

平成 21 年 6 月 4 日現在

研究種目：若手研究(B)
 研究期間：2006～2008
 課題番号：18700188
 研究課題名(和文) 高度な情報検索システムのための認識誤りの質に着目した音声認識戦略
 研究課題名(英文) Automatic Speech Recognition focusing on significance of errors for Sophisticated Information Retrieval Systems
 研究代表者
 南條 浩輝 (NANJO HIROAKI)
 龍谷大学・理工学部・助教
 研究者番号：50388162

研究成果の概要：音声入力型の情報検索および音声を検索対象とした情報検索のための音声認識戦略について研究を行った。具体的には、各単語やフレーズに対して検索に対する影響度をモデル化し、検索への影響が大きいものから優先的に音声認識する方法について研究を行った。音声入力型テキスト検索および講演音声からの重要文抽出において本音声認識戦略が有効であることを確認した。各単語の影響度のモデル化も半自動で行えることがわかった。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	1,400,000	0	1,400,000
2007年度	1,100,000	0	1,100,000
2008年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,400,000	270,000	3,670,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声認識，情報検索，探索アルゴリズム，ベイズリスク

1. 研究開始当初の背景

記憶装置の大容量化と高速ネットワークの普及により、インターネットを通じて大規模なデータに容易にアクセスできるようになった。WEB 上には自然言語で記述された大規模データ(音声データも含む)が大量に蓄積され、必要とする情報を素早く確実に取り出す技術、すなわち、高度な言語理解処理をとともなう情報検索の手法が注目されている。特に音声を含む映像データなどの高度な検索手法は確立されていない。また、音声入力による検索に対する要望も大きい。このような高度な情報検索のために、音声認識は必須の技術であり、その精度向上が強く望まれて

いる。

従来、音声認識は『音声認識誤りの量』に着目しており、全ての単語やフレーズを区別せず一様に扱っていた。この結果、情報検索にとって重要な単語を間違って認識しても、たくさんのあまり重要でない単語(たとえば助詞など)が正しく認識されるほうが、音声認識結果として優先されるという問題があった。すなわち『各音声認識誤りの質』を高めることが望まれており、「情報検索に致命的な影響を与える単語やフレーズの誤りを重要視し、そのようなものの誤りを最小にする」音声認識戦略が必要であった。

2. 研究の目的

(1) 損失を最小化する“期待損失最小化音声認識”手法の確立

音声認識をパターン認識における期待損失最小化の枠組み、すなわち、分類誤り（認識誤り）に対して損失（損失は0か1の二値ではなく、認識誤りの『質』に応じた連続値）を定義し、損失の平均値が最も小さくなるように識別を行うという枠組みでとらえ、質を高める音声認識を実現する。この実現により、情報検索にとって重要度の高い語句からきちんと音声認識されていくことが期待できる。

さらにその際、全く新しい音声認識戦略を設計するのではなく、これまでの音声認識研究を包括する一般的な形式で音声認識戦略を表現することを目指す。

(2) 情報検索における情報損失の観点からの音声認識誤りの質の定義方法の確立

損失基準の定義は重要な問題である。各語句の認識誤りに対する損失は情報検索タスクによって異なる。はじめに情報検索システムに関する知識を用いて損失をモデル化する方法を確立し、次に知識を用いないで損失のモデル化を行う自動化手法を確立する。

3. 研究の方法

(1) 実情報検索タスクの整備

実際に大規模な情報検索タスクを整備し、そのタスクで提案手法の有効性を実証する。具体的には、WEB 検索タスク、講演音声ドキュメント検索タスク、要文抽出タスクを設計し、提案手法の有効性を評価するための基盤システムを構築する。

(2) 損失を最小化する“期待損失最小化音声認識”アルゴリズムの実装

音声認識はパターン認識の一部であり、入力音声に対してそれが属する語クラスを求める問題である。このとき、あるクラスに属するデータを誤ったクラスに分類した時の損失を定義し、その期待損失、すなわちベイズリスクを最小とする決定を行うように音声認識を定式化し、実装する。

具体的には、各語句の重要度を考慮した音声認識の評価尺度（誤り率）を定義し、その評価尺度を最小化するような音声認識を行う。

(3) 情報検索における語句の重要度の自動推定

情報検索タスクによって、各語句の重要度は異なる。人手による設定は対象とする情報検索システムについての深い知識が必要でありコストが大きい。これを自動化する。具体的には、語句の重要度を考慮した音声認識の評価尺度が検索の評価尺度と強い相関をもつように語句の重要度を決定する。すなわち、音声認識において誤り率を最小化することが情報検索の精度低下を最も少なくすることと同一視できるように語句の重要度を決定する。

