

音声全文検索・書き起こしサービス PodCastle

多数のユーザーの協力によって音声認識性能が日々向上



後藤 真孝

ごとう まさたか (左)
 m.goto@aist.go.jp
 情報技術研究部門 上席研究員
 (兼) メディアインタラクション
 研究グループ長(つくばセンター)

メディアコンテンツを計算機で自在に扱える技術の確立を目指し、音楽・音声の音響信号の自動理解と、それに基づくユーザーインタフェースの研究を中心に、さまざまな研究課題に取り組んでいます。

緒方 淳

おがた じゅん (右)
 jun.ogata@aist.go.jp
 情報技術研究部門
 メディアインタラクション研究
 グループ
 研究員 (つくばセンター)

機械による自動音声認識・理解に関する研究に取り組んでいます。産総研入所後は、実用化を意識し、実環境向けの音声認識技術やインタフェース、さらにインターネットサービスの開発を行っています。

関連情報：

- 参考文献
 後藤 真孝 他：人工知能学会論文誌，25(1)，104-113，(2010)。
- 用語説明
 ＊ポッドキャスト：インターネット上で音声データを公開するために用いられている方法の一種。一連のエピソードと呼ばれる音声データ(MP3ファイル)に加え、その流通を促すためのメタデータRSSが付与される。
- プレス発表

2008年6月12日「ユーザーが協力すると性能が向上する音声情報検索システムを実現」

2011年10月12日「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」

テキストで書き起こす必要性

インターネット上の動画共有サービスや音声配信サービスの普及に伴い、誰でも視聴できる音声を伴う動画データや音声データが日常的に生成・蓄積され増え続けています。しかし、動画音声データはテキスト(文字)データと異なり、それ自体を索引として使えないため、発言内容などの詳細な情報の検索(全文検索)ができませんでした。

この全文検索を可能にするために、これまで音声認識技術を使って音声を自動的にテキスト化する取り組みはありましたが、音声認識誤りを避けることはできず、ユーザーが誤りに気づいても正しい書き起こしを作成するために協力し合うことはできませんでした。

動画共有サービスに対応したPodCastleを公開

産総研では、音声認識誤りをユーザーが訂正できる独自のインタフェースを開発し、2008年に日本語のポッドキャスト*を対象とした音声全文検索サービスを一般公開しました。16万件以上の音声データが登録され、実際に58万単語以上が訂正されました。それらを学習することで音声認識性能の向上が可能なることを実証しました。さらに動画や英語への対応が求められてきました。

今回私たちは、ユーザーに協力してもらうことで性能が日々向上する音声情報検索技術を開発

し、インターネット上にある日本語と英語の動画音声データを対象とした音声全文検索・書き起こしサービス「ポッドキャストル (PodCastle)」(日本語版 <http://podcastle.jp>、英語版 <http://en.podcastle.jp>)を2011年に一般公開し、実証実験を開始しました(図1)。

このPodCastleでは、普及が進んでいる代表的な動画共有サービス(ニコニコ動画、YouTube、Ustream)に新たに対応しました。そして、複数のユーザーが協力し合って、話者名や改行を入力しながら読みやすい書き起こしを作成できる機能を実現しました(図2)。さらに、エジンバラ大学音声技術研究所の協力を得て、英語の動画音声データも音声認識と検索ができるように改良しました。誰でも無料で利用できる便利なサービスなので、多くの方々に使っていただければと願っています。

今後の予定

PodCastleの持続的な研究開発・運用を通じて社会に貢献していくとともに、この研究で開発した、動画音声データに対する検索・書き起こしシステムを、産業界と連携して実用化し、動画視聴の効率化や聴覚障がい者支援、会議議事録作成など、さまざまな応用に展開していく予定です。また日本語、英語以外の言語にも今後対応していくことを検討しています。



図1 実証実験中の音声全文検索・書き起こしサービス「ポッドキャストル (PodCastle)」



図2 音声認識の誤認識を訂正し、書き起こしを作成できるインタフェースの画面例