



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
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16016号
Issue Date	2024-03-25
Doc URL	https://hdl.handle.net/2115/92395
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Rintaro_Yanagi_abstract.pdf, 論文内容の要旨



学位論文内容の要旨

博士の専攻分野の名称 博士（情報科学） 氏名 柳 凜太郎

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。

現在、画像・映像に代表されるマルチメディアコンテンツは個人の携帯端末上や Web 上に広く存在しており、そのデータ量は爆発的に増加している。このような膨大なデータの中から目的のマルチメディアコンテンツを簡便かつ正確に検索可能とする技術は、データを探索する時間を大幅に削減することが可能であることから、様々な研究がなされている。特に、視覚的な情報と言語的な情報の関連性を学習することで、テキストのクエリからユーザの所望する画像・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盛んに研究が行われている。従来研究によりマルチメディア情報検索の高精度化が進む一方で、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から未だ十分な検索精度に達していない。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し、両者を比較することで、目的のマルチメディアコンテンツを検索する。上記の枠組みにより、クエリテキストに関連したマルチメディアコンテンツを高精度に検索可能な枠組みが実現されている。一方、従来の手法は、クエリテキストから目的のマルチメディアコンテンツを一意に特定可能であることを前提に設計が行われている。そのため、クエリテキストに該当する検索候補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。ユーザは検索候補の内容を把握しておらず、検索目的のマルチメディアコンテンツ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、マルチメディア情報検索技術の更なる高精度化には、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。一方で、近年のマルチメディア情報検索は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。ユーザと検索システムの間で正確に情報を共有するためには、透明性の低い深層学習モデルの内部構造をそのまま取り扱うのではなく、我々が日常的に取り扱うテキストや画像などのマルチメディアコンテンツを用いることが望ましい。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について提案する。具体的に、深層生成モデルを用いて、検索システム側が把握する情報を画像やテキス

トの形式でユーザに提示することで、ユーザによるクエリテキストの修正を補助し、検索順位を改善する。本論文では、1) 画像生成モデルの導入により、ユーザと検索システムのクエリに対する解釈を共有するマルチメディア情報検索手法、2) 質問文生成モデルの導入により、絞り込みに必要な情報をユーザに要求することが可能な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の解決する。まず、1) では、テキストの内容を反映した画像を生成可能な画像生成モデルを用いることで、ユーザと検索システムの間でクエリテキストに対する解釈を共有することが可能な枠組みについて検討している。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。特に、クエリとしてテキストを利用する場合、クエリに多様な意味合いが含まれやすく、上述した解釈の共有は必要不可欠である。このような解釈の共有を図るため、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。ユーザの与えるクエリテキストに十分な情報が含まれず、検索候補中にクエリテキストに該当するマルチメディアコンテンツ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。上記の課題を解決するために、2) では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検索手法を考案した。考案した手法により、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。以上のように、本論文では、様々な深層生成モデルをユーザと検索システムの情報の共有に活用することで、ユーザによる情報の想起を補助し、検索順位を改善することが可能な技術を実現する。

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、関連研究としてマルチメディア情報検索に関する従来研究を紹介し、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し、解釈の共有が検索の高精度化に貢献することを示す。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、質問応答に基づく対話型検索の枠組みが検索精度の向上に貢献可能であることを確認するために、定型文形式の質問文を生成可能な再検索手法について検討する。第7章では、第6章における検証結果を受けて、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し、検索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ことの有効性を示す。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。さらに、本手法を多様な画像データベースに適用する実験を行うことで、その有効性を示している。