



Title	深層生成モデルに基づく対話型マルチメディア情報検索に関する研究 [論文内容及び審査の要旨]
Author(s)	柳, 凜太郎
Citation	北海道大学. 博士(情報科学) 甲第16016号
Issue Date	2024-03-25
Doc URL	http://hdl.handle.net/2115/92395
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	theses (doctoral - abstract and summary of review)
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	Rintaro_Yanagi_review.pdf (審査の要旨)



[Instructions for use](#)

学位論文審査の要旨

博士の専攻分野の名称 博士 (情報科学) 氏名 柳 凜太郎

審査担当者 主査 教授 長谷山 美紀
副査 特任教授 荒木 健治
副査 特任教授 坂本 雄児
副査 教授 土橋 宜典
副査 教授 小川 貴弘

学位論文題名

深層生成モデルに基づく対話型マルチメディア情報検索に関する研究

(A study on interactive multimedia information retrieval based on deep generative models)

本論文は、深層生成モデルに基づく対話型マルチメディア情報検索に関する研究の成果をまとめたものである。

視覚情報と言語情報の対応関係を学習することで、テキストのクエリからユーザの必要とする画像や映像を検索する技術 (以降、マルチメディア情報検索) は、ユーザがクエリを簡便に用意可能であることから様々な研究が行われている。一方で、マルチメディア情報検索はテキストおよび画像・映像という異なる知覚情報を比較することの困難さやテキスト情報に含まれる曖昧性から未だ十分な検索精度に到達していない。

従来のマルチメディア情報検索手法では、クエリテキストおよび検索候補を同一の特徴量空間に射影することで、高精度に目的の画像・映像を検索する。従来手法は、目的の画像・映像を一意的に特定可能なクエリテキストが与えられることを前提に設計されている。そのため、検索候補中にクエリテキストに該当する画像・映像が複数存在する場合、一度の検索で高精度な検索結果を得ることは困難である。一般的に、ユーザは検索目的の画像・映像を想起しながらクエリテキストを考案するため、クエリテキストには十分な情報が含まれていない可能性が高い。そのため、クエリテキストに不足している情報を補完することで、検索順位を改善することが可能な技術について検討する必要がある。

クエリテキストに不足している情報を補完するためには、検索候補を絞り込むための手掛かりとなる情報をユーザと検索システムで共有する必要がある。現在のマルチメディア情報検索はユーザにとって内部構造を理解することが困難な (以降、透明性の低い) 深層学習モデルを利用しており、ユーザと検索システムの間で正確に情報を共有することは困難となる。ここで、近年、テキストや画像などを生成可能な深層生成モデルに関する研究が盛んに行われている。このような深層層生成モデルを活用し、ユーザにとって理解が容易なテキストや画像の形式で検索システム側が把握する情報をユーザと共有することで、クエリテキストに不足している情報を容易に補完することが可能となり、効果的な検索順位の改善に繋がると期待される。

そこで本論文では、深層生成モデルによる情報共有を介した、対話型マルチメディア情報検索手法を提案する。本論文では、1) 画像生成モデルを応用したマルチメディア情報検索手法、2) 質問文生成モデルを応用したマルチメディア情報検索手法の2つのアプローチにより、従来研究に存在し

た問題を解決する。マルチメディア情報検索において、「クエリテキストに込められたユーザの検索意図」と「クエリテキストに対する検索システムの解釈」を一致させることは、ユーザの必要とする情報を検索するために重要である。そこで、1)ではユーザにより与えられたクエリテキストを画像に変換し、変換後の画像を「クエリに対する検索システムの解釈」としてユーザに共有し、クエリテキストの修正を支援する手法を考案している。次に、2)では、画像や映像に関連した質問文を生成することが可能な質問生成モデルを用いることで、検索候補を絞り込むために必要な情報を質問応答の形式でユーザに問い合わせる枠組みについて検討している。考案した手法により、ユーザにより与えられたクエリテキストに十分な情報が含まれていない場合にも、ユーザは提示された質問に回答するだけで大幅に検索結果を改善することが可能である。

以下に本論文の構成を示す。第1章では、本論文の研究背景および目的を述べる。第2章では、従来研究を紹介し、本論文で解決すべき課題を明らかにする。第3章では、画像生成モデルに基づくマルチメディア情報検索手法を提案する。第4章では、画像生成モデルに基づいて、クエリテキストの入力を補助することが可能な対話型検索手法を提案する。第5章では、上記により提案した手法が、絵画等の多様なドメインに対しても応用可能であることを確認する。次に、第6章では、定型文形式の質問文を生成可能な再検索手法を提案する。第7章では、質問生成モデルに基づいて、非定型な質問文を生成可能な再検索手法を提案する。最後に第8章では、本研究の成果を要約し、論文全体のまとめとする。

以上を要約すると、本論文は、検索システム側がユーザに種々の情報を提示することが可能な枠組みを深層生成モデルに基づいて構築することで、ユーザの検索行動を効果的に補助することが可能なマルチメディア情報検索を実現している。この貢献は、情報科学分野の発展に寄与するものと認められる。したがって、本論文における著者は、北海道大学博士(情報科学)の学位を授与される資格を有するものと認める。