様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成　25　年　5　月　31　日現在

機関番号：82626
研究種目：若手研究(B)
研究期間：2011～2012
課題番号：23700225
研究課題名（和文）Ｗｅｂ音声インデキシングのための言語的特性の変動に頑健な音声認識に関する研究
研究課題名（英文）A Study of Speech Recognition under Various Linguistic Conditions for Web Audio Indexing
研究代表者
緒方　淳　（OGATA JUN）
独立行政法人産業技術総合研究所・情報技術研究部門・研究員
研究者番号：10392599

研究成果の概要（和文）：Web 音声インデキシングのための音声認識の高度化に関する研究を行った。本研究では、事前のコーパスを前提とした従来の音声認識の問題点を解決するために、不特定多数のユーザからの協力（集合知）を活用した言語モデリングを提案、開発した。評価実験の結果、開発した言語モデリング手法が、実際の Web 音声コンテンツに対する音声認識において有効であることが示された。

研究成果の概要（英文）：In this research, we studied automatic speech recognition for Web audio indexing. To overcome difficulties in preparing task-specific corpora in advance, we proposed and developed a language modeling method on the basis of wisdom of crowds and web-text resources. The experimental results have shown that the proposed language modeling can significantly improve the recognition performance in transcription of web-audio content.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	3,300,000	990,000	4,290,000

研究分野：総合領域
科研費の分科・細目：情報学、知能情報処理・知能ロボティクス
キーワード：音声認識・理解、情報検索

１．研究開始当初の背景

YouTube に代表される動画共有サービス、ポッドキャストの普及により、Web 上では大量の音声コンテンツが日常的に生成・蓄積されるようになった。特に最近は、米国大統領選挙など、社会的関心の高い出来事はこうしたコンテンツとして積極的に公開・利用されるなど、政治・経済・社会の様々な場面で大きい影響を与えるようになっている。このような Web 上の膨大なコンテンツからユーザが欲しい情報を自由に検索したり、快適な鑑賞を行うためには、コンテンツに含まれる音声情報を計算機が的確に理解し、索引情報を付与する「Web 音声インデキシング」が重要となる。そして、この Web 音声インデキシングを実現するために必要不可欠となるのが音声認識技術である。音声認識技術は、近年、放送ニュース、講演、会議など、実環境のタスク(音声認識が利用される環境、状況)を想定した研究開発が多くなされ、いずれの場合においてもそれぞれのタスクに合致した膨大なコーパス(音声データとその書き起こ

し)を利用することで大きな改善が得られている。一方、本研究で対象とする Web 音声データは、その発話内容や録音環境などが多種多様であるという特徴を持っているため、出現する全てのタスクに対してコーパスを事前に構築することは現実的に不可能である。したがって従来の音声認識技術では、Web 音声データを正しく認識することは困難であった。

２．研究の目的

本研究では、Web 上の様々な言語資源や Web サービスを通じた集合知(不特定多数のユーザからの協力)を利用した音声認識手法を構築することで、Web 音声データに対する音声認識性能をいかに向上させることができるかを探求する。

３．研究の方法

本研究では特に、音声認識を構成する要素としてもっとも重要な言語モデルに着目する。具体的には Web 音声データにおける言語的特性の変動について下記の課題に取り組むことで、Web 音声データに対する音声認識性能を向上させ、Web 音声インデキシングの実用性を高めることができるかを明らかにする。

- 幅広い話題・語彙への対処：
  Web 音声データは、政治、経済、スポーツ、芸能といったように話される話題が多岐に渡る。また、タスクもニュース、講義、雑談といったように多様で、事前に絞り込むことができない。
- 日々生まれる新しい言葉への対処:
  Web 音声データには日々更新・アップロードされていくという特徴がある。そのため、従来の音声認識とは違い、話題や語彙が日々変わっていき、流行語や新出語が次々と生まれる。
- 様々な発話スタイルへの対処：
  Web 音声データは、ニュースにおけるアナウンサー口調、雑談における自由な発声など、発話スタイルも様々である。特に雑談等のくだけた発声は、従来の音声認識研究で構築されたコーパスではカバーできず認識は困難である。

上記の課題を解決するために、Web の総合的ニュースサイト(Web ニュース)上の大量のテキスト記事や Web キーワード辞書サービスといった Web 上の言語資源を有効活用した言語モデリングの構築に取り組む。さらに、本研究の特徴的なアプローチとして、申請者らが開発、運用している音声情報検索 Web サービス「PodCastle」を通じた集合知に基づく言語モデリングの構築に取り組む。PodCastle は、ユーザが音声認識誤りを容易に訂正できる機能を持っており、これにより日々訂正情報（集合知）が蓄積され、音声データとともにその書き起こしが収集できる。本研究では、PodCastle がより多くのユーザに活用されるとともに、こうした訂正情報をより多く収集することができるように、Web サービスとしての機能拡張やインタフェースの改良にも取り組む。

