに、データランカの課題9(データ質問型クエリ), データサマライザの課題10(典型的なデータ要約方法の発見)に取り組み、データ検索エンジンが提供する検索機能およびデータ要約についての研究開発を行った。

データランカの課題8については、情報アクセスのための評価キャンペーンであるNTCIR-15においてData Searchタスクを運営し、事実検証型クエリのデータセットの構築および検索手法の分析を行なった。本試みでは、日英の2つの言語を対象としたデータセットを構築し、両言語に対して訓練用およびテスト用のトピックをそれぞれ96個用意した。本タスクには、主催者チームを含む合計6チームが参加し、合計で74検索システムの結果が得られた。提出された検索システムのアルゴリズムを分析したところ、表ヘッダ、ニューラル言語モデル、コミュニティ質問応答サイトにおけるカテゴリ情報の利用などが主に有効であったことがわかった。

データランカの課題9では、表形式データの見出しの認識および見出しの階層認識タスクに取り組んだ。この問題に対して、表形式データを表現するスプレッドシートをグラフデータに変換し、グラフニューラルネットワークを適用することで、スプレッドシートの大域的な特徴を考慮して見出しの認識を行う方法を提案した。また、教師なし表現学習を行うことによって有効なセル表現を学習し、少数の教師データであっても効果的な学習が可能となる方法を提案した。

データサマライザの課題10では、データに対して適切な要約方法を発見するというタスクに取り組んだ。このタスクではクエリと表形式データが与えられており、我々は両者の関係性から互いのどの部分に着目して要約方法を推定すべきかを決定するような、双方向アテンションを用いたモデルを提案し、その有効性を示した。

### 【代表的な原著論文情報】(最大5件)

- 1) Makoto P. Kato, Wiradee Imrattanatrai, Takehiro Yamamoto, Hiroaki Ohshima, Katsumi Tanaka: Context-guided Learning to Rank Entities. Proceedings of the 42nd European Conference on IR Research (ECIR 2020), pp. 83-96, 2020.
- 2) 藤岡 周平, 加藤 誠, 吉川 正俊: 表からの量的データ属性間の関係抽出. 情報処理学会論文誌データベース(TOD) 13, pp. 10-21, 2020.