

音声ドキュメントのセマンティックコンテンツ化と音声対話による高度利用化の研究 (051306006)

Study on Semantic Content of Spoken Document and Utilization through Spoken Dialog

研究代表者

中川聖一 豊橋技術科学大学 工学部情報工学系
Seiichi Nakagawa Faculty of Engineering Toyohashi University of Technology

研究分担者

新田恒雄† 増山繁†† 秋葉友良††† 土屋雅稔†††† 山本和英†††††
小暮悟†††††† 西崎博光††††††††
Tsuneo Nitta† Shigeru Masuyama†† Tomoyoshi Akiba††† Masatoshi Tsuchiya††††
Hidekazu Yamamoto††††† Satoru Kogure†††††††† Hiromitsu Nishizaki††††††††
† 豊橋技術科学大学工学研究科知識情報工学専攻 †† 豊橋技術科学大学工学部知識情報工学系
††† 豊橋技術科学大学工学部情報工学系 †††† 豊橋技術科学大学工学部情報工学系
††††† 長岡技術科学大学電気系 †††††† 静岡大学 情報学部情報科学科
†††††††† 山梨大学大学院医学工学総合研究部工学部コンピュータメディア工学科
†,††,†††,††††† Faculty of Engineering, Toyohashi University of Technology
†††††††† Department of Electrical Engineering, Nagaoka University of Technology
†††††††† Faculty of Engineering, Shizuoka University
†††††††††† Interdisciplinary Graduate School of Medical and Engineering, University of Yamanashi

研究期間 平成 17 年度～平成 19 年度

本研究開発の概要

講義や講演、ニュース音声などの大量のマルチメディアデータを、音声言語処理技術を用いて（半）自動的にインデキシングし、音声対話によるインタラクションによりユーザの所望の内容を表す部分を精度よく検索（情報検索、質問応答）する技術を開発する。これらの研究開発に必要な音声データとして、講義音声を取り上げ、講義音声を収録し、その音声を書き起こし、フィラー・言い直しなどのラベルを付与し、データベースを整備する。話し言葉音声の音声認識、音声対話、音声検索、質問応答の要素研究を行ない、それらを統合したコンテンツ処理システムを開発する。またデータを効率よく圧縮（要約の自動生成）し、これらの講義・講演資料（パワーポイント等）から効率よく学習するためのコンテンツを生成するシステムも開発し、評価実験により、システムの有効性を検討する。

Abstract

Recently, e-Learning systems for the purpose of browsing when preparing and reviewing lectures have been developed, most of which are now available for education purposes. However, with many of these systems, users cannot directly access the specific points that they need from video and slides. The most important and most basic technology for doing this is robust speech recognition of spontaneous speech. In addition, techniques such as spoken retrieval and speech summarization may improve the usability of such systems. In this project, we developed such techniques. Lastly we constructed and evaluated the “Classroom Lecture Browsing System”. From questionnaire results, our system was considered valuable by the majority of our subjects.

1. まえがき

インターネット上には、映像ビデオや音声情報が大量に発信されている。これらを有効に利活用するためには音声言語処理技術を用いて、内容を要約し、インデキシング化により検索できるようにする必要がある。本研究では音声ドキュメントの例として講義音声を例に取り、研究を行なった。

2. 研究内容及び成果

2.1 音声ドキュメント処理

豊橋技術科学大学、静岡大学、山梨大学の 3 大学で、実際に開講された講義音声（16 話者、22 科目、114 講義、3860 分）を収録した。なお、収録マイク自体の影響等を調べるために、ピンマイクと同時にハンドマイクとヘッドセットマイクでも収録した。これらのデータに対し、手作

業による書き起こし、重要文の抽出等、評価実験用のデータベースを構築し、公開した。これらのデータに対し、話し言葉音声認識の高精度化の研究を行った。それらは、コンテキスト依存音節単位の音響モデルや話し言葉に頻繁に出現するフィラー（間投詞）のモデル化、講義対象に適応した言語モデルの構築等である。コンテキスト依存音節単位モデルは、世界標準となっているトライホン単位モデルと同等の性能が得られることがわかった。

講義・講演音声に含まれる音響的・言語的な情報を分析し、これらの特徴がどのように聞き手の受ける印象や理解度などに影響を与えているかを調査した。その結果を元にして、発話速度・声量・ポーズ・明瞭性・抑揚などを中心に音声の分析を行い、音声中の物理特徴量と聴講者の印象との相関を調査した。この結果から、話し方の違いにおける聴講者の印象は異なり、声に含まれる音声の物理特徴量

と比較的高い相関があることが分かった。これを基にして、講師音声の特徴を一目で分かるように図示する方法を開発した。

2.2 コンテンツの高度化

講義音声コンテンツの高度化として、要約に関しては、重要文抽出に基づく方法に対して、重複する内容を含む文の選択を排除する方法（冗長な文の検出）、1文ごとの抽出ではなく複数文の纏まりを抽出する方法を考案し、我々の従来の要約方法の精度を向上でき、人間の要約と遜色ない精度を得た。

