

数理・情報のフロンティア
2019 年度採択研究代表者

2020 年度 年次報告書

磯沼 大

東京大学大学院 工学系研究科
大学院生(博士課程)

談話構造に基づく教師なし生成型要約

§ 1. 研究成果の概要

自動要約は膨大な文書からの知識抽出に有用な技術だが、従来法は大量の教師データ(10 万件単位の見本の要約)を必要とするため実用性に乏しい。この問題意識から、本研究は文書の潜在構造、特にトピック構造に着目した教師なし要約生成手法の開発に取り組んでいる。

本年度では、前年度に開発した木構造トピックモデルにより文書のトピックとその木構造を推定し、各トピックの要約文を生成する手法を開発した。木構造トピックモデルでは、文書から推定したトピックが木構造に配置され、子が親のサブトピックとなる構造を持つ。こうした特性は、様々な粒度の情報抽出に有用であり、教師なし要約生成に応用すれば、根からは文書全体に関する要約文が、葉に近づくにつれより詳細なトピックに関する要約文が得られることが期待される。

得られた要約と人手で作成した要約を比較評価した結果、既存の教師なし要約手法より元文書の内容を網羅し、かつ一貫性の高い要約が得られることが確認された。また得られた木構造について、親子関係にある要約文は一貫したトピックを持ち、葉に近づくにつれより詳細なトピックに関する要約文が得られるといった性質が定量的に確認された。以上の成果は、計算言語学の国際ジャーナル TACL に採択された他、国内学会・研究会で発表を行い受賞した。

【代表的な原著論文情報】

- 1) Masaru Isonuma, Junichiro Mori, Danushka Bollegala, Ichiro Sakata. “Unsupervised Abstractive Opinion Summarization by Generating Sentences with Tree-Structured Topic Guidance.” Transactions of the Association for Computational Linguistics, 2021. (to appear)