

プレスリリース
2017年6月28日

国立研究開発法人情報通信研究機構



ニューラル機械翻訳で音声翻訳アプリ VoiceTra が更なる高精度化を実現

～話し言葉の翻訳精度が大幅アップ・洗練された表現～

【ポイント】

- 日・英、双方向の話し言葉の翻訳を対象として、ニューラル機械翻訳を開発
- 従来技術（統計翻訳）に比べ大幅に高い9割前後の翻訳精度を達成
- 多言語音声翻訳アプリ VoiceTra にニューラル機械翻訳技術を組み込んで公開

国立研究開発法人情報通信研究機構(NICT、理事長: 徳田 英幸)の先進的音声翻訳研究開発推進センターは、ニューラルネットワーク(深層学習)を用いた機械翻訳の実用技術を開発しました。

ニューラル機械翻訳(NMT)^{*1}技術を用いることにより、従来の統計翻訳(SMT)技術に比べ、精度が大幅に向上しました。本日から、NICTが開発し、公開している音声翻訳アプリ VoiceTra(ボイストラ)^{*2}に同技術が組み込まれており、精度改善を直接ご体験いただけます。

* 本件には、総務省の情報通信技術の研究開発「グローバルコミュニケーション計画の推進-多言語音声翻訳技術の研究開発及び社会実証-I。多言語音声翻訳技術の研究開発」における社会実証用コーパスの構築の成果を用いています。

【背景】

訪日観光客数は2016年に2400万人を超え、2020年には4000万人に達する勢いになり、多言語音声翻訳へのニーズは日本各地で高まっています。これに応えるために、NICTが開発したVoiceTraの音声翻訳技術の製品化が、既に始まっています。

一方で、近年の人工知能(AI)の進展は目覚ましく、ニューラルネットワーク(深層学習)に基づく実用化事例も増え、機械翻訳への適用も活発化し始めました。2013年から、ニューラル機械翻訳(NMT)の基礎研究を推進してきたNICTも、2016年12月にその実用化に着手しました。

【今回の成果】

- NICTは、対訳コーパスから自動翻訳の構築にニューラルネットワークを利用した機械翻訳を2013年から研究を開始し、2016年12月から実用化を加速しました。
- 実用化には、様々な分野において話し言葉の対話で使われる固有名詞や専門用語辞書の追加が不可欠であり、この手段として「単語とその訳語及びそのクラス(意味分類)」を利用する手法を実装しました。
- 翻訳精度(意味が通じる率を翻訳者が評価)を大幅に改善し、多分野(防災、買物、タクシー等)で9割前後の高精度を達成しました。
- 既にニューラルネットワーク化されている音声認識部分に加えて、今回、翻訳のニューラルネットワーク化を実現し、一般公開されている多言語音声翻訳アプリ VoiceTra のニューラルネットワーク化を前進させました。

【今後の展望】

今回は、日英翻訳の双方向のみですが、グローバルコミュニケーション計画³の目標とされる全10言語(日本語、英語、中国語、韓国語、タイ語、インドネシア語、ベトナム語、ミャンマー語、スペイン語、フランス語)をカバーできるよう順次拡大を進めてまいります。

さらに、積極的な技術移転により、民間での製品化を推進してまいります。

また、現在、特許等の書き言葉のために翻訳エンジンのニューラルネットワーク化⁴も推進しています。

<用語解説>

*1 ニューラル機械翻訳(Neural MT)

脳の神経回路を模したニューラルネットワークを用いた自動翻訳技術。膨大な対訳データから学習したニューラルネットワークを用いて翻訳することで、従来技術よりも高い翻訳精度が実現可能になることが確認されつつある。

