

論文 / 著書情報  
Article / Book Information

題目(和文)	高品質教師データが得られないドメインにおける要約手法の研究
Title(English)	
著者(和文)	狩野竜示
Author(English)	Ryuji Kano
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11758号, 授与年月日:2022年3月26日, 学位の種別:課程博士, 審査員:奥村 学,熊澤 逸夫,中山 実,篠崎 隆宏,船越 孝太郎
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11758号, Conferred date:2022/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

## 論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	狩野電示	
論文審査 審査員		氏名	職名	氏名	職名
	主査	奥村学	教授	船越孝太郎	准教授
	審査員	熊澤 逸夫	教授		
		中山実	教授		
篠崎隆宏		准教授			

### 論文審査の要旨 (2000 字程度)

本論文は「高品質教師データが得られないドメインにおける要約手法の研究」と題し、全5章より構成されている。

第1章「序論」では、既存の要約研究の課題として、新聞記事などの高品質な教師データが整備されているドメインに特化して行われてきたことをまず指摘している。また、高品質な教師データが存在しないドメインにおける要約手法として、2つのアプローチがあることを述べている。1つ目は教師データを使用しない教師なし手法であり、2つ目は品質の低い教師データを使って効率的に学習を行う手法である。そして、本論文ではこれら2つのアプローチにおいて、それぞれ新しい手法を提案していることを説明している。

第2章「関連研究」では、これまでの要約モデルの発展に触れながら、上述した2つのアプローチに関して、これまでにどのような関連研究が行われてきたかを詳細に論じている。

第3章「会話文における暗黙的引用を利用した抽出型教師なし要約」では、1つ目のアプローチである「教師なし要約」として提案する手法について論じている。まず、既存の教師なし要約手法の課題が述べられ、既存手法とは異なる要約の指標に基づいたモデルが提案されている。既存の教師なし要約手法の多くは、高頻度に言及される話題は重要であるという仮定に基づき、頻度を元に要約としての文のふさわしさを定量化している。しかし、実際には重要な話題が必ずしも多く言及されるとは限らない。そこで新たな手法では、返信による言及のされやすさを重要度の指標として提案している。提案したモデルは、要約対象の本文と返信の対を用いて、正しい返信とランダムサンプリングにより得られた偽返信を判別する学習を行う。その際、Gumbel Softmaxにより本文の一部の文を抽出し、正しい返信と偽返信の判別に使用する。モデルは学習の過程で返信によって言及されやすい文を本文から抽出するようになるため、これを要約とみなして出力することができる。提案手法をメールデータセット及び、ソーシャルメディアデータセットで評価し、従来手法と同等もしくは上回る性能を発揮することを確認している。また、提案手法が、従来手法では抽出できない重要文を抽出できていることを定量的、定性的両方の分析で確認している。

第4章「要約データの適切性定量化を利用したカリキュラムラーニング」では、2つ目のアプローチである「低品質教師データの効率的学習」として、カリキュラムラーニングを要約データセットに利用した手法を提案している。カリキュラムラーニングは、ある指標に基づき学習するデータの順序を変更することでモデルの性能を上げる手法であり、ノイズを含むデータセットに対する有効性が示されてきた。しかし、既存手法では、高品質教師データと低品質教師データ両方が必要という課題があったため、カリキュラムラーニングを要約タスクに応用するにあたって、ノイズを含むデータセットからノイズを定量化する手法を提案し、既存手法よりも高い性能を発揮することをメール及びソーシャルメディアデータセットを用いて確認している。また、これまでカリキュラムラーニングに用いられてこなかった抽出率や含意判定確率が要約データセットに対するカリキュラムラーニングに有効であることも合わせて示している。

第5章「結論と今後の課題」では第3章と第4章で提案した手法に対する課題と今後の展望について述べている。

以上を要するに、本論文は、要約技術が適用可能な対象を教師データが未整備のドメインに拡大するための新しい手法を2つの方向性で提案し、その有効性を検証しており、工学上貢献するところが大きい。よって博士(工学)の学位を授与するに十分な価値を持つものと認められる。

注意:「論文審査の要旨及び審査員」は、東工大リサーチポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。