4. 研究成果

(1) 実情報検索タスクの整備

以下の情報検索タスクを整備し、実際に検索実験を行えるようにした

1. WEB 検索タスク

NTCIR3-WEBの音声検索タスクを導入した。100G サイズの検索タスクであり、ベクトル空間モデルによる検索システムを整備した。

2. 講演音声ドキュメント検索タスク

CSJの2702講演から、検索要求に合致する講演や発話の一部を検索するタスクを整備した。ベクトル空間モデルによる検索システムを整備した。これは近いうちに公開を行う予定である。

3. 重要文抽出（重要文検索）タスク

CSJの講演から重要文を抽出するタスクを整備した。談話標識を手掛かりとした重要文を求めるシステムを整備した。

(2) 期待損失最小化音声認識の実装

以下の3種類の期待損失最小化音声認識を実装し比較を行った。

1. N-best 探索

2. ラティス型単語グラフ探索

3. コンフュージョンネットワーク（CN）型単語グラフ探索

2のラティス型単語グラフ探索は計算量が多いわりに高い精度が得られないこと、CN型単語グラフ探索は精度に優れること、1のN-best探索も比較的高速に動作し、精度についても効果があることがわかった。

オープンソースの音声認識エンジンJuliusに組み込むことを検討中であり、実現できれば多くの研究者の間で利用できるものとなる。

(3) 語句の重要度の自動推定

単語重みは、その単語が音声認識で誤認識

されたときの検索に及ぼす影響に基づいて決定する必要がある。この問題に対し、音声認識の誤り率と情報検索精度の低下率が等しくなるように、最急降下法に基づいて単語の重要度の推定を行う方法（教師あり学習法）を実現した。なお本手法は、以下の3種類のデータを用意するだけで適切な重要度を推定するものである。

1. 当該情報検索の典型的なクエリ集合とその書き起こし
2. 各クエリの音声認識結果
3. 各クエリに対する検索の正解ラベル

WEB 音声検索タスクにおいて、提案音声認識手法により、音声認識精度と情報検索の精度の両方を改善することができた。

また、3 の検索正解ラベルを必要としない学習法も実現し、教師ありと比べると若干劣るものの、効果があることを確認した。

本成果に関しては論文誌にまとめた上でプログラム等の公開を目指している。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計9件)

Tomoyosi Akiba, Kiyooki Aikawa, Yoshiaki Itoh, Tatsuya Kawahara, Hiroaki Nanjo, Hiromitsu Nishizaki, Norihito Yasuda, Yoichi Yamashita, and Katunobu Itou. "Construction of a test collection for spoken document retrieval from lecture audio data." IPSJ-Journal, Vol.50. No.2, pp.501-513, 2009. 査読あり

Tomoyosi Akiba, Kiyooki Aikawa, Yoshiaki Itoh, Tatsuya Kawahara, Hiroaki Nanjo, Hiromitsu Nishizaki, Norihito Yasuda, Yoichi Yamashita and Katunobu Itou. "Test Collections for Spoken Document Retrieval from Lecture Audio Data.", Proceedings of the Sixth International conference on Language Resources and Evaluation (LREC 2008), 2008. 査読あり

南條浩輝,河原達也,七里崇,"音声理解を指向したベイズリスク最小化枠組みに基づく音声認識." 電子情報通信学会論文誌, Vol. J91-D, No.5, pp.1314-1324, 2008. 査読あり

Takashi Shichiri, Hiroaki Nanjo and Takehiko Yoshimi. "Minimum Bayes-Risk Decoding with Presumed Word Significance for Speech Based

Information Retrieval.", In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008), pp.1557-1560, 2008. 査読あり
Takashi Shichiri, Hiroaki Nanjo, and Takehiko Yoshimi. "Automatic Estimation of Word Significance oriented for Speech-based Information Retrieval", Proceedings of The Third International Joint Conference on Natural Language Processing (IJCNLP 2008), pp.204-209, 2008. 査読あり
七里崇, 南條浩輝, 吉見毅彦. "音声理解を指向した音声認識のための単語重要度の自動推定.", 情報処理学会研究報告 SLP-66-14 (NL-179-14), pp.79-84, 2007. 査読なし

秋葉友良, 相川清明, 伊藤慶明, 河原達也, 南條浩輝, 西崎博光, 安田宣仁, 山下洋一, 伊藤克亘. "音声ドキュメント検索テストコレクションの試作と基本検索性能評価.", 第1回音声ドキュメント処理ワークショップ, pp73--80, 2007. 査読なし

