

氏 名 後藤 淳

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 1680 号

学位授与の日付 平成26年3月20日

学位授与の要件 複合科学研究科 情報学専攻  
学位規則第6条第1項該当

学位論文題目 自然言語処理を利用した放送メディアアクセス技術に関する研究

論文審査委員 主 査 教授 山田 誠二  
准教授 相原 健郎  
准教授 宮尾 祐介  
教授 相澤 彰子 国立情報学研究所  
教授 佐藤 真一 国立情報学研究所

論文内容の要旨  
Summary of thesis contents

自然言語処理を利用した放送メディアアクセス技術に関する研究

近年、放送と通信の連携が急速に進んでいる。これまで放送は文字通り放送局から一方的に送られてきた映像や音声を受信し視聴するものであった。ブロードバンド環境の普及により、この状況が変わってきている。2013年3月には、放送と通信の連携を推進する新しい放送方式であるハイブリッドキャストの技術仕様が、放送局やメーカーなどで組織されたIPTVフォーラムより公開され、今後、この新方式の受信機が普及していくことが予想される。これにより、放送で送信される番組関連情報を用いて、様々なサービス事業者が作成したアプリケーションをテレビ受信機上で利用できるようになる。

放送と通信の連携が進むなか、放送メディアを取り巻く環境の大きな変化の1つに、番組アーカイブと呼ばれる放送済みの番組を蓄積した巨大なデータベースの登場が挙げられる。ユーザは番組アーカイブにアクセスすることで、大量の過去の番組をいつでも視聴できるようになった。もう1つの大きな変化は、Twitterなどのマイクロブログの登場が放送メディアの役割を果たし始めたことである。これまで、放送が伝えてきた情報、例えば、イベント、災害、事故等のより詳細な情報が、マイクロブログを通してリアルタイムに直接多くの人々に届くようになった。

本研究の目的は、視聴中の番組を基点に、番組アーカイブやマイクロブログから通信により容易に関連するコンテンツを取得するアクセス技術を確立することである。放送と通信の連携を進展していく上で、重要な役割を持つ番組アーカイブやマイクロブログを活用するには、大量のデータから必要な情報を効果的に引き出すことができるアクセス手段が不可欠である。情報を使いこなせる人とそうでない人との間に起こるデジタルデバイドが問題となるなかで、放送と通信の連携時代の放送メディアへのアクセスにおいても、誰でも容易に利用できる環境の構築を支援することは急務である。本論文では放送と通信の連携が進む新時代の放送メディアを取り巻く環境において現状の課題を確認し、自然言語処理技術を用いて、ユーザに負担をかけずに意図する情報を効果的に取得するメディアアクセス技術の確立を目指す。

第1章では本研究の背景として、これまでの放送メディアの進展について整理し、近年出現した番組アーカイブやマイクロブログなどの通信を用いた新しいメディアの現状についてまとめる。これらのサービスに効率的にアクセスできる手段を提供することの社会的なニーズについて述べる。さらに、番組アーカイブとマイクロブログへのアクセスに関する現状の技術についてまとめ、取り組むべき課題を確認にする。

第2章では番組アーカイブ内に格納されている大量の番組からユーザが見たい番組を取得するメディアアクセス手段を検討する。家庭用のレコーダなどで利用されている番組検索では、番組に付与されているジャンルや出演者などのメタデータを利用している。しかし、番組アーカイブのように検索対象の番組数が大規模となると、既存のメタデータを利用した番組検索では、検索結果を適量に絞り込めず、ユーザが意図する番組を取得することは難しい。本章では、まず番組検索に関する先行研究について紹介し、本研究の対象を

(別紙様式 2)  
(Separate Form 2)

明確にする。次に、メタデータに含まれる番組内容を示した概要（番組概要）の類似度に基づく関連番組検索手法を提案し、その性能を評価する。さらに提案手法の拡張として、Wikipedia の更新履歴など外部の情報を利用した注目度による番組検索結果のリランキングについて、その有効性について考察する。最後に、番組オンデマンドサービスへの適用事例や、閲覧中のウェブページなどの外部コンテンツから直接関連性のある番組を提示する応用例について述べ、今後の拡張性について考察する。

第 3 章では、内容をより反映した番組検索を実現するため、番組に登場するエンティティの間の関係性をシステムに把握させる手法について検討する。本研究では、対象とするエンティティを人に限定し、番組の内容をコンパクトに記述している番組概要の文章から登場人物間の関係を取得する。まず人物間の関係取得に関する先行研究について紹介し、本研究の対象を明確にする。次に、番組概要から人物表現間の関係を取得し、複数の関係を接続した関係グラフを生成する手法を提案する。続いて、関係グラフ生成で用いられている人物表現抽出、照応関係抽出、関係抽出の性能を評価するとともに、関係グラフを用いることで、構文解析結果から直接取得できない人物間の関係を獲得できることを検証し、取得した関係を利用した番組検索への応用について考察する。また、この関係グラフを用いて、登場人物相関図を生成する番組内容の可視化の応用を検討する。

