

令和元年6月21日現在

機関番号：13904

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00153

研究課題名(和文) 音声言語と文字言語の統一のおよび相補的利用による音声言語情報アクセスの研究

研究課題名(英文) A study of Information Access by utilizing spoken and written language impartially and complementaly

研究代表者

秋葉 友良 (Akiba, Tomoyosi)

豊橋技術科学大学・工学(系)研究科(研究院)・准教授

研究者番号：00356346

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：近年の情報通信技術の発展により、音声言語と文字言語が相互交換的に利用され記録されるようになってきた。このような状況のもとで、本研究は情報アクセスにおいてテキストと音声に分け隔てなく扱い、かつ相補的に活用する技術として、自由発話音声で表現した情報要求を利用する技術、自由発話音声とキーワードを併用する技術、音声とテキストが相互参照的に混在する対象を検索する技術、不均質なテキストが混在するデータを分類器学習に利用する技術、を開発した。

研究成果の学術的意義や社会的意義

言葉は、それを伝える媒体の違いから、話し言葉(音声言語)と書き言葉(文字言語)に分類され、それぞれ別の機能を担ってきた。しかし、情報の電子的な記録の促進や情報通信技術の発展により、音声言語と文字言語の機能分担が不明瞭になり相互交換的に利用され記録されるようになった。本研究の成果により、このような不均質な情報の利用が促進されるとともに、音声言語と文字言語のそれぞれの利点を活かす技術の開発が進むと考えられる。

研究成果の概要(英文)：Attribute to the recent development on the information processing and communication technology, people started to offer information in spoken language and in written language exchangeably and interactively, which then have been recorded electronically. This study aims to develop a technology to handle both speech and text without distinction and to make use of each advantage complementarily in information access. We developed the method of making use of spontaneously spoken query to express one's information need, the method of using spontaneously spoken query and text keyword query at the same time, the method of document retrieval targeting heterogeneous document collection of speech and text related mutually, and the method of dealing with heterogeneous training data to train a classifier targeting microblogs.

研究分野：音声言語情報処理

キーワード：情報検索 音声ドキュメント検索 文字言語 音声言語 自由発話音声 機械学習

1. 研究開始当初の背景

言葉は、それを伝える媒体の違いから、話し言葉(音声言語)と書き言葉(文字言語)に分類される。一方、言葉が持つ2つの重要な機能として、相互作用の(interactional)なもの、業務的(transactional)なものがある。前者は、人間が言葉を用いて他人に気持ちを伝えたり、社会的関係を構築・維持したりするための機能である。後者は、言葉を用いて知識や技術・情報を伝達する機能である。従来の人間の言語使用では、相互作用の機能は音声言語によって、業務的機能は文字言語によって、主に分担されてきた。つまり、口頭の会話によって社会関係が構築され、文字化された文書によって情報が伝達・記録されてきた。

しかし、計算機が日常的に利用されている現代社会においては、情報の電子的な記録や、情報通信技術の発展により、音声言語と文字言語の機能分担が不明瞭になりつつある。例えば、音声の電子的な記録と、音声の文字化、すなわち音声認識、が可能になったことにより、音声言語を業務的に利用することが可能になりつつある。これを実現する技術を「音声ドキュメント処理」と呼ぶ。申請者は、音声ドキュメント処理の研究を通して、将来的には音声ドキュメントも従来のテキストドキュメントも、シームレスに蓄積され業務的に利用されることになるであろうと予測した。

また逆に、文字を使った個人またはグループ間で即時かつ相互に通信する手段(電子メール、電子掲示板、ブログ、ツイッター、チャットなど)が確立されたことから、文字言語を相互作用的に利用することが可能になった。申請者は、音声言語の持つ表現力に着目して、自由発話音声による情報要求からの検索システムの開発を進めてきた。一方、文字言語の相互作用的利用の拡大が進めば、情報要求の表現媒体としても、音声とテキストを相互交換的に使ったり、それらを併用して豊富な要求を表現する利用形態が進むであろうと予想した。

2. 研究の目的

以上の背景の元、話し言葉音声と書き言葉テキストは、将来的には両者を分け隔てることなく、相互交換的に、かつ相補的に、相互作用のおよび業務的な機能の両方に利用されていくものと考えた。このような展望を見据えると、音声とテキストの両モダリティを統一的に、かつ各モダリティの特長を活かしつつ、互いの利点を併用する音声言語情報アクセス技術の開発が急務であると考えた。

