

## 論文内容の要旨

博士論文題目 Bilingual Dictionary Extraction via Multilingual Topic Models  
(多言語トピックモデルを用いた対訳辞書の抽出)

氏名 劉曉東

(論文内容の要旨)

機械翻訳、情報抽出などの自然言語処理のタスクでは、対訳辞書は極めて重要な役割を果たしている。本研究では、トピックモデル及びIBMモデルなどの単語アラインメント手法をを組み合わせ、対照コーパスから対訳辞書を抽出するフレームワークを提案する。

多言語トピックモデルを利用することにより、まずドキュメント整列の対照コーパスをトピック整列の対訳コーパスに転換する。このトピック整列のコーパスは、機械翻訳によく使われる対訳コーパスに構造的に類似するため、単語整列モデルを利用し、対訳辞書を抽出することが可能となる。

このフレームワークの次の2つの利点を持つ。

- (1) プロセスのブートストラップとして、シード辞書が不必要である。
- (2) 二つの言語以上の多言語対照コーパスも同様に利用できる。

ウィキペディアのデータセットを使っての大規模実験では、提案手法が先行研究より高精度の辞書を抽出できることが実証された。さらに、より多くの言語をデータセットに追加すれば、本手法の性能はより一層高くなることを示した。

氏名	劉曉東
----	-----

(論文審査結果の要旨)

平成27年1月23日に開催した公聴会の結果を参考に平成27年2月18日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

劉曉東は、本博士論文において、多言語の対照コーパスを用いて対訳辞書の自動構築を行う手法を提案した。本論文の貢献は次のようにまとめることができる。

1. 必ずしも文対応が付かないが、同様の内容を記述した二言語の対照コーパスを対象に、多言語トピックモデルを利用して、言語間で共通のトピックモデルを学習し、二言語の文書を共通のトピックを持つ語の集合として疑似的な並列コーパスとし、対訳単語対を抽出する手法を提案した。
2. 対照コーパスから対訳を抽出する他手法に比較して、高い精度で対訳単語対を抽出できること、また、トピックモデルを利用する副作用として単語の多義性を考慮した対訳抽出となっていることを示した。
3. 二つ以上の言語の対照コーパスを用いることにより、対訳単語対をより高い精度で抽出可能であることを示した。

多言語対照コーパスから対訳辞書を高い精度で抽出する手法を提案した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士（工学）の学位論文として価値あるものと認める。