

科学技術・大学

ニユーラル機械翻訳
は、ニューラルネット
ワークを用いて翻訳を
実現する仕組みであ
る。そのモデルは、特
定の言語対のデータを
使用して訓練される。

ユニアーサルコミュニケーション研究所
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
有期研究員
ダブレ・ラジ

18年京大院博士後期課程修了。同年NICTに着任。
現在に至る。以来、低資源言語対の機械翻訳、多言語機
械翻訳の品質向上に関する研究に従事。博士(情報学)。

複数の言語対のデータを使用する
多言語多段階学習(NICT提供)

載
載
(火曜日に掲
載)

情報通信研究機構

NICT 先端研究

(201)

しかし、高品質な翻訳を実現するには数百万台の翻訳の実例データが必要であり、翻訳の実例が小規模(例えば数千文)しか存在しない言語間の翻訳は非常に難しい。

この低資源言語対の問題に対しても、ニユーラル機械翻訳のアプローチがあ

は困難であることが知られていたが、我々はモードルを訓練する。こ

の言語対に専用のモードルを取得する。この言語対に専用のモードルを段階的に訓練することで、英語からアジア語など、難易度を軽減

によって、より高い品質を継続する。これは多言語間の翻訳が可能になる。能の改善幅のほとんど

我々は図のように英語→中国語のデータとモデル洗練によるもので、難易度を軽減できた。これらは、少量のデータのみで困難な工程を管理するための方法である。

新たに追加した英語と新しいN個の言語対(例えば英語→ベンガル語や英語→クメール語など)の少量のデータを導入し、(N+1)のような段階的な訓練によって、より高い品質を向上させた。性能のデータのみで困難な問題に対処するよりも、複数の言語を使用しながら、より小規模で管理しやすい工程を通じて対処する方が良いことを示している。

ニユーラル 機械翻訳

多段階訓練で品質向上

1. 事前訓練 英語→中国語

2. 多言語対の洗練

- 英語→中国語
- 英語→日本語
- 英語→クメール語
- 英語→マレー語
- 英語→インドネシア語
- 英語→ベンガル語
- 英語→タガログ語
- 英語→ベトナム語

3. 言語対ごとの洗練 英語→クメール語