例えば、検索対象文書について、テキスト文書と音声ドキュメントが分け隔てなく蓄積されるようになると、文書の不均質性が問題になると考えられる。音声ドキュメントは一般に長く多様な情報が記録される一方、より冗長になりがちである。テキスト文書は、十分に推敲された密度の濃い情報が記録されるが、例示や言い換えなどの多様な記述に欠ける傾向がある。また、テキスト文書と音声ドキュメントを併用して文書が作成されるようになると、相補的に情報を表現して検索のために参照される仕組みが必要になるであろう。情報量の観点から正規化して両者を分け隔てなく検索するとともに、相互に関連付けられた文書を効果的に検索する技術が必要である。

また、情報要求の表現方法として、音声言語と文字言語とを併用して作成された検索クエリに対する検索技術が必要である。自由発話音声で表現された検索クエリは、豊富な手がかりを含む一方、冗長かつ曖昧性が高い。一方、テキストで表現された検索クエリは、確実な手がかりを含むが多様な手がかりに欠ける。両者の利点を相補的に利用して、検索性能を改善する技術が必要である。

3. 研究の方法

情報アクセスには、(A)人の情報要求を正確かつ漏れなく汲み上げる仕組み、(B)情報要求に対して適合情報を高精度に漏れなく見つける技術、(C)検索者に適合情報を適切に提示する方法の3項の組み合わせが必要であり、それぞれについて研究課題が存在する。本研究では項目(A)(B)に注力して、以下の課題を設定し研究を推進した。

- (A) 自由発話音声とテキストで表現された情報要求を活用した音声言語情報アクセスシステムの開発
- (B) 音声ドキュメントとテキストドキュメントが混在した情報レポジトリに対する検索手法の開発

4. 研究成果

4.1 音声クエリの特徴を利用した検索手法の開発

音声クエリには、検索要求の内容を表す言語情報に加えて、発話内容を修飾するパラ言語情報・非言語情報が含まれる。研究代表者は、テキストクエリには現れない情報を活用した手法として、自由発話音声に含まれるピッチ、パワー、話速などの音響情報を抽出し、発話セグメントの検索貢献度を予測し、検索に利用する手法の開発を行ってきた。この手法について、発

話の時系列関係を考慮に入れるように拡張し、再帰型ニューラルネットワーク(LSTM)を利用した検索貢献度予測手法を開発した。時系列情報を考慮せずに各単語の貢献度予測を行う手法(SVR)との比較実験の結果を以下の表に示す。

表：単語貢献度予測による検索性能評価(MAP)

Query	Document	ベースライン	SVR			LSTM			オラクル
			音響	言語	両方	音響	言語	両方	
manual	manual	0.232	0.231	0.239	0.238	0.238	0.236	0.246	0.307
manual	ASR	0.203	0.204	0.209	0.212	0.208	0.210	0.213	0.240
ASR	manual	0.192	0.190	0.201	0.194	0.191	0.192	0.196	0.246
ASR	ASR	0.182	0.185	0.202	0.198	0.187	0.191	0.197	0.217

第1,第2列の manual および ASR の別は、検索クエリ(Query)と検索対象文書(Document)の表現として、正解書き起こしテキスト(manual, 音声認識 100%に相当)、あるいは音声認識結果(ASR, 認識誤りを含む)を使ったかの違いである。各列に、それぞれの条件での、各手法の検索性能(MAP 値。大きいほど良い)が示されている。手法の「オラクル」は、検索貢献度の予測が正しく行われた場合の検索性能で、提案手法の上限値を示している。ベースラインに比べて、検索貢献度予測を行う提案手法(SVR, LSTM)は、検索性能を改善していることがわかる。また、Query に正しい認識結果を使った場合(manual)は、時系列情報を予測に用いる手法(LSTM)の性能が高い。一方、音声認識結果を使った場合は、時系列情報を用いない手法(SVR)の方が良好な結果を示しており、提案手法は認識誤りの扱いに課題が残されていることがわかった。

4.2 音声クエリとテキストクエリを併用した検索手法の開発

表現手段としての音声の特長は、思考を即時に表現可能であること、思考の中断にならないこと、誰でも容易に労力なく短時間で、専用の入力機器を必要とせずどこでも利用できること、にあると考えられる。この特長を活用し、検索者には情報要求を自由に発想することを許し、それを即時に音声発話して表現することを可能にするような、自由発話音声を入力とする情報アクセス手法を検討してきた。一方、テキストクエリは、認識誤りのない確実な手がかりを与えるという利点がある。そこで、自由発話音声クエリと、それを推敲して整った表現へと編集したテキストクエリのペアに対して、検索性能の違いを分析した。各クエリを構成要素(語)に分解して、それぞれの検索結果に対する貢献度を調べ、各クエリの有する利点と欠点を明らかにした。その結果、平均的にはテキストクエリの方が検索性能は高いが、個別に比較すると自由発話クエリがテキストクエリを大きく改善している場合も多く見られ、音声クエリの特徴である冗長性や多様性が効果的に機能する場合があるという知見を得た。