図 1 動画へ対応した PodCastle インタフェースの画面例：YouTube、ニコニコ動画、Ustream といった主要な動画共有サービスのコンテンツの全文検索・閲覧・編集（音声認識結果の訂正）が可能。

図 2 タブレット・スマートフォン向け PodCastle インタフェースの画面例

４．研究成果

（１）音声情報検索 Web サービス「PodCastle」の機能拡張

申請者らは Web 音声インデキシングの一環として、Web 上の音声コンテンツの 1 つであるポッドキャストを対象とした音声情報検索 Web サービス「PodCastle」の開発を行ってきた(http://podcastle.jp)。PodCastle は、ポッドキャストを音声認識技術によって自動的にテキスト化・索引付けすることで、それらをユーザが全文検索できるだけではなく、Web ブラウザを通じて詳細な閲覧、編集も可能にする「ソーシャルアノテーションシステム」である。本研究では、より一層多くのユーザの参加、利用を促して幅広いタスク、発話スタイルのテキストを収集するために、

近年急速に普及が進んでいる動画共有サイトの動画データも扱えるようにPodCastleを拡張した。これによりポッドキャスト同様に動画の全文検索、閲覧、編集が可能となった。動画共有サイトとして YouTube、ニコニコ動画、Ustream といった現在主流のサービスをカバーし、対象とすることで、膨大かつ幅広い種類のWeb上のコンテンツを扱えるようになった(図1)。

また、タブレットPC、スマートフォンといった近年普及が著しい携帯デバイスにおいても、PodCaslte の円滑な利用が可能となるように、インタフェースの拡張も行った（図2)。ただし、現状は iPad、iPhone といったiOSのみの対応となっている。

（2）Web 上の言語資源を活用した言語モデリングの高度化

本研究ではWeb音声インデキシングの性能を向上させるために、Web 上の言語資源を有効活用した言語モデリングの検討を行った。提案する手法は、様々なトピックをカバーするWebニュースステキストをベースにしてメイン言語モデルを構築し、さらにその特性を活かして、認識対象ごとのトピックに合致するよう言語モデルのパラメータを最適化することで動的な言語モデル適応を行うものである。

（3）言語モデルが日々継続的に育つ仕組みの確立

Web ニュースステキストには、音声認識の言語モデリングにおいて有用となり得る、2 つの大きな特徴があるといえる。まず、一般的なニュースアグリゲーションWebサイトでは、様々なニュース配信サービスからの幅広い内容に関するニュース記事が集約されており、それらの記事はユーザが閲覧しやすいように複数のトピック、カテゴリごとに分類されている。そして2つ目としては、日常的に記事が更新される仕組みにより、一般社会における最新のトピック・語彙がカバーされている点である。本研究では、言語モデルにおけるトピックの多様性に対処するために、ニュースアグリゲーションWebサイトの1つである Yahoo! Japan ニュースの膨大なニュース記事を利用する。Yahoo! Japanニュースでは、全てのニュース記事が、6 メイントピック、25サブトピック（以降トピックと呼ぶ）からなる階層構造上に分類されている。本研究では、こうした日々配信される最新のニュースステキストをもとに,メイン言語モデルを日常的に自動更新可能とした.これにより,言語モデルが日々移り変わる世の中の情勢や話題を逐次追従していく（日々育っていく）仕組みを実現した。

（4）Web ニュースステキストを活用した動的言語モデリング手法

Web ニュースステキストに基づくトピック言

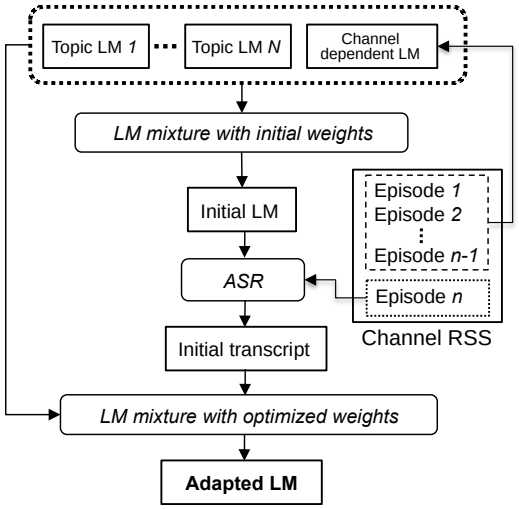

図 3 動的言語モデリングの全体図

語モデルを利用して、適応言語モデルを動的に生成する（図3)。本手法は、各トピック言語モデルを用いたモデルレベル混合手法に基づいている。モデルレベル混合では、複数の要素モデルの N-gram 確率を重み付きで線形補間する。

