

音声認識辞書構築コスト削減に向けた取り組み

北出祐・西光雅弘・三木清一
荒井一博・江森正

要旨

人と人の自由な会話を認識するための音声認識辞書を構築するには、対象とする会話音声の大量の書き起こしテキストが必要となります。このようなデータを用意するには多大なコストを要します。そこで我々は、認識対象とする音声と関連する情報を、既に保有しているデータから抽出したり、外部より自動で収集したりすることで、音声認識辞書の構築コストを低減させる方式を開発しました。

キーワード

●音声認識辞書 ●議事録作成支援ソリューション ●コンタクトセンター ●インターネット

1. まえがき

音声認識技術は人間が話した音声をテキスト化する技術で、音の特徴を表現する音響モデルと、単語の表記や読み、単語の並びを格納した音声認識辞書が用いられます（図1）。本稿では、音声認識辞書の構築に関する取り組みについて紹介します。

我々は現在、コンタクトセンター向け音声認識製品VisualVoiceや、議事録作成支援ソリューションVoiceGraphyといった、人と人の自由な会話を精度よく認識できる、会話音声認識システムの開発を行っています。自由な会話では、どのような単語の並びで文が発話されるかを、ルールのような形であらかじめ決めておくことができません。これに対し、認識対象とする会話を書き起こしたテキストから、単語と單

語のつながり方を統計的に学習することで、音声認識辞書を構築する手法があります。しかし、このような手法で品質の高い音声認識辞書を構築するには、大量の書き起こしテキストを集め必要があります。これには、音声の録音や書き起こしなど、多大なコストがかかります。

本稿では、音声認識辞書構築コスト削減の取り組みについて、3つのトピックを紹介します。

まず1つ目は、コンタクトセンターにおける顧客とオペレータの通話音声を認識するための音声認識辞書構築における取り組みです。挨拶のような共通に用いられる表現を抽出・活用することで音声認識辞書構築コストの削減を実現しています。

2つ目は、決算説明会や株主総会などの一般企業での会議向けの議事録作成支援ソリューションでの取り組みです。会議内容に関する文書を、コンタクトセンター通話などのほかの会話音声の書き起こしテキストと組み合わせることで効率良く高精度な音声認識辞書を構築する手法を開発しました。

3つ目は、テキストデータの収集など、音声認識辞書を構築することそのものにかかる人手を削減する取り組みです。インターネットからのテキスト自動収集手法などを紹介します。

2. コンタクトセンター通話音声認識

コンタクトセンターにおけるオペレータと顧客の会話には、以下の特徴があります。1つは、コンタクトセンターごとに業種や業務内容が異なるため、会話の内容が大きく異なること

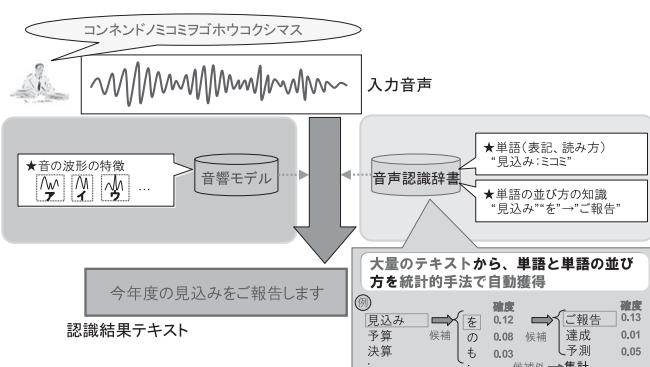


図1 音声認識の仕組み

研究開発

音声認識辞書構築コスト削減に向けた取り組み

です。例えば、製品の障害対応業務の場合には、製品名や状況説明のような、業務特有の単語が多数含まれます。一方、「お電話ありがとうございます」といった、業種・業務内容に依存しない、コンタクトセンターであれば共通に用いられる表現もまた多く存在します。このような特徴に基づき、複数のコンタクトセンターの会話の書き起こしテキストから共通に用いられる表現を抽出し、コンタクトセンター向け共通音声認識辞書を開発しました。この共通音声認識辞書に対し、各コンタクトセンターに特有の単語を追加することで、音声認識辞書を効率良く構築することができます。

