様式 C-19

科学研究費補助金研究成果報告書

平成２２年　５月２８日現在

研究種目：若手研究（B）
研究期間：2007～2009
課題番号：１９７００１８４
研究課題名（和文）　ウェブドキュメントを利用した音声認識結果の信頼度推定と音声認識の高精度化
研究課題名（英文）　Confidence Estimation and Improvement of Speech Recognition Results Using Web Documents

研究代表者
高橋　伸弥（TAKAHASHI SHINYA）
福岡大学・工学部・助教
研究者番号：４０３３０８９９

研究成果の概要（和文）：ニュース音声を音声認識した結果に含まれる単語をキーワードとして検索したウェブ上の文書から音声認識用言語モデルを学習することで音声認識処理を高精度化することを試みた。その際、検索結果の文書内に含まれる単語の出現頻度を用いて計算した文書間の類似度により分類した結果から、元々の認識結果の信頼度を推定する方法を提案した。更に口語文章に含まれる長単位での定型表現をモデル化するためにネットワーク文法を自動構築する方法について検討し、高精度化のための手法を考案した。

研究成果の概要（英文）：This research attempted to achieve the high accuracy speech recognition for news speech using a language model constructed from Web documents collected with keywords in speech recognition results. In addition, the method estimating confidence of speech recognition results from clustering results of collected Web documents was proposed. Moreover, the method constructing network grammar automatically was investigated for multi-word expression in oral sentences.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００７年度	1,200,000	0	1,200,000
２００８年度	900,000	270,000	1,170,000
２００９年度	800,000	240,000	1,040,000
年度			
年度			
総　計	2,900,000	510,000	3,410,000

研究分野：音声認識
科研費の分科・細目：知覚情報処理・知能ロボティクス
キーワード：音声認識、言語モデル適応、スペクトラルクラスタリング、ウェブドキュメント、ニュース音声、ネットワーク言語モデル、話し言葉認識

１．研究開始当初の背景

研究を開始した 2007 年当時は、インターネットの急速な普及とそのブロードバンド化に伴い、ウェブ上でも大容量の映像情報が提供されるようになってきた時期である。ハードディスクを組み込んだ大容量家庭用録画機器が登場したことや、パソコン上で簡単にテレビ番組が録画できるようになったこともあり、膨大な映像データの中から見たい情報を検索する機能の実現は必要不可欠な

ものとして様々な機関で研究が行われていた。

このような背景のもと、音声ドキュメントを対象とした第1回ワークショップが開催されたのも研究開始直前の2007年2月であり、本報告者を含む研究グループも本研究の核となる部分について、このワークショップで発表を行った。

２．研究の目的

音声ドキュメントの中でもニュース映像は、その内容の有用性から、検索用のデータベースとして保存する価値が高いものとして多くの研究が行われている。

高精度なニュース映像検索機能を実現するためには、ニュース映像の内容を表す適切な検索用キーワード（索引語）を予め付与しておく必要があるが、日々大量に生み出されているニュース映像に人手で索引語を付与することは困難である。これに対し、ニュース映像の音声を音声認識し、その認識結果から索引語として適切な語を抽出することが試みられている。このとき誤って認識した語を索引語としたのでは検索性能が著しく低下してしまうため、誤認識をいかに低減するかが重要となる。

そこで本研究では、上記のような問題に対処することを目的として、①認識結果の信頼度を推定する方法および②高精度な言語モデルを構築する方法を検討した。

３．研究の方法

上記のような問題に対する解決策の１つとして、認識対象の音声の内容（対象トピック）に言語モデルを適応させる方法がある。トピックが既知である場合には、関連したテキストを多量に収集し、統計的言語モデルを構築する方法が最もストレートな方法である。また多量のテキストを収集することが困難な場合には、既存の言語モデルと融合する手段がとられている。しかし、ニュースのように、様々な分野にまたがり、予めトピックを特定することはできない場合には、誤認識を含んだ音声認識結果から、その分野を推測する必要がある。