塾考の上作成された長いテキストクエリの効果は確認できたものの、ユーザが長いクエリを作成する負担は大きく、利便性の点で現実的ではない。そこで、長く多くの手がかりを含むが確実ではない自由発話音声クエリと、短いが確実な手がかりとなるテキストキーワードクエリを、併用する情報要求の表現方法を検討した。上記クエリセット中のテキストクエリを被験者に提示し、情報要求を表す短いキーワードクエリを新たに作成してもらうことで、自由発話音声クエリとキーワードクエリから成る新たなクエリセットを構築した。

このクエリセットを用いて、自由発話音声クエリとテキストキーワードクエリを併用する情報検索手法の開発を行った。以下の表に、音声クエリ、キーワードクエリ、および両方を併用した場合の検索性能を示す。

表：音声クエリとキーワードクエリ、およびその併用による検索性能評価(MAP)

query	document	音声クエリ	キーワード	併用	併用(自動)
manual	manual	0.232	0.248	0.283	0.282
manual	ASR	0.203	0.188	0.218	0.221
ASR	manual	0.192	0.248	0.264	0.262
ASR	ASR	0.182	0.188	0.219	0.222

第1,第2列の manual および ASR の別は、検索クエリ(Query)と検索対象文書(Document)の表現として、正解書き起こしテキスト(manual, 音声認識 100%に相当)、あるいは音声認識結果(ASR, 認識誤りを含む)を使ったかの違いである。各列に、それぞれの条件での、各手法の検索性能(MAP 値。大きいほど良い)が示されている。「音声クエリ」「キーワード」の列は、音声クエリ、キーワードクエリをそれぞれ単独で使った場合の検索性能を、「併用」の列は、両者を併用した時の検索性能が示されている。キーワードクエリは、特にクエリを音声認識した場合と比べて良好な結果を示している。しかし、豊富な手がかり表現を含む音声クエリを併用した場合、すべての場合で検索性能が改善しており、両者を相補的に利用する効果が示されている。

上記の併用手法では、音声クエリ用の検索モデルとキーワードクエリ用の検索モデルを線形補間することで検索結果を得た。どちらのクエリをどれだけ重視するかを決める重み係数は、

上記の実験では定数とした。しかし、具体的な音声クエリおよびキーワードクエリの内容、および両者の関係でどちらに重みを置くかは変化するはずである。そこで、音声クエリおよびキーワードクエリを分析した上で、重み係数を動的に推定する手法の開発を行った。上記の「併用(自動)」の列に、重みの動的推定を行った場合の検索性能を示す。自動推定を行った場合でも、評価セットの最適固定値を用いた場合(「併用」と同等の性能が達成でき、特に検索対象が音声認識結果であるときに自動推定の効果が大きいことがわかった。

4.3 音声とテキストが混在するリポジトリに対する検索手法の開発

学会講演を対象とした情報検索評価用テストコレクションを整備し、検索対象として講演音声と発表用プレゼンテーションスライドに含まれるテキスト情報が混在するテストコレクションを構築した。この不均質なテストコレクションに対する文書検索手法を検討した。ベースラインとなる手法として、以前に開発した講演音声とプレゼンテーションスライドに含まれるテキスト情報を併用する手法を実装した。プレゼンテーションスライドをより良く活用する手法として、音声とテキストを相補的に併用し、かつプレゼンテーションスライドの構造を利用する検索手法を開発した。プレゼンテーションスライドの構造を利用する手法としては、スライドの各構成要素の役割を考慮し情報の統合を行う手法と、スライドの種類(例えば、箇条書きテキストで言語情報で情報を伝えるスライド、図・表を中心に視覚的な情報を提供して音声で説明をするスライド、など)を考慮して音声との関係を利用する手法、を検討した。評価実験により、各手法単独で若干の検索性能が改善し、両手法を併用することでより大きな効果が得られることが分かった。

