

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)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報通信 情報通信	系 コース	申請学位 (専攻分野)： Academic Degree Requested	博士 Doctor of	(工学)
学生氏名： Student's Name	村上聡一郎		指導教員 (主)： Academic Supervisor(main)	奥村学	
			指導教員 (副)： Academic Supervisor(sub)		

要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

第 1 章「序論」では、様々な形式のデータの概要や特徴を分かりやすく説明するテキストを生成する研究課題、すなわち Data-to-Text 生成課題の研究動向をまず説明する。加えて、Data-to-Text 生成課題において既存研究で広く用いられるニューラルベース手法ではしばしば入力データの事実と異なる内容や重要性の低い内容が言及される問題があることを指摘する。また、それらの問題を解決する方向性として本研究では Data-to-Text モデルの高度化に取り組むことを述べる。

第 2 章「関連研究」では、Data-to-Text 生成課題における研究動向を詳細に解説する。また、生成テキストの正確性の課題に対する既存研究の取り組みを解説した上で、それらの手法の問題点を述べる。具体的には、表データからテキストを生成する課題等で中心に扱われてきた内容選択モデルやコピー機構は、入出力データが単語表層で対応づくことを前提とした手法であるため、数値や画像等のデータを入力として受け取るタスクに適用できない問題があることを指摘する。その後、本研究で取り組む具体的なタスクとして、入力データとして数値データを受け取る「株価の市況コメント生成」や「天気予報コメント生成」の既存研究を述べる。

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、まず株価の市況コメントにおける 3 つの特徴として、(1) 株価の市況コメントでは株価の短期的・長期的な値動きに言及される点、(2) 「前引け」「大引け」のように市況コメントの配信時間に依存する表現が用いられる点、(3) 株価の終値や増減幅等の数値表現が言及される点を説明する。その後、提案手法として、これらの特徴を捉えてテキスト化するためのエンコーダ・デコーダモデルについて解説する。実験では、自動評価および人手評価を通して、ベースライン手法と比較して品質が高く、正確な市況コメントが生成できることを示す。

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、まず天気予報コメントにおける 3 つの特徴として、(1) 天気予報コメントでは降水量や気圧、雲量等の複数の物理量とその時間変化を考慮する必要がある点、(2) コメントでは配信時間や対象エリアに依存した表現が用いられる点、(3) 天気予報コメントでは「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求められる点を説明する。その後、提案手法として、上記の特徴を捉えテキスト化するために内容選択モデルを導入したエンコーダ・デコーダモデルについて述べる。実験では、数値データを対象とした内容選択モデルの導入により、提案手法ではベースライン手法と比べて、情報性の高い天気予報コメントを生成できることを定量的および定性的な分析を通して示す。

5 章「結論と今後の課題」では、本論文のまとめを行い、3 章および 4 章で取り組んだ研究の今後の課題に加え、Data-to-Text 生成課題全体における今後の方向性について議論する。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(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。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

(博士課程)
Doctoral Program

論文要旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報通信 情報通信	系 コース	申請学位 (専攻分野)： Academic Degree Requested	博士 Doctor of	(工学)
学生氏名： Student's Name	村上聡一郎		指導教員 (主)： Academic Supervisor(main)	奥村学	
			指導教員 (副)： Academic Supervisor(sub)		

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

In this dissertation, we address the problem of data-to-text generation, which is the task of automatically producing description from non-linguistic data such as structured or numeric input data. Specifically, data-to-text models based on neural networks are known to suffer from the hallucination problem in that they are prone to output descriptions that are false or inconsistent with the facts of input data. To address the problem, we design and enhance an encoder-decoder model that appropriately captures the characteristics of the input data to accurately generate descriptive texts. In this study, we tackle two tasks of generating text from numerical data: market commentary generation and weather forecasts generation. These tasks require generating comments that accurately convey numerical values, such as movement of stock prices, or weather information, such as sunny or cloudy. For the former task, we allow an encoder-decoder model to choose an arithmetic operation such as subtraction or rounding instead of generating a word. For the latter task, we introduce weather labels representing weather information, such as sunny and rain, to help an encoder-decoder model explicitly describe useful information and improve the correctness. Experimental results show that the proposed method contributes to improving the quality of generated text in each task. This dissertation concludes with a discussion of data-to-text models that consider external information as a direction for future research in the field of data-to-text generation.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(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。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).