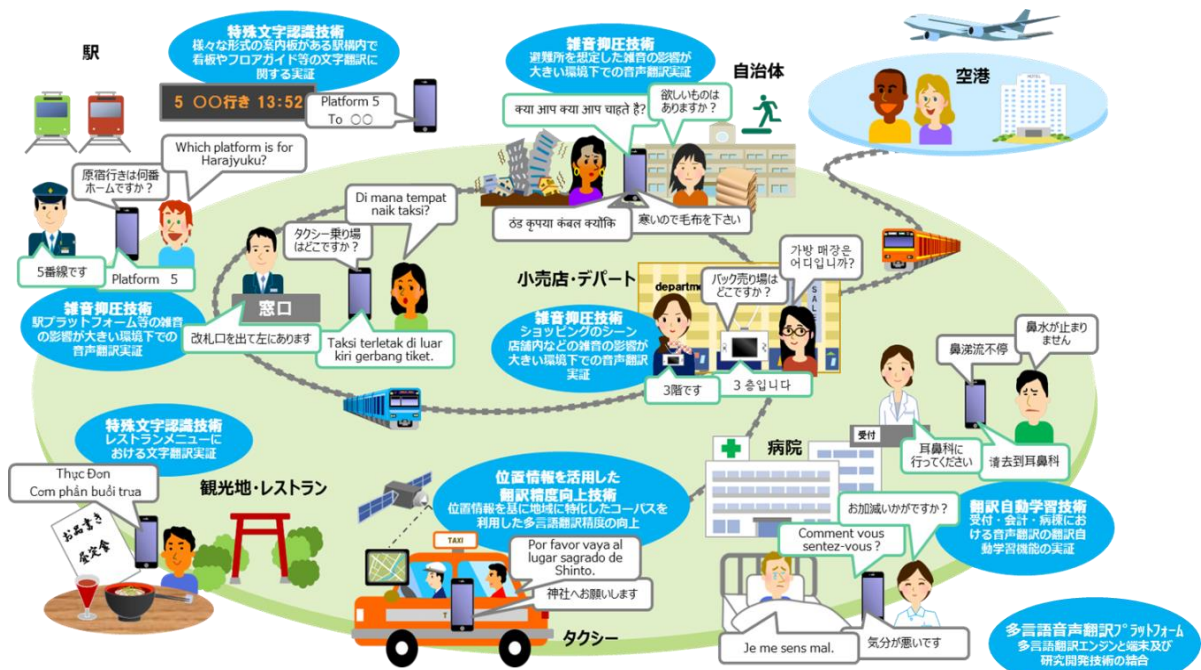
*2 VoiceTra(ボイストラ)

話しかけると外国語に翻訳する音声翻訳アプリ。見やすい画面で簡単に操作でき、翻訳結果が正しいかどうかも確認できる。翻訳できる言語は 31 言語(中国語、ポルトガル語の方言を含む)。うち、21 言語が音声で入力可能、16 言語が音声出力可能。ダウンロード、利用は無料。

<http://voicetra.nict.go.jp/index.html>

*3 グローバルコミュニケーション計画

http://www.soumu.go.jp/main_content/000285578.pdf



*4 特許等の書き言葉のための翻訳エンジンのニューラルネットワーク化

NICT は、国立研究開発法人産業技術総合研究所と情報通信分野における連携・協力の推進に関する協定を締結し、人工知能に関する研究等を中心とした情報通信分野における研究協力を行う。また、本協定に基づき、共同研究「特許文献専用のニューラル機械翻訳とそれを可能とするシステム構築に関する研究」を開始している。

2017年5月26日

国立研究開発法人産業技術総合研究所と情報通信分野における連携・協力の推進に関する協定を締結

<http://www.nict.go.jp/info/topics/2017/05/170526-1.html>

今回の成果の詳細

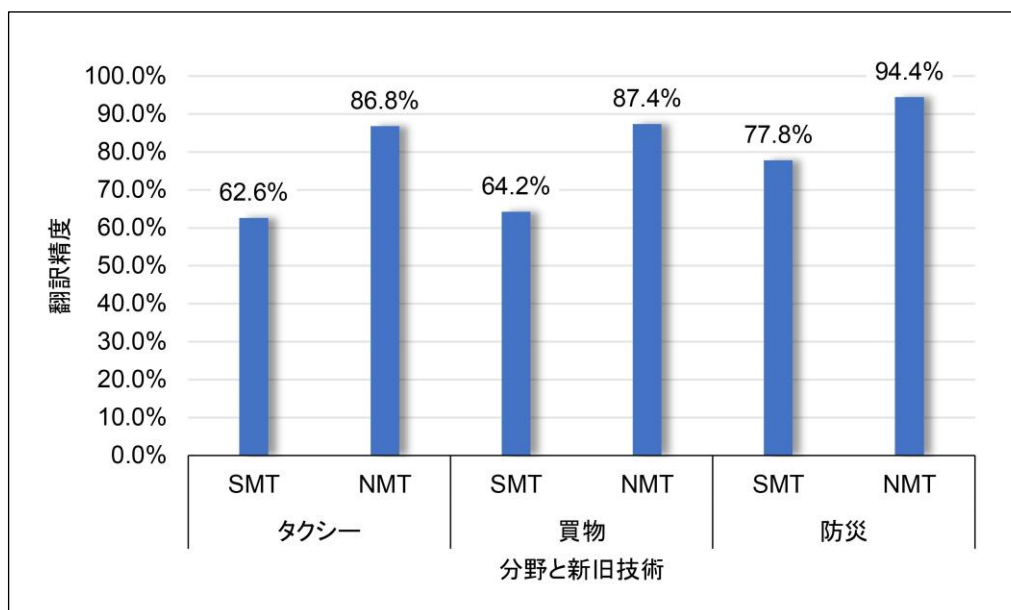
ニューラル機械翻訳(NMT)は、従来技術より、平均的に精度が高いです。しかしながら、NMT に特徴的な誤訳のパターンがあり、その様な点の解決が今後の課題となっています。例えば、「ハム」と「ハムサンド」のような関連語の混同(例 9)、ある表現の繰り返し(例 10)、「テレビ」のような訳語の抜け落ち(例 12)等があります。

