

8言語をリアルタイムに識別でき、言語設定が不要

～何語を話しているか分からない相手の言葉も即座に識別・認識・翻訳～

【ポイント】

- 8言語の短い音声をリアルタイムで識別し、翻訳の言語設定が不要のシステムを開発
- 発話時間1.5秒の音声を識別率9割以上、識別時間0.15秒以内のリアルタイム識別を実現
- 何語を話しているか分からない外国人の言葉も即座に識別し、音声認識や自動翻訳が可能に

国立研究開発法人情報通信研究機構(NICT、理事長: 徳田 英幸)は、人の発話は何語であるのかを入力音声のみから識別する8言語(日、英、中、韓、タイ、ミャンマー、ベトナム、インドネシア)の言語識別技術を開発しました。従来の一般的な方式では、10秒程度の長い発話でないと識別が困難であるため、実際の音声アプリケーションではあまり使われていませんでした。これに対して、言語識別に必要な発話の特徴を精度よく抽出し、かつ、高速演算できるニューラルネットワークを提案し、1.5秒程度の短い発話でも0.15秒以内に即座に識別できる技術を開発しました。これにより、何語を話しているか分からない外国人の言葉も即座に識別し、言語設定の必要もなく音声認識や自動翻訳ができるようになります。

この成果について、10月25日(木)～27日(土)のけいはんな情報通信フェアにてデモ展示します。

【背景】

NICTでは、多言語音声認識、機械翻訳、音声合成の研究開発に取り組んでおり、スマートフォン用アプリケーション「VoiceTra」を実証実験のために無料公開しています。音声アプリケーションの大半は、入力される言語が何語であるかをあらかじめ指定する必要があるため、ユーザにとって不便でした。さらに、相手が話している言語が分からない場合には、何語かを指定すること自体が困難でした。

これを解決する方法として、入力発話は何語なのかを識別する言語識別技術があります。従来の一般的な方式では、10秒程度の長い発話でないと識別が困難であるため、実際の音声アプリケーションでは使いにくいという課題があります。



【今回の成果】

今回NICTは、この課題を解決するために、知識蒸留^{*1}を用いて、言語識別に必要な発話の特徴を精度よく抽出できる長い発話用のニューラルネットワークを変換して、短い発話でも識別精度が高く、かつ、リアルタイムで識別可能な小規模ニューラルネットを構築する方式を提案しました。

この方式により、1.5秒程度の短い発話でも即座(0.15秒)に識別できる技術を開発し、8言語(日、英、中、韓、タイ、ミャンマー、ベトナム、インドネシア)で90%以上の識別率を実現しました。この技術により、これまで必要であった入力言語の事前指定が不要となるため、何語を話しているか分からない外国人の言葉も即座に識別し、音声認識、機械翻訳ができるようになります。

【今後の展望】

今後は、識別言語数の拡張、識別精度向上と共に、この機能を VoiceTra 等の音声翻訳アプリケーションに実装するとともに、民間企業にライセンスする予定です。

この成果について、10月25日(木)～27日(土)のけいはんな情報通信フェアにてデモを展示します。

<http://khn-fair.nict.go.jp/>

＜用語解説＞

*1 知識蒸留

規模の大きいニューラルネットは識別精度が高い代わりに、計算に時間を要する問題があり、リアルタイム処理が求められる実際のアプリケーション等には向きません。一方、高速演算可能な小規模ネットワークは、そのまま学習させただけでは識別精度が悪いという問題があります。

これを解決するために、図1左のように、まず長い発話を高精度に識別可能な大規模ネットワーク(Teacher)を学習します。このネットワークは、入力する発話が長い場合、言語識別に必要な音声の特徴を精度よく抽出できますが、計算には時間がかかります。次に、図1右に示す高速演算可能な小規模ネットワーク(Student)を学習する際、Teacherの優れた音声特徴抽出とできるだけ同じ結果が出るように学習を行います。このような学習法を知識蒸留学習と言います。これにより、高速演算が可能でかつ識別率の高い小規模ネットワークを構築できます。

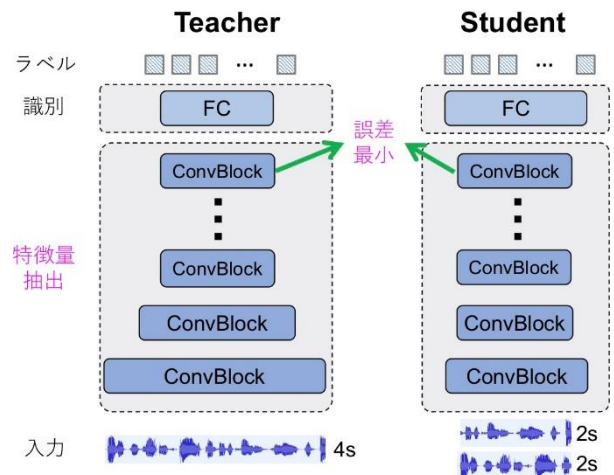


図1: 提案法における知識蒸留学習

＜性能比較＞

提案法の性能を評価するために、従来の一般的な方式と提案法を10言語(日、英、中、韓、タイ、ミャンマー、ベトナム、インドネシア、フランス、スペイン)の誤識別率にて比較した結果を図2に示します。いずれも、リアルタイムに識別できる方式です。結果から、提案法は短い発話識別において従来法よりも高精度に識別できることが分かります。

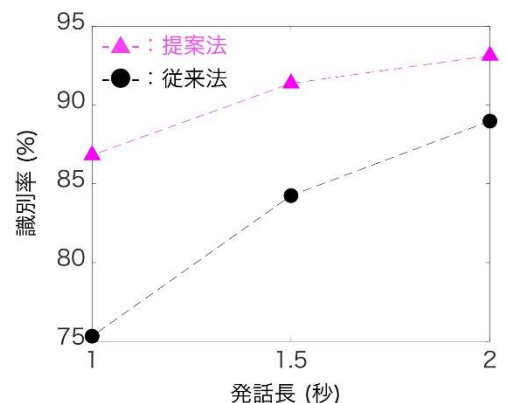


図2: 10言語識別の識別率の比較

＜関連技術との比較＞

言語識別の公開アプリケーションとしては、Google社から提供されている音声翻訳アプリに搭載された2言語識別機能及び4言語識別APIがあります。

<https://cloud.google.com/speech-to-text/docs/multiple-languages>

これらは、スマートフォン用音声翻訳アプリケーションやスマートスピーカにおいて、指定した2言語及び4言語の中から言語識別を行う機能です。これに対して、今回NICTの開発した方式は、2言語や4言語ではなく、倍の8言語を即座に識別できます。

＜ 本件に関する問い合わせ先 ＞

国立研究開発法人情報通信研究機構
先進的音声翻訳研究開発推進センター
先進的音声技術研究室
河井 恒
Tel: 0774-98-6314
E-mail: ict@khn.nict.go.jp

＜ 広報 ＞

広報部 報道室
廣田 幸子
Tel: 042-327-6923
Fax: 042-327-7587
E-mail: publicity@nict.go.jp