入力音声(各チャンネルのエピソード)ごとの動的プロセスとして、まず、初期言語モデル（25トピックの各Webニュース言語モデルを均一重みで線形補間したモデル）を用いて音声認識を行い、初期認識結果を生成する。そして、初期認識結果を用いて各トピック言語モデルの混合重みを動的に算出する。すなわち、初期認識結果のテキストを前述のヘルドアウトセットとして、混合結果のモデルが最小のパープレキシティを示すように EM アルゴリズムにより混合重みを推定する。そして、算出した混合重みを基にトピック言語モデルを混合し、入力音声のトピックに適応化した最終的な言語モデルを出力する。

（5）Web サービスを通じた集合知をした動的言語モデリング手法

PodCastle を通じて得られる集合知（ユーザによる訂正情報、書き起こし）を利用することで、Web音声コンテンツごとのトピック、ドメイン、発話スタイルに特化した動的言語モデリングを実現した。ここでは、認識対象エピソードと同じチャンネル内の他の(過去の)エピソードデータを利用して言語モデルを構築し(チャンネル依存言語モデル)、これを前述の動的言語モデリングシステムに組み込んだ。この理由としては、同一のチャンネル中の各エピソードは、同じ言語的特性(トピック、発話スタイル等)を持っている可能性が高いことが挙げられる。さらに、チャ

ンネルを構成するRSSの仕組みにより、認識対象となる各エピソード音声データがどのポッドキャストに属するのか、すなわち、各音声ごとにどの言語モデルを動的言語モデリング時に適用すべきかが自明であるという利点もある。拡張システムでは、まず事前にチャンネル依存言語モデルを、認識対象エピソード以外の過去のエピソードを利用して学習しておく。この際、過去のエピソードの音声認識結果のテキストをもとに言語モデルの学習が行われるが(教師なし)、PodCastle を通じてユーザによる訂正がなされていれば、より正確な書き起こしから学習が行われ、より精度の高い言語モデルが構築できる。

以上の動的言語モデリング手法を実装し、実際のWeb音声コンテンツ（本実験ではポッドキャスト8番組）を用いて認識実験を行ったところ、最終的に9.9%のエラー削減率が得られた(表 1)。特に、集合知を活用した動的言語モデリングに大きく性能を向上可能なことが示された。

本研究で構築した言語モデリング手法は、実運用中のWebサービスと密接に連携することを想定したものであり、今後Webサービスの利用が広がるにつれて、さらなる音声認識性能の向上が期待できる。今後は、言語モデルだけでなく、その他の要素技術についても、こうしたWebサービスと連携した手法の確立を行っていく。

ベースライン	動的LM（教師なし）	動的LM（集合知活用）
35.3 %	32.7 %	28.9 %

表1 音声認識実験結果(単語誤り率)

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）

〔学会発表〕(計 1件)

① Jun Ogata and Masataka Goto: “PodCastle: Collaborative Training of Language Models on the Basis of Wisdom of Crowds”, Interspeech2012, 2012/09/13, Portland, USA.

〔その他〕
広報

① 産総研プレスリリース「インターネット上の動画音声データの検索・書き起こしシステムを実現」、2011/10/12、http://www.aist.go.jp/aist_j/press_release/pr2011/pr20111012/pr20111012.html

報道

① 日刊工業新聞 2011年10月13日(木) 21面
② 日刊スポーツ新聞 2011年10月13日(木) 18面
③ フジサンケイビジネス 2011年10月13日(木) 6面
④ 日本情報産業新聞 2011年10月24日(月) 2面
⑤ 産経新聞 2011年12月9日(金) 23面
⑥ I/O 平成24年1月号 (2012年1月1日発行) 「コンピュータの未来技術[第69回] PodCastle」, Vol.37, No.1, pp.84-86

ホームページ

① PodCastle(ポッドキャッスル) http://podcastle.jp/

６．研究組織
(1)研究代表者
緒方 淳 (OGATA JUN)
独立行政法人産業技術総合研究所・情報技術研究部門・研究員
研究者番号： 10392599