<改善の具体例>

例 1	原文	I'll wait until the terrace is available.
	SMT	テラスがあるまで待ちます。
	NMT	テラスが空くまで待ちます。
例 2	原文	Hunger is the best sauce.
	SMT	空腹であれば。
	NMT	空腹は最高のタレです。
例 3	原文	Yes, you can't take a reserved taxi or a pickup taxi.
	SMT	はい貸切タクシーの送迎タクシーに乗ることはできませんか。
	NMT	はい貸切タクシーや送迎タクシーはご利用いただけません。
例 4	原文	There are more and more shops which accept electronic money even though credit cards are not acceptable.
	SMT	電子マネーカードは使えないものでも受け付けている店も増えています。
	NMT	クレジットカードが使えなくても電子マネーを使える店は増えてきています。
例 5	原文	How can I find out which drugs I cannot use?
	SMT	どうやって調べる薬が使えません。
	NMT	使えない薬はどうやって調べればいいですか。
例 6	原文	Let's search a car accessory store by a car navigation system.
	SMT	カー用品店のカーナビで検索してみましょう。
	NMT	カーナビでカー用品店を探してみましょう。
例 7	原文	Of course we plan what to do on the assumption to get damaged.
	SMT	もちろん被害想定をどうするかを計画しています。
	NMT	もちろん被害を受けることを想定して計画しています。
例 8	原文	This area is the area where inundation has been assumed in the Great Kanto Earthquake type.
	SMT	この地域は浸水想定地域は関東大震災のタイプです。
	NMT	この地域は関東大震災型で浸水が想定されている地域です。

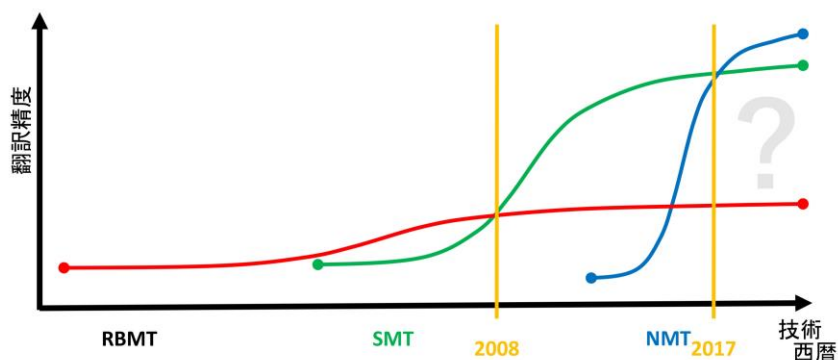
<改悪の具体例>

例 9	原文	Sandwich with ham and cheese, please.
	SMT	ハムとチーズのサンドウィッチをお願いします。
	NMT	ハムサンドとチーズのサンドイッチをください。
例 10	原文	Then, how about using an airport pickup service by a hired car?
	SMT	それではハイヤーの空港送迎サービスを利用してはいかがでしょうかですか。
	NMT	ではハイヤーのハイヤーをご利用されてはいかがでしょうかでしょうか。
例 11	原文	This bed of down is very light and warm.
	SMT	この羽毛布団はとても軽くて暖かいですよ。
	NMT	このダウンのベッドはとても軽くて暖かいですね。
例 12	原文	My house's TV doesn't have such equipment.
	SMT	私の家のテレビはそのような装備はありません。
	NMT	私の家にはそういった設備はありません。



多分野(タクシー、買物、防災等)で高い翻訳精度を実現

RBMTからSMT/NMTへ



自動翻訳の研究開発は、ルールベース翻訳(RBMT)から始まり、本資料で説明した統計翻訳(SMT)やニューラル機械翻訳(NMT)に世代交代して、高精度化してきました。第一世代のRBMTが文法規則と辞書に基づく手法で精度改善の壁にぶつかっていたところ、第二世代の対訳コーパスに基づく統計翻訳(SMT)が大きな精度の差をもたらし、取って代わりました。第二世代と同様に、対訳コーパスを使う第三世代のニューラル機械翻訳(NMT)が出現したところですが、前記の欠点が克服された場合、完全にSMTもNMTに取って代わられます。

< 本件に関する問い合わせ先 >
 先進的音声翻訳研究開発推進センター
 隅田 英一郎
 Tel: 0774-98-6350
 E-mail: ltg-info@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