コンタクトセンターの通話音声を用いて音声認識実験を行い、音声認識辞書構築の効率化を確認しました。書き起こしテキスト量をこれまでの1/10に削減すると、削減しない場合と比べ音声認識率は約4.2%低下します。一方、この1/10に削減した書き起こしと、今回開発した共通音声認識辞書を組み合わせて用いることで音声認識率の低下を約1.6%にとどめることができました。今後、認識精度を更に高めるため、応対履歴や製品カタログといった、書き起こしテキストと比べよりコストの低いテキストを用いて、コンタクトセンター特有の単語を効率良く音声認識辞書に追加する手法を検討していきます。

3. 会議音声認識

会議音声を認識する場合には、会議の書き起こしテキストから音声認識辞書を構築することが好ましいですが、そのためには多大なコストが必要となります。一方、決算説明会や株主総会など、会議で議題となるような、決算情報などの企業経営に関する情報は、ウェブなどから比較的容易かつ大量に入手できます。また、多くの企業について、過去の決算説明会・株主総会などの議事録入手することが可能ですが、しかし、これらのほとんどは整形・要約された「書き言葉」であり、「話し言葉」である実際の会議音声を認識するための音声認識辞書の学習データとしては不適切です。書き言葉と話し言葉を比較すると、出てくるキーワードは同じですが、文末などの言い回し（言葉遣い）が大きく異なります。これに対し、ウェブなどで入手できる書き言葉テキストデータを有効に活用しながら、既に保有する話し言葉テキスト（例えばコンタクトセンター通話書き起こし）を利用して、会議音声認識のための音声認識辞書を効率的に構築する方法を開発

しました。

企業情報に関するテキストと、会議音声とは異なる会話音声の書き起こしとを組み合わせる場合、異分野の書き起こしをそのまますべて用いると、その書き起こしに含まれる固有名詞などの分野依存の単語（以下、異分野固有単語）が認識結果に出力され、認識精度が低下したり、ユーザが違和感を覚えるような不自然な認識結果が得られたりします。

これに対し、書き起こしから異分野固有単語を抽出し、それらを取り除いた単語列のみを用いて音声認識辞書を構築します。これにより、異分野固有単語を音声認識辞書に追加することなく、話し言葉に共通に出現する言い回し表現を強化することができ、会議音声の認識精度を大幅に改善することができます。

先に述べた方法の有効性を確かめるため、以下の3種類の音声認識辞書を比較評価しました。1つはベースラインとなる音声認識辞書で、ウェブなどの書き言葉テキストを学習に用いて構築した音声認識辞書です。2つ目は、それに、異分野の会話音声の書き起こしテキストをそのまま全て加えたものです。最後は、異分野の会話音声の書き起こしを選択的に加えて構築した音声認識辞書です。評価には、NECなど企業の決算説明会の音声（5会議分）を用いました。

ベースラインと、異分野の会話音声の書き起こしを全て追加した場合を比較すると、後者で4%以上の認識誤り削減を確認しました。更に、今回開発した手法を用いることで、会話音声の書き起こしをすべて追加する場合と比べ、誤挿入される単語が削減され、異分野固有単語が削除されることを確認しました。今後は、書き起こしの選択的追加方法の更なる改良などに取り組んでいきたいと考えています。

4. 音声認識辞書構築作業の効率化

音声認識辞書を構築する際、図2に示すような作業が必要となります。図2の(1)～(3)の方法により、コストの高い作業を自動化することを検討しました。

(1)検索キーワードの自動抽出と、インターネットからのテキスト自動収集

認識する対象の会話の種類が決定した後、その内容が含まれる少量のテキスト（以下、シードテキスト）を見本として用意し、そのテキストから頻度の高い単語を検索キーワードとしてシードテキストと関連のあるテキストをウェ