一方、高精度な音声認識を実現するには、ニュース映像に含まれる音情報の中から明瞭な音声区間を切り出しておく必要がある。音情報には音声だけでなくＢＧＭなどの音楽や効果音、雑音などが多く含まれているため、これらをそのまま音声認識に入力したのでは、信頼できる索引語を抽出することができない。さらに音声区間であっても複数の話者が同時に発話していたり、音声以外の背景雑音が重畳されていたりといった不明瞭な音声区間であれば認識率は極端に低下してしまう。音声／非音声区間の識別に関しては、音声波形の零交差数を利用する方法やケプストラム特徴を利用する方法が提案されているが、信頼のできる明瞭な音声区間のみを切り出す方法に関してはあまり検討されていない。

そこで本研究では、ニュース映像中のトピックごとに高精度な索引語を自動的に付与することを目的として、「ウェブドキュメントを利用した音声認識結果の信頼度を推定する方法」を提案した。具体的には、

(1)音声認識結果から抽出した索引語候補によるウェブ上のニュース記事の収集

(2)収集記事のクラスタリングに基づく音声認識結果の信頼度推定

(3)信頼度による明瞭な音声区間の判定方法

(4)信頼度の大きさを考慮した適応言語モデルの更新方法

を検討した。ここでの基本的な考え方は、収集したニュース記事をクラスタリングした結果、互いに類似した多数の記事が同一クラスタとして分類されたならば認識結果の信頼度は高く、クラスタの分散が大きいならば認識結果の信頼度は低いとするものである。この信頼度は、収集したテキスト内における索引語候補単語（群）の共起関係により計算するものと見做すこともできる。これは、ニュース映像として放送されるトピックと同一内容のニュース記事が新聞でも報道されることから、ウェブ上の新聞社サイトにも類似した内容の記事が存在することが予想でき、認識結果から抽出した索引語候補で類似記事を検索・収集した際に、認識精度が高ければ正解文に類似した記事を多数収集することができるが、逆に誤認識を多く含むような場合には、対応する記事が見つからず、類似度の低い記事、正解文と関連の無い記事を収集してしまうだろうという考え方に基づいている。

具体的な実現方法としては、①既存の言語モデルを用いてニュース音声を認識し、②認識結果から索引語候補群を抽出した後、③それらを検索キーとしてウェブ上の類似記事の検索を行う。④さらに収集記事群をクラスタリングして、⑤クラスタの相対的な大きさによって認識結果の信頼度を推定する。ここで信頼度の低い音声区間は不明瞭な音声区間であったとして索引語付けを行わないこととし、信頼度の高い場合には、最も類似するクラスタ内の記事を用いて言語モデルの適応を行い、再び音声認識を行う（図１）。

上記のようなウェブ上のテキストを利用した言語モデル適応は、既に多くの研究例がある。その多くは既知のトピックに対するも

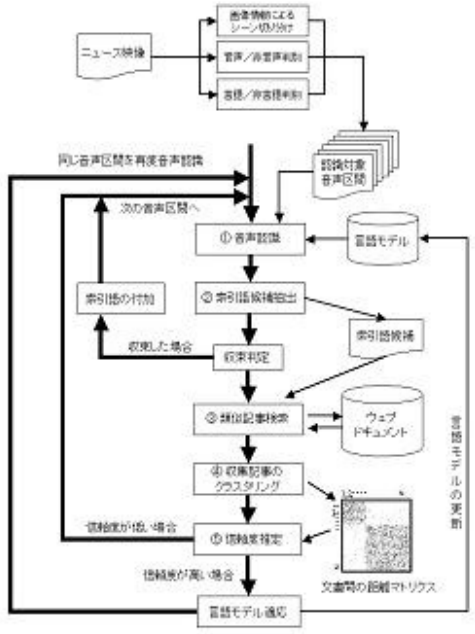

図 1　提案手法の概要

のであり、検索キーワードを人間が指定するものとなっている。本研究と同様、トピックを未知として教師無し適応を行っている研究もあるが、検索結果の大量のテキストを既存のコーパスに追加することで未知語へ対応するというものであり、検索結果の信頼性は検討されていない。