4.4 マイクロブログからの罹患判定手法の開発

音声ドキュメントとテキストの中間的な文書表現として、Twitterなどに代表されるマイクロブログに着目した。マイクロブログは、テキストとして発信される一方、話し言葉的な言語表現が利用され、またパラ言語情報に相当する表現として顔文字などの非言語的な文字が現れる。そこで、マイクロブログを対象とした検索の問題として、ツイッター投稿をもとに投稿者の罹患判定に取り組んだ。

罹患判定を機械学習による分類問題として定式化する場合、投稿に対して罹患・非罹患をラベル付けした学習データが必要となる。しかし、投稿のプライバシー保護の観点から、実際の投稿にラベル付けした学習データの入手は難しく、現状では擬似的に作成したマイクロブログに対するラベル付き学習データのみが利用可能である。擬似的なマイクロブログは、実際のツイートと異なり、言語表現は書き言葉に近く、非言語的言語の使用も行われていない。そこで、マイクロブログにおける投稿へのリプライの構造と、罹患患者への定型的応答「お大事に」を利用して、自動的に罹患患者の投稿を収集する手法を開発した。

さらに、擬似的に作成したラベル付きマイクロブログと自動的に獲得したラベル無しマイクロブログから成る不均一な学習テキストを用いて、クラス分類を行う手法を開発した。実際のツイッター投稿を対象に行った実験結果を以下の表に示す。

表：学習データの違いによる罹患判定性能

	ベースライン	人手作成	自動獲得データ	併用
分類精度	0.559	0.640	0.660	0.714

ベースラインは、全てのテストデータを罹患と判定した場合の精度である。人手作成は、人手で擬似的に作成したマイクロブログ約2,000件から構成されるラベル付きデータで学習した分類器である。自動獲得データは、提案手法により自動収集した約200,000件の(不確実な)ラベル付きデータで学習した分類器である。提案手法は、品質は低いものの自動的にラベル付きデータを収集できるので、大量の学習データを低コストで収集することができる。大量のデータを利用することで、人手作成と同程度の分類性能を達成できた。また、ノイズを多く含むが構築コストが低い自動獲得データを、整った人手構築データと併用することで、分類性能が改善できることを確認した。

5. 主な発表論文等

〔雑誌論文〕(計 1 件)

田崎広人, 秋葉友良,
自然発話音声クエリの音響・言語特徴を利用した音声内容検索
電子情報通信学会論文誌, Vol. J101-D, No.3, 2018/3. (査読あり)

〔学会発表〕(計 27 件)

- (1) 白土大樹, 秋葉友良, 増山繁,
注釈者のクラスタリングに基づく議会発言の事実検証可能性の推定
言語処理学会第 25 回年次大会, 2019/3
- (2) 浅川玲音, 秋葉友良,
罹患患者への定型的応答を利用した罹患ツイートの自動獲得と RNN 罹患判定器学習への適用
言語処理学会第 25 回年次大会, 2019/3
- (3) 森田知熙, 秋葉友良, 塚田元,
双方向ニューラル機械翻訳の反復的な教師なし適応の検討
言語処理学会第 25 回年次大会, 2019/3
- (4) 佐橋広也, 秋葉友良, 中川聖一,
科学技術論文抄録と講義音声の英日機械翻訳のリスクアリングの検討
言語処理学会第 25 回年次大会, 2019/3
- (5) 高田凌平, 秋葉友良, 塚田元,
ニューラル機械翻訳におけるトピック情報の利用
言語処理学会第 25 回年次大会, 2019/3
- (6) 浅川玲音, 秋葉友良,
罹患患者への定型的応答を利用したツイート罹患判定のデータ拡張,
第 5 回自然言語処理シンポジウム, 2018/12.
- (7) 森田知熙, 秋葉友良, 塚田元,
双方向の逆翻訳を利用したニューラル機械翻訳の教師なし適応の検討,
第 5 回自然言語処理シンポジウム, 2018/12.
- (8) 佐橋広也, 西村友樹, 秋葉友良, 中川聖一
統計的翻訳とニューラル翻訳に基づく翻訳候補文の分散表現と逆翻訳によるリスクアリングの
検討,
音声言語情報処理研究会(SIG-SLP), 2018/5.
- (9) 田崎 広人, 秋葉 友良
自然発話クエリを利用した内容検索におけるテキストクエリ併用の効果
日本音響学会春季研究発表会, pp. 193-196, 2018/3.
- (10) 田崎 広人, 秋葉 友良
自然発話クエリを利用した内容検索における RNN による単語貢献度予測
日本音響学会春季研究発表会, pp. 189-192, 2018/3.
- (11) 浅川 玲音, 秋葉 友良
疾病サーベイランスのための罹患ツイートの自動獲得と事実性判定への利用
言語処理学会第 24 回年次大会, pp.396-399, 2018/3.
- (12) 西村友樹, 秋葉友良, 塚田元
大規模単言語資源を用いた大語彙ニューラル機械翻訳の検討
言語処理学会第 24 回年次大会, pp.21-24, 2018/3.
- (13) 佐橋広也, 西村友樹, 秋葉友良, 中川聖一
統計的翻訳とニューラル翻訳による翻訳候補の文の分散表現に基づくリスクアリングの検討
言語処理学会第 24 回年次大会, pp.260-263, 2018/3.
- (14) Reine Asakawa, Tomoyosi Akiba
AKBL at the NTCIR-13 MedWeb Task
In Proceedings of the 13th NTCIR Conference, pp.52-55, 2017.
- (15) Hiroto Tasaki, Tomoyosi Akiba,
Incorporating Acoustic Features for Spontaneous Speech driven Content Retrieval
In Proceedings of International Conference on Speech Communication and Technology
(Interspeech), pp.2894-2898, 2017.