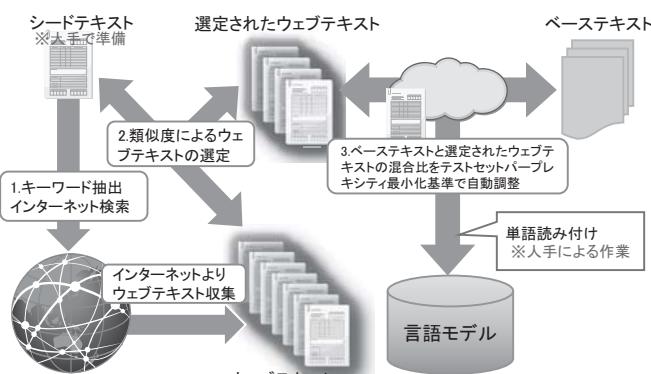


図2 音声認識辞書構築作業

から自動的に収集します。従来は、会話の内容と関連するテキストを人手で大量に集める必要がありました。今回開発した方式では少量のシードテキストを準備するだけで済みます。

(2) 類似度によるウェブテキストの選定

(1) で収集したウェブテキストはシードテキストに存在するキーワードを含むという意味で関連のあるデータになりますが、必ずしもシードテキストと内容が類似しているとは限りません。そこで、シードテキストと、自動収集されたウェブテキストとの類似度を計算し、類似度の高いテキストを選定して音声認識辞書の学習に用います。

(3) テストセットパープレキシティ最小化基準による混合比の決定

(2) で選定したウェブテキストを、ほかのテキストに混ぜる際の重みを決定するため、音声認識辞書の評価尺度の1つであるテストセットパープレキシティを最小化する基準を用います。この基準で混合比を決定することで、自動的に精度の高い音声認識辞書を構築することが可能となります。

今回の手法の有効性を確認するため、NECなど企業の決算説明会を対象に評価実験を行いました。シードテキストとしては、対象とする企業のウェブテキスト（幹部メッセージなど）を用いました。実験の結果、従来の人手で収集したテキストを用いて学習した音声認識辞書と比べ、今回の手法で構築した音声認識辞書の方がテストセットパープレキシティが小さくなり、より精度が高い音声認識辞書となっていることを確認しました。また、実際に音声認識辞書を構築するのにかかる期間を比較したところ、従来と比べおよそ1/4に短縮さ

れることを確認しました。今後、コンタクトセンターやほかの会議についても評価を行い、今回開発した方式を更に改良していくたいと考えています。

5. むすび

本稿では、自由な会話音声を認識するための音声認識辞書の構築コストを削減する取り組みについて紹介しました。具体的には、認識対象の性質に応じ、共通する表現が多い場合（例えばコンタクトセンター向け）にはそれらを表す共通音声認識辞書を用意したり、内容に関するテキストデータを比較的容易に大量に入手可能な場合（例えば決算説明会など一般企業の会議）にはそれをベースとして話し言葉表現を選択的に追加することを行っています。また、見本となる少量のテキストを元に、外部から類似するテキストを自動で取得して音声認識辞書の学習に用いる方法を紹介しました。今後は、データの選択方法や、自動取得方法を更に洗練することで、構築される音声認識辞書の質を高めることを検討し、合わせて、さまざまな実データで検証を行って有効性や頑健性を確認していきます。また、その成果を活かし、今以上に高い音声認識性能を持つ製品をご提供していきたいと考えています。

執筆者プロフィール

北出 祐
共通基盤ソフトウェア研究所
日本音響学会会員

西光 雅弘
共通基盤ソフトウェア研究所
日本音響学会会員

三木 清一
共通基盤ソフトウェア研究所
主任研究員
情報処理学会
日本音響学会各会員

荒井 一博
NEC情報システムズ
先端技術ソリューション事業部

江森 正
NEC情報システムズ
先端技術ソリューション事業部
エキスパート
日本音響学会会員