また、本研究で提案する音声認識結果の信頼度推定法は、誤り訂正等で多く用いられている単語信頼度とは異なり、認識結果（索引語候補集合）全体の信頼度となっている。さらに単語の事後確率や音響スコア、N-best 認識結果などといった音声認識器内部の情報をもとに算出されるのではなく、ウェブ上のテキストという外部コーパスを用いる点に特徴がある。

４．研究成果

(1)2007 年度

ウェブドキュメントを収集するための検索キーワードには、ニュース音声に対する音声認識結果中の名詞句を用いることとし、スペクトラルクラスタリングと呼ばれるクラスタリング手法を用いて、誤認識された語句から収集された文書と正認識された語句から収集された文書とを分類することにより、音声認識用言語モデルの精度を向上させることを試みた。小規模なデータに対する実験を行い、本提案手法の基本的なアイデアの有効性を確認し、それらを国際会議他で発表した。

図 2 に、あるニューストピックに対する実験結果を示す。この図は、認識結果から収集

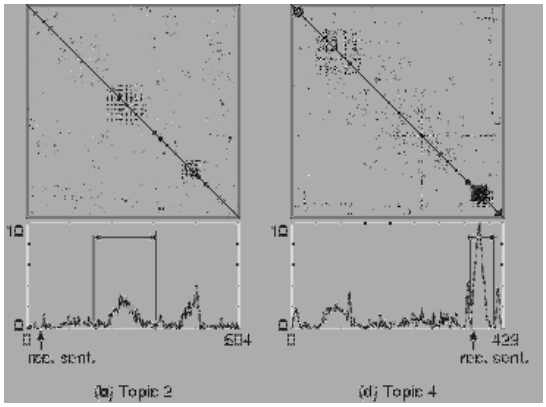

図 2　クラスタリング結果

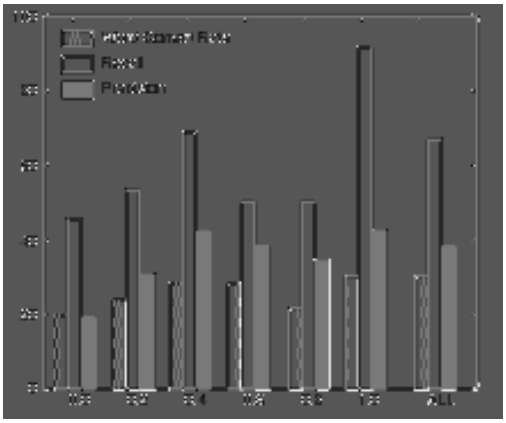

図 3　言語モデル適応結果

されたニュース記事と認識結果との関係を示したものである。左の図は、認識結果に誤りが多く含まれ、その結果、類似したニュース記事が収集できなかったケースを示している。また右の図は、認識結果と類似した記事が収集できたケースである。図 2 の右のケースに対する言語モデル適応の認識結果を図 3 に示す。横軸は既存の言語モデルと類似記事のみから作成した新規言語モデルとを混合する際の重みの値であり、1.0 は新規モデルのみを使用したことを示している。また ALL はクラスタリングせず収集記事全てを使用した際の結果である。実験の結果、類似記事から作成した言語モデルを使用することで再現率（Recall）が大幅に改善されることが示された。

更に、複数のトピックから構成されるニュース音声に対する、同様のアプローチに基づいたトピック切り分け手法を考案し、クラスタリングによりトピック境界を検出することを試みた。具体的には、複数のトピックから得られた誤りを含む検索キーワードを用いてウェブドキュメントを収集し、それらをクラスタリングすることで同一のトピック