伊藤克亘, 相川清明, 秋葉友良, 伊藤慶明, 河原達也, 南條浩輝, 西崎博光, 安田宣仁, 山下洋一. "音声ドキュメント検索評価のためのテストコレクションの試作.", 電子情報通信学会技術研究報告 SP2006-108(NLC2006-52), pp.137-142, 2006. (情報処理学会研究報告 SLP-64-25, 2006.) 査読なし
南條浩輝,河原達也."単語グラフを利用したベイズリスク最小化音声認識とそれに基づく重要文抽出.", 電子情報通信学会技術研究報告 SP2006-106, (NLC2006-50), pp.125-130, 2006. (情報処理学会研究報告 SLP-64-23, 2006.) 査読なし

〔学会発表〕(計13件)

Tomoyosi Akiba, Kiyooki Aikawa, Yoshiaki Itoh, Tatsuya Kawahara, Hiroaki Nanjo, Hiromitsu Nishizaki, Norihito Yasuda, Yoichi Yamashita and Katunobu Itou. "Test Collections for Spoken Document Retrieval from Lecture Audio Data.", the Sixth International conference on Language Resources and Evaluation (LREC 2008), 2008年5月28-30日, Marrakech, Morocco.
Takashi Shichiri, Hiroaki Nanjo and Takehiko Yoshimi. "Minimum Bayes-Risk Decoding with Presumed Word Significance for Speech Based Information Retrieval.", IEEE International Conference on Acoustics,

Speech, and Signal Processing (ICASSP 2008), 2008年3月30日~4月4日, Las Vegas, U.S.A.

南條浩輝, 七里崇, 吉見毅彦."情報検索システムの音声インターフェースのための単語重要度の自動推定とそれに基づく音声認識", 日本音響学会研究発表会, 春季, 2008年3月17日~19日, 千葉工業大学

Takashi Shichiri, Hiroaki Nanjo, and Takehiko Yoshimi. "Automatic Estimation of Word Significance oriented for Speech-based Information Retrieval", The Third International Joint Conference on Natural Language Processing (IJCNLP 2008), 2008年1月7日~12日, Hyderabad, India

南條浩輝."多言語音声メディアコンテンツのアーカイブのための音声認識手法", 日本音響学会研究発表, 秋季, 2007年9月19日~21日, 山梨大学

七里崇, 南條浩輝, 吉見毅彦."情報検索を指向した音声認識のための単語重要度の自動推定", 日本音響学会研究発表会, 秋季, 2007年9月19日~21日, 山梨大学

七里崇, 南條浩輝, 吉見毅彦."音声理解を指向した音声認識のための単語重要度の自動推定.", 情報処理学会音声言語処理研究会, 2007年5月24日~25日, 法政大学

南條浩輝, 河原達也."単語グラフを用いたベイズリスク最小化音声認識の重要文抽出における評価", 日本音響学会研究発表会, 春季, 2007年3月13日~15日, 芝浦工業大学

秋葉友良, 相川清明, 伊藤慶明, 河原達也, 南條浩輝, 西崎博光, 安田宣仁, 山下洋一, 伊藤克亘."音声ドキュメント検索テストコレクションの試作と基本検索性能評価.", 第1回音声ドキュメント処理ワークショップ, 2007年2月26日~27日, 豊橋技術科学大学

伊藤克亘, 相川清明, 秋葉友良, 伊藤慶明, 河原達也, 南條浩輝, 西崎博光, 安田宣仁, 山下洋一."音声ドキュメント検索評価のためのテストコレクションの試作.", 電子情報通信学会音声研究会, 2006年12月21日~22日, 名古屋大学

南條浩輝, 河原達也."単語グラフを利用したベイズリスク最小化音声認識とそれに基づく重要文抽出.", 電子情報通信学会音声研究会, 2006年12月21日~22日, 名古屋大学

Hiroaki Nanjo, Tatsuya Kawahara. "Towards an efficient archive of spontaneous speech: Design of

computer-assisted speech transcription system.", 4th Joint Meeting Acoustical Society of America and Acoustical Society of Japan, 2006年11月28日~12月2日, Hawaii, U.S.A.
南條浩輝, 河原達也."音声認識エンジン Juliusの単語グラフを利用したベイズリスク最小化音声認識", 日本音響学会研究発表会, 秋季, 2006年9月13日~15日, 金沢大学

6. 研究組織

(1) 研究代表者

南條 浩輝 (NANJO HIROAKI)

龍谷大学・理工学部・助教

研究者番号: 50388162