講義音声の要約や講義資料と講義音声認識結果の解析によるキーワードを抽出してインデックス化し、音声と対応付けを自動的に行う等の機能を備えたコンテンツ視聴システムのプロトタイプに対して、複合語からなるキーワードの抽出により精度を向上させ、評価実験によりこれらの機能の有効性を実証した。

下記に構築した講義音声のコンテンツとブラウジングシステムを示す。機能として、要約率の選択機能、音声の再生速度の選択機能、スライドと書き起こし文からのキーワード抽出機能、キーワードからのビデオの頭出し機能、再生音声と同期した自動書き起こし文のハイライト表示機能、などがある。



図 講義音声のコンテンツ化とブラウジングシステム

2.3 音声検索と質疑応答システム

音声ドキュメント検索では、大語彙連続音声認識の認識結果と正解書き起こしテキストの間の差異を、統計的翻訳モデルでモデル化し、認識結果から正解テキストを統計的に予測して検索を行う、認識誤りに頑健な音声ドキュメント検索手法を開発した。認識結果をそのまま利用した従来の検索手法よりも、特に文書長の短いパッセージを検索するタスクにおいて、高い性能を示すことが分かった。

音声ドキュメントを対象としたオープンドメイン質問応答の性能改善手法を検討した。先の音声ドキュメント検索によって得られた検索パッセージ候補に対して回答抽出を適用すれば、音声ドキュメントを対象としたオープンドメイン質問応答が実現できる。認識誤りを含む音声ドキュメントの自動書き起こし文書から回答タイプを推定する手法について、推定するタイプごとに有効な特徴量と判定の閾値を自動選択する手法を適用し、質問応答性能を改善した。

2.4 音声対話インタフェース

本プロジェクトで開発を進めてきたマルチモーダル対話システムを、2.3節で述べた音声検索／質問応答システムと組み合わせ、質問応答対話システムとして完成させた。マルチモーダル対話システムの音声対話の部分を質問応答システムと組み合わせることにより、テキスト／音声ドキュメントに対する質問応答システムの利便性と正答率の向上を実現した。新聞記事（毎日新聞 1991 年～2001 年、読売新聞 2000 年～2001 年）、およびニュース音声（NHK ニュース 1996 年 6 月 1 日～7 月 14 日）を対象に実験を行ったところ、新聞記事では正答率が 44%から 68%と大幅に上昇し、ニュースを対象にした実験でも正答率が 17.3%から 19.2%に小幅ながら上昇することが確認でき、提案システムの有効性が確認できた。

3. むすび

音声ドキュメントの代表例として、大学の講義音声を取り上げ、パワーポイント資料と同期した映像・音声を収録し、公開した。音声認識、音声要約、インデキシング化、音声検索、音声質問応答の研究を行い、これらの技術を統合したブラウジングシステムを開発し、評価実験により有効性を示した。

なお、本研究に関連する「音声ドキュメント処理ワークショップ」を平成 19 年 2 月と平成 20 年 2 月に豊橋技術科学大学で開催した。

【誌上発表リスト】

- [1] L. Wang, N. Kitaoka, S. Nakagawa, “Robust distant speaker recognition based on position-dependent CMN by combining speaker-specific GMM with speaker-adapted HMM”, Speech Communication Vol.49, No.6, pp.501-513 (2007.6)
- [2] 中川聖一、富樫慎吾、山口優、藤井康寿、北岡教英、“講義音声ドキュメントのコンテンツ化と視聴システム”、電子情報通信学会論文誌 Vol. 91-D, No.2, pp.238-249 (2008.2)
- [3] L. Wang, S. Nakagawa, N. Kitaoka, “Robust speech recognition by combining short-term spectrum and long-term spectrum based position-dependent CMN with conventional CMN”, IEICE Trans. Vol.E91-D, No.3, pp.457-466 (2008.3)

【申請特許リスト】

- [1] 中川聖一、北岡教英、富樫慎吾、山口優、プレゼンテーション解析装置およびプレゼンテーション視聴システム、日本、出願番号；特願 2006-340825
- [2] 中川聖一、話者認識システム、日本、出願番号；特願 2007-61123
- [3] 入部百合絵、夏目大伍、桂田 浩一、新田恒雄、視線の軌跡情報に基づいた動画像推薦システム、日本、出願番号；特願 2008-28860

【報道発表リスト】

- [1] 新聞報道：日経産業新聞、H19.5.31、講義録画し文書要約、豊橋技科大、自動システム、高度な内容にも対応

【本研究開発課題を掲載したホームページ】

<http://www.slp.ics.tut.ac.jp/CJLC/>