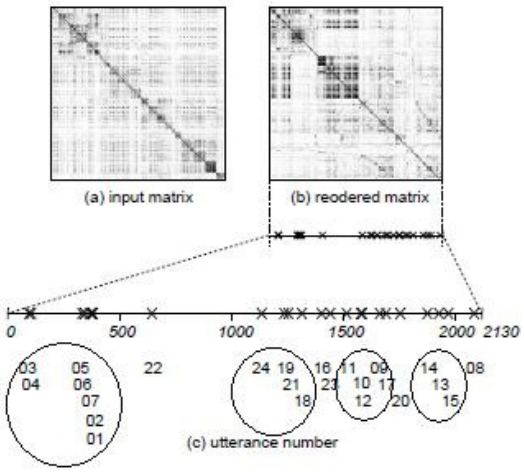

図 4　スペクトラルクラスタリングによるトピック分割

に関するドキュメントとそれ以外とに分類することでトピック境界を検出しようというものである。小規模なデータに対する実験により、その有効性と改善点を検討し、結果を発表した。

図 4 にクラスタリング結果を示す。図は、トピック境界が不明な連続したニュース音声の文章を音声認識した結果を用いて収集したウェブ文書をスペクトラルクラスタリングによりクラスタリングした結果である。連続したニュース音声が同一クラスタ内にある場合は同一トピックである可能性が高いことから、図に印をつけたように複数のトピックに分割することができた。

上記以外の研究成果として、提案手法の前処理としての音声区間切り出しおよび音声・非音声識別の高精度化を検討するため、音素認識における混同行列を利用した字幕テキストの自動対応付けを試み、その有効性について検討したものを国内外の学会で発表した。

(2)2008 年度

前年度までに検討したウェブドキュメントの収集方式および収集された文書をトピックごとにクラスタリングする方式について比較検討を行い、提案手法で用いるスペクトラルクラスタリングの有効性を確認した。またクラスタリングされた収集文書集合の分散から、検索に用いた語句の認識精度を推定することにより、収集文書の信頼性を判断することを検討し、信頼性の高い文書を用いることで認識精度を高めることができることを示した。

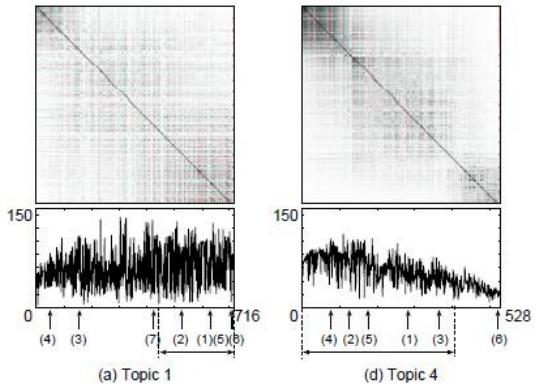

図 5 発話ごとのクラスタリング結果

表 1 単語正解率と F 値の比較

Topic	既存言語モデル		トピック特化言語モデル	
	WCR	F 値	WCR	F 値
1	62.5	49.5	84.4	81.4
2	68.8	62.3	71.5	74.7
3	72.4	72.0	74.1	79.5
4	64.7	55.7	71.1	68.5

図 5 は、トピック内に含まれる各発話が、収集したニュース記事をクラスタリングした結果のどの位置になるかを示したものである。図からわかるように認識対象のトピックから外れる語句を多く含む発話、すなわち認識精度が低いと推定される発話はクラスタ外に位置づけられることから、より類似した収集記事のみを利用してトピックに特化した言語モデルを作成する方がより効果的であると言える。表 1 は既存の言語モデルと上記のトピック特化言語モデルとで認識精度にどの程度差が生じるかを比較したものである。ここで、表中の WCR は単語正解率を示す。また F 値とは認識結果に含まれる名詞の再現率と適合率の調和平均である。表から分かるように提案手法により作成したトピック特化言語モデルを用いることで F 値が大幅に向上した。