第 4 章では、新しいメディアとしての存在感を増しているマイクロブログから、テレビなどの既存メディアが伝えきれない情報を取得するための質問応答技術に基づくアクセス手段について検討する。マイクロブログの投稿記事一つ一つの情報量は少ないため、その対象が少なければ容易に情報を取得することができるが、様々な人から時々刻々とマイクロブログに投稿される膨大な情報を網羅的に取得し、その全体像を把握することは難しい。本章では、まず、質問応答技術についての先行研究を紹介するとともに、マイクロブログからの情報取得に適用する際の課題を明確にする。次に、マイクロブログの情報を網羅的に把握するための質問応答に基づく情報取得システムを提案する。続いて、東日本大震災の際に発信された Twitter の大規模データを用いてシステムの性能を評価し、課題と今後の展望を述べる。

第 5 章では、本論文の成果をまとめる。

本論文の成果は、放送と通信の連携に伴い、登場した新しいテレビ視聴環境における課題を確認し、番組アーカイブやマイクロブログから必要な番組や情報を取得するメディアアクセス技術を提案するとともに、その有効性を検証し実用化への道筋を示したことにあつた。本研究により、放送と通信の連携・融合時代の放送メディアへのアクセス方法の確立に向けて大きな貢献を行えたと考えられる。

博士論文の審査結果の要旨  
Summary of the results of the doctoral thesis screening

自然言語処理を利用した放送メディアアクセス技術に関する研究

(論文審査結果) [2014年1月24日実施]

5名の審査委員全員出席の下、博士請求論文(以下、「論文」)の内容についてスライドを用いて約45分間の口頭発表を行った後、質疑応答を30分間行った。

論文では、放送と通信の連携に向けた近年の急速な動きを背景に、番組アーカイブやマイクロブログから視聴中の番組に関連するコンテンツを効果的に取得するためのメディアアクセス技術について論じている。そして、自然言語処理技術を用いて、大量のデータからユーザの興味にあった情報を的確に取得する情報検索・獲得手法を提案するとともに、実験により有効性を示して課題と今後の展望を述べている。

論文は日本語で記述されており、以下の通り全5章から構成されている。

第1章はイントロダクションである。まず研究の背景である放送と通信の連携の動きと現状について述べ、次に番組アーカイブとマイクロブログへのアクセスに関する現状の技術をまとめて、取り組むべき課題を明確にしている。

第2章では、番組アーカイブ内に格納されている大量の番組からユーザが見たい番組を取得するための情報推薦手法について論じている。番組タイトルや出演者に基づく通常の検索では候補番組の適切な絞り込みが困難であることから、番組の内容を簡潔にまとめた「番組概要」の利用を検討し、その特性を踏まえた関連番組検索手法を提案して、実験によってその有効性を示している。

第3章では、内容をより反映した番組検索を実現するため、番組に登場するエンティティ間の関係性をシステムに把握させる手法について述べている。対象とするエンティティとして人物に焦点をあて、番組概要の文章から登場人物間の関係を自動的に獲得する手法を提案している。表現抽出、照応関係抽出、関係グラフに基づく関係抽出の手法を工夫することで既存の自然言語処理ツールよりも大幅に性能を改善し、構文解析結果からは直接取得できない人物間の関係が獲得できることを示している。

第4章では、マイクロブログからテレビなどの既存メディアが伝えきれない情報を取得するための情報提示手法について述べている。マイクロブログに投稿される膨大な情報を網羅的に把握するため、質問応答に基づく情報取得システムを新たに提案し、大規模な言語資源を利用した質問文の意味的に等価な言い換えによって網羅的な検索を可能にしている。また、東日本大震災の際に発信されたTwitterの大規模データを用いて提案手法の有効性を示している。

第5章は結論である。放送と通信の連携に伴い登場した新しいテレビ視聴環境において、提案するメディアアクセス技術が必要な番組や情報の取得に有効であると結論づけるとともに、将来の課題を論じている。

口頭発表では論文の内容に沿って、まず研究の背景および位置づけを述べた後に、言語情報を利用した関連番組検索、番組内容把握のための人物表現間の関係抽出とその応用、質問応答に基づくマイクロブログからの情報獲得について説明があった。これら3つの各テ

(Separate Form 3)

ーマについて、具体例による紹介も含めながら、問題設定、提案手法、実験による評価結果を逐次説明した。最後に研究全体のまとめを行った。

質疑応答では、論文及び口頭発表の内容に関して、理論解析に関する定義や証明を中心に質問があり、的確な回答がなされた。

論文の内容は言語処理学会論文誌「自然言語処理」に原著論文として採録されており、また一部の内容は Intelligent User Interface (IUI)などの査読付き国際会議において発表済である。

以上に基づき審査した結果、5名の審査委員全員一致で、本学位請求論文は学位を授与するのに十分なレベルであるものと判定された。

(試験結果) [2014年1月24日実施]

口述による試験を実施した結果、出願者はその博士論文を中心としてそれに関連がある専門分野及びその基礎となる分野について博士(情報学)の学位の授与に十分な学識を有するものと判断し、合格と判定した。