

論文 / 著書情報
Article / Book Information

題目(和文)	Data-to-Textモデルの高度化に関する研究
Title(English)	
著者(和文)	村上聡一郎
Author(English)	Soichiro Murakami
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12190号, 授与年月日:2022年9月22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,白井 清昭
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12190号, Conferred date:2022/9/22, 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		村上 聡一郎	
		氏名	職名		氏名	職名	
論文審査 審査員	主査	奥村学	教授	審査員	船越孝太郎	准教授	
	審査員	熊澤 逸夫	教授		白井清昭	准教授	
		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「Data-to-Text モデルの高度化に関する研究」と題し、全5章より構成されている。

第1章「序論」では、ニューラル言語生成技術に基づく Data-to-Text モデルは、制御が難しく入力データの事実と異なる内容や重要性の低い内容のテキストが生成される問題があることをまず指摘している。また、既存研究の方向性として、モデルを高度化する取り組みがあり、その中心として表データからのテキスト生成等で用いられてきた内容選択モデルやコピー機構によるアプローチがあることを述べている。さらに、これらのアプローチは、入力データと出力テキストが表層単語に基づいて対応づくことを前提とした手法であることから、数値や画像等を入力として扱うタスクに適用することが難しいことを述べている。そして、本論文では時系列数値データを入力として扱うタスクとして株価の市況コメント生成および天気予報コメント生成を例に挙げ、上述の2つのアプローチが抱える問題を踏まえた上で、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 Data-to-Text 生成課題の研究動向を述べて、上述した2つのタスクに関して、これまで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「時系列株価データからの市況コメントの自動生成」では、既存研究で用いられてきたコピー機構は市況コメントで言及される株価の増減幅やそれらを切り上げ・切り捨てた価格などの数値表現を生成することができないことをまず指摘している。また、市況コメントには株価の短期的および長期的な値動きに関する言及や市況コメントの配信時間に依存する表現が用いられる特徴があることを論じている。これらの特徴に対して提案したモデルでは、株価の値動きを捉えるために短期的及び長期的な株価データを与え、さらに市況コメントの配信時間を考慮するために時間帯情報を導入している。さらにコピー機構では生成できない数値表現を生成する方法として、数値の演算操作を推定して計算することで株価の数値表現を生成する手法を提案している。実験では、自動評価および人手評価を通して、提案したモデルが上述の特徴を捉えたより質の高い市況コメントを生成できることを確認している。また、提案手法が、従来手法では生成できない数値表現を、多くの事例において正しく生成できていることを定量的および定性的に分析し確認している。

第4章「数値気象予報からの天気予報コメントの自動生成」では、天気予報コメントにおいて「晴れ」「雨」「曇り」「雪」といった情報を適切かつ明示的に言及することが重要であることをまず述べた上で、そのために内容選択モデルを導入することの必要性を述べている。また、既存研究で広く用いられてきた内容選択モデルは、数値データを対象とした本研究へ適用することが難しいことを指摘している。加えて、天気予報コメント生成では、数値気象予報に含まれる複数の物理量やコメントの配信時間および対象エリアを考慮することの必要性を論じている。これらの特徴に対して提案モデルでは、気象データのエンコード手法を比較検討した上で、天気予報コメントの配信時刻や対象エリアを表すメタ情報を導入している。さらに数値データに対する内容選択モデルを導入するために、「晴れ」「雨」「曇り」「雪」の各気象情報に対応する手がかり語に基づいて、内容選択モデルの教師ラベルを作成する方法を提案している。実験では、人手評価および自動評価を通して、内容選択モデルの導入により生成テキストの情報性が改善することを確認している。

第5章「結論と今後の課題」では提案手法に対する結論と課題、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、Data-to-Text モデルにおける生成テキストの正確性の課題に対して、モデルの高度化手法として新たに2つの手法を提案し、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、東工大リサーチポータル(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。