またニュース文書を対象として収集した文書の中には 3 連接以上の熟語や固有名詞を含んだ名詞句が現れるケースが多く見られることから、従来の統計的言語モデルで用いられる 3 gram モデルをそのまま利用するのではなく、熟語や名詞句をそのままの形で言語モデルに未知語として登録する手法が有効であるとの予想を得ることができた。

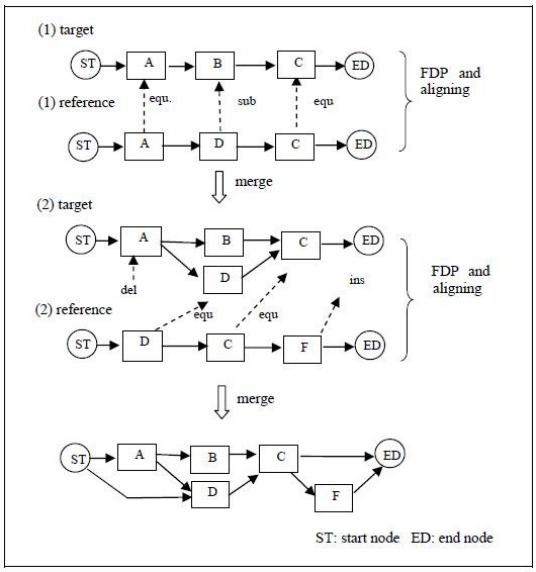

図 6　ネットワーク文法の自動構築

また上記の提案手法では学習データが小～中規模になってしまうことから、統計的言語モデルの学習データとしては不十分であると考えられるため、既存の統計的言語モデルと併用するためのネットワーク文法の自動構築手法についても検討した。さらに、ネットワーク文法構築の際の問題点である未知語および未知の文パターンに対する対応策を検討し、その有効性について検討したものを国内外の学会で発表した。

図 6 は中規模のテキストコーパスからネットワーク文法を自動構築するアルゴリズムの概要を示したものである。図のように単語をノードとするネットワークに入力文を対応付けしていくことで最終的なネットワークを得ることができる。ここで対応付けのアルゴリズムには DP マッチングを用いる。この時、学習コーパスに含まれない単語はネットワーク上に現れないことから、名詞を対象として意味素性を付与しておくことで未知語に対応することとした。実験の結果、クロ

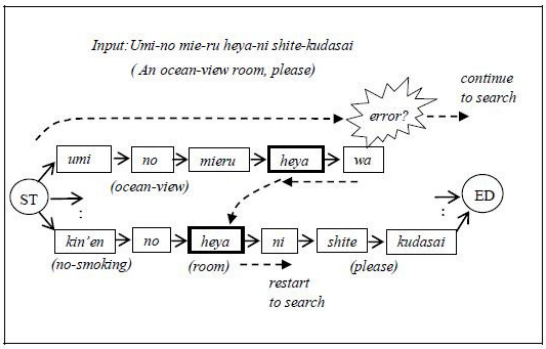

図 7　動的代替パス探索アルゴリズム

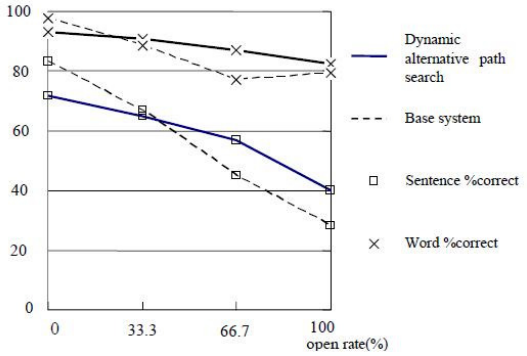

図 8　動的代替パス探索による認識結果

ーズドデータに対しては高い認識性能を得ることができたが、文パターン自体がネットワーク上のパスとして現れないクローズドデータに対しては十分な認識率を得ることができないことが示された。