(16)Tomoyosi Akiba, Bon Sy, Ayman Zeidan,
ICD-10 Code Retrieval based on Distributional Semantics of Diagnosis Descriptions
In Proceedings of International Conference on Advanced Informatics: Concepts, Theory and
Applications (ICAICTA), 2017.

(17)Tomoki Nishimura, Tomoyosi Akiba.
Addressing Unknown Word Problem for Neural Machine Translation using Distributed
Representations of Words as Input Features
In Proceedings of International Conference on Advanced Informatics: Concepts, Theory and
Applications (ICAICTA), 2017.

(18)Koya Sahashi, Norioki Goto, Hiroshi Seki, Kazumasa Yamamoto, Tomoyoshi Akiba,
Seiichi Nakagawa
Robust Lecture Speech Translation for Speech Misrecognition and Its Rescoring Effect
from Multiple Candidates
In Proceedings of International Conference on Advanced Informatics: Concepts, Theory and
Applications (ICAICTA), 2017.

(19)田崎 広人, 秋葉 友良
自然発話クエリの音響・言語特徴を利用した確率的検索モデルによる音声内容検索
日本音響学会春季研究発表会, pp. 197-200, 2017/3.

(20)水上仁志, 秋葉友良
統計的機械翻訳における未知の一般語と固有名への対処
言語処理学会第 22 回年次大会, pp.1042-1045, 2017/3.

(21)Tomoyosi Akiba,
“Taking advantage of spontaneous speech for document retrieval: Lessons from the
organization of evaluation tasks”
5th Joint Meeting of Acoustical Society of America and Acoustical Society of Japan, 2016/11

(22)河谷浩志, 大野哲平, 秋葉友良
距離順音声検索語検出に基づく音声ドキュメントの漸進的内容検索
日本音響学会秋季研究発表会, pp.77-80, 2016/9.

(23)Hitoshi Mizukami, Tomoyosi Akiba
Effects of Class-based Statistical Machine Translation on Unknown Names
In Proceedings of International Conference on Advanced Informatics: Concepts, Theory and
Applications (ICAICTA), 2016.

(24)Kosuke Yamauchi, Tomoyosi Akiba
Repeated Event Discovery from Image Sequences by Using Segmental Dynamic Time
Warping: Experiment at the NTCIR-12 Lifelog Task
In Proceedings of the 12th NTCIR Conference, pp.382-385, 2016.

(25)Sho Kawasaki, Hiroshi Oshima, Tomoyosi Akiba
Graph-based Document Expansion and Robust SCR Models for False Positives:
Experiments at the NTCIR-12 SpokenQuery&Doc-2
In Proceedings of the 12th NTCIR Conference, pp.196-199, 2016.

(26)Kosuke Yamauchi, Tomoyosi Akiba
Repeated Event Discovery from Image Sequences by Using Segmental Dynamic Time
Warping: Experiment at the NTCIR-12 Lifelog Task
In Proceedings of the 12th NTCIR Conference, pp.382-385, 2016.

(27)Tomoyosi Akiba, Hiromitsu Nishizaki, Hiroaki Nanjo, Gareth Jones
Overview of the NTCIR-11 SpokenQuery&Doc-2 Task
In Proceedings of the 12th NTCIR Conference, pp.167-179, 2016.

〔図書〕(計 1 件)

秋葉友良, 「音声ドキュメント処理」, pp.795-796,
人工知能学会編, 人工知能学大辞典, 共立出版, 2017.