

TSUBAME 共同利用 平成28年度 学術利用 成果報告書

日中・中日機械翻訳実用化プロジェクト
Developing Chinese-to-Japanese and Japanese-to-Chinese Practical Machine Translation

中澤 敏明
Toshiaki Nakazawa

科学技術振興機構
Japan Science and Technology Agency
<http://jst.go.jp>

邦文抄録(300 字程度)

本プロジェクトは科学技術文献に対する実用的な日中・中日機械翻訳システムを構築し、言語障壁を取り除くことで、日中間の科学技術交流・発展を促進することを目的としている。基礎解析技術の向上や言語資源の構築はもちろんのこと、機械翻訳エンジンの性能向上も重要な課題である。そこで TSUBAME を使ってニューラル機械翻訳を様々なパラメータで訓練し、最適なパラメータ設定の検討を行う。

英文抄録(100 words 程度)

This project aims at developing a practical Japanese-Chinese and Chinese-Japanese machine translation system for scientific and technological documents, removing the language barriers and enhancing the science and technology exchange between Japan and China. The key issues are improving the sentence analysis technology, constructing language resources and improving the quality of a machine translation. We try to find the best parameter set of neural machine translation system using TSUBAME by training the system with various settings.

Keywords: machine translation, neural network, Japanese-Chinese

背景と目的

これまでの機械翻訳研究においては、大規模な対訳コーパスから統計的に翻訳知識を学習する統計的機械翻訳が主流であったが、近年の深層学習の発展が機械翻訳研究にも大きな影響をもたらしており、ニューラルネットワークを利用した機械翻訳研究が盛んになっている。現状ではニューラルネットワーク翻訳モデルの精度はあまり安定しておらず、翻訳言語対や対訳コーパスの量などによって、その精度に大きな違いが見られる。本研究課題では特に日中・中日の翻訳において、高精度なニューラルネットワーク翻訳モデルの構築を目的とする。

概要

本プロジェクトは、平成 18～22 年度に実施された科学技術振興調整費「日中・中日言語処理の開発研究」プロジェクトの成果をもとに、日中の担当大臣の合意に基づき、日中両国で予算を負担し、共同の技術開発により実用化を目指す JST の事業「日中・中日機械翻訳

実用化プロジェクト」として、平成 25～29 年度の 5 年間で実施されている。

科学技術振興調整費「日中・中日言語処理技術の研究」(平成 18～22 年度)の成果(日中対訳コーパスや翻訳エンジン)を母体として、それらの成果を最大限活用して翻訳エンジンの性能向上、対訳コーパスの増強、日本語・中国語の形態素解析システムの向上などの研究を進め、科学技術文献に対する実用的な日中・中日機械翻訳システムを構築し、言語障壁を取り除くことで、日中間の科学技術交流・発展を促進する。

近年のニューラルネットワーク技術の進歩により、機械翻訳においてもニューラルネットワークを利用した手法の精度が著しく向上している。そこで我々も独自のニューラル機械翻訳エンジンを開発し、TSUBAME を利用して様々なパラメータ設定で実験を行い、最適な設定を検討する。

結果および考察

TSUBAME を用いて 300 万ペアからなる大規模な

中日対訳コーパスの形態素解析、構文解析などのデータの下準備を並列で行い、さらに GPU を利用してニューラル機械翻訳モデルのトレーニングを行なった。様々なパラメータでの実験を行い、最適な設定の検討を行った。構築したモデルを用いて機械翻訳の国際評価ワークショップに参加し、科学技術翻訳タスクにおいてトップの成績を収めた。

まとめ、今後の課題

本プロジェクトではニューラル機械翻訳エンジンを開発し、大規模中日対訳コーパスを利用することで高精度な翻訳システムを構築することができた。今後は構築したモデルを実サービスに組み込み、活用することを考える。