(3) 2009 年度

前年度までの検討結果を受け、本年度は話し言葉を認識するためのネットワーク文法の検討を行った。具体的には、小～中規模の例文パターンから自動的に構築したネットワーク文法における、学習パターンに存在しない未知の文パターンに対する対応策を検討した。ここで用いた方法は、音声認識中の音響スコアが低いパスに対しては、ネットワーク中の可能な接続を探索するという方法である。その際、代替パスの接続方法として、認識中に動的に探索する方法と、順向きのパス探索結果と逆向きのパス探索結果とを比較して接続箇所を決定する方法の 2 通りについて検討を行った。その結果、動的に代替パスを探索する手法ではある程度の有効性が得られたが、双方向のパス探索結果を比較する方法では十分な性能が得られないという結論を得た。

図 7 は動的代替パス探索のアルゴリズムを示したものである。このアルゴリズムによる認識結果を図 8 に示す。ここで、図 8 の横軸はテストデータが学習コーパス中に含まれる割合（オープンデータ率）を示している。また実線が提案手法による結果、破線が従来手法による結果となっている。図からわかるように、提案手法は従来手法に比べ、オープンデータ率の増加に対する認識率の低下を抑えることができている。

さらに上記の手法を応用し、学習文パターン中には存在しない言い淀みや冗長語を動的に探索する手法について検討し、予備実験として小規模な例文パターンを対象として提案手法を用いない場合との比較を行った。さらに話し言葉における疑問・肯定を判別するためにピッチ抽出を利用した方法につい

ても検討した。

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 5 件）

① T. Morimoto, S. Takahashi, Speech Recognizer With Dynamic Alternative Path Search and It's Performance Evaluation, Intelligent Automation and Computer Engineering (Lecture Notes Elec. Eng.), 2010, 査読有, Vol. 52 (in Printing)
② T. Morimoto, S. Takahashi, Automatic Construction of a FSA Language Model and Speech Recognition on it with Dynamic Alternative Path Search, Proc. of the Int. Multi-Conference on Engineer and Computer Science, 2009, 査読有, Vol.1, pp.611-615
③ S. Takahashi, Topic Specific Language Model Based on Graph Spectral Approach for Speech Recognition, Trends in Intelligent Systems and Computer Engineering (Lecture Notes Elec. Eng.), 査読有, Vol.6, 2008, pp.497-514
④ S. Takahashi, T. Morimoto and Y. Nishimoto, Automatic Closed-Caption Alignment Using Pronunciation of Speech Recognition Transcripts for Public Relations TV Program, Proc. of the Int. Multi-Conference on Engineer and Computer Science, 2008, 査読有, Vol.1, pp.259-263
⑤ S. Takahashi, T. Morimoto and N. Tsuruta, News Topic Specific Language Model Based on Spectral Clustering and Web Crawling, IAENG International Journal of Computer Science, 査読有, Vol.34, No.2, 2007, pp.208-213

〔学会発表〕（計 3 件）

① 高橋伸弥、森元逞、FSA 言語モデルの自動構築と動的代替パスサーチによる音声認識、情報処理学会音声言語情報処理研究会、2008/10/24、東京工業大学
② 高橋伸弥、森元逞、西本由之、類似音素行列を用いた音声認識結果とキャプション文字列との自動対応付けに関する検討、情報処理学会全国大会、2008/3/13、筑波大学
③ 高橋伸弥、森元逞、鶴田直之, Web 上の類似文書にするクラスタリング結果を用いたニュース音声のトピック分割と音声認識結果の信頼度の判定, 電気関係学会九州支部第 60 回連合大会, 2007/09/19, 琉球大学

６．研究組織
(1)研究代表者
高橋　伸弥（TAKAHASHI SHINYA）
福岡大学・工学部・助教
研究者番号：４０３３０８９９