

磯沼 大

東京大学大学院工学系研究科
大学院生（博士課程）

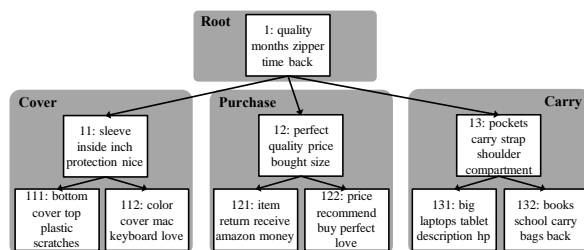
談話構造に基づく教師なし生成型要約

§ 1. 研究成果の概要

本研究では、文書のトピック構造を捉え、各トピックに関する要約文を教師データなしに生成する手法の開発に取り組んでいる。当年度では、文書のトピック構造を捉える手法として木構造トピックモデルに着目し、文書要約への応用にあたりその学習スピードの向上に取り組んだ。

木構造トピックモデルは、文書から推定したトピックを木構造に配置し、子のトピックが親のサブトピックとなる構造を持つ（右図）。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。

既存の木構造トピックモデルは大規模文書への適用が困難なことから、本研究では並列学習が可能な木構造トピックモデルを開発した。提案法は従来法に対し学習時間を約 15 倍短縮し、要約生成など、大量の文書を要するアプリケーションへの応用が可能である。



本研究成果は計算言語学分野の国際会議 ACL2020 に short paper として採択された。