

情報通信研究機構

NICT
先端研究

155

ニューラル機械翻訳
(ニューラルMT)
は、ディープラーニン
グ(深層学習)テクノ
ロジーの進化に伴い、
近年、翻訳品質が向上
している。
ただし、SNSに見
られるテキストなど、
ユーザーが生成したテ
キストを翻訳すると、

ニューラルMTは依然として大抵の場合に適切な翻訳を出力できず、重要な情報を世界規模で迅速に共有することが難しい。

ニューラルMTは通常、プロの翻訳者によって作成された非常に大量の翻訳データを用いて学習する。この学習データには、ニュー

スや議会議での討論など

の形式的なテキストが、学習して利用する独自のアプローチを提案した。私のアプローチは、あらかじめ学習された多言語の言語モデルによって、大量の翻訳されたユーザー生成テキストを学習し、その特徴を学習し、それを通常のニューラルMTの学習用の形式的なテキストに適用する。

変換されたデータを

ニューラルMTは、表の例のように、ユーザーが生成したテキストの翻訳方法を学習できるように

ニューラルMTはこれらの特徴をうまく翻訳する

（火曜日に掲載）

ニューラル
機械翻訳 ユーザー生成文法学習

先進的音声翻訳研究開発推進センター・
先進的翻訳技術研究室 テニユアトラック研究員

マリ・バンジャマン

フランスのParis Sacalay大学にてLIMSIIIC
NRS研究所から博士号を取得。その後、2016年にNICT
のASTRECE研究所に着任。以来、単言語データと合成データを
より有効活用して機械翻訳の品質を向上させる研究に取り組む。



翻訳の例

(a)	翻訳するテキスト	#COVID19 in NYC, plz dont go out and stay @home!!!
(b)	標準ニューラルMTの翻訳	#NYCのCOVID19、PLZは外出せず@home!
(c)	私の研究による翻訳	#COVID19 NYC, 外出しない、家にいてください!!!

形式的なテキスト

変換

ユーザー生成テキスト

形式的なデータをユーザー生成テキストのデータに変換して、ユーザー生成テキスト用のニューラルMTを学習