

論文 / 著書情報
Article / Book Information

| | |
|-------------------|---|
| 題目(和文) | 統計的機械翻訳における翻訳モデルの学習手法に関する研究(要約) |
| Title(English) | |
| 著者(和文) | 上垣外英剛 |
| Author(English) | Hidetaka Kamigaito |
| 出典(和文) | 学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10542号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:高村 大也,新田 克己,中本 高道,奥村 学,長谷川 晶一 |
| Citation(English) | Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10542号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,, |
| 学位種別(和文) | 博士論文 |
| Category(English) | Doctoral Thesis |
| 種別(和文) | 要約 |
| Type(English) | Outline |

論文要約

本論文では、統計的機械翻訳において、翻訳モデルを学習する際の問題を解決するための手法を提案する。対象とする翻訳モデルの学習の際の問題点は次の二つである。

一つ目は既存の単語アライメントの学習手法では、言語学的に遠い言語対における機能語と内容語の対応が誤り易いという問題である。この問題を解決するために、本論文では事後確率正則化学習法の制約として、頻度に基づいて機能語と内容語を識別し、明示的に扱う手法を提案した。評価の結果、提案した手法は、従来手法である対称化制約と比較し、日本語と英語の言語対において単語アライメントおよび翻訳の精度が向上する事を確認した。

二つ目の問題は、翻訳時に使用するフレーズテーブルやルールテーブルが、単語アライメントからのフレーズ対やルールを網羅的に抽出することで生成されるために、翻訳時には不要な対応まで含んでしまい、サイズが増大するという問題である。

本論文ではこの問題に対処するために、階層型Pitman-Yor過程に基づき、階層的なバックオフを行うモデルを提案した。提案手法と従来のベイズ的手法を比較するためにドイツ語/フランス語/スペイン語/日本語から英語へ翻訳する実験を実施した結果、提案手法は同等のルールテーブルサイズの中で、BLEUにおいて同等かより高い翻訳精度を示した。また、提案手法と従来のヒューリスティックな手法との比較評価についてもドイツ語から英語への翻訳を行う実験を実施し、提案手法は10分の1以下のルールテーブルサイズで、BLEUにおいてより高い翻訳精度を示すことを確認した。さらに提案手法とヒューリスティックな手法に統計的な検定に基づくルールテーブル削減手法を適用したものの比較評価についてもドイツ語から英語への翻訳を対象に実施し、その結果より少ないルールテーブルサイズで、同等の翻訳精度を示すことについても確認した。

上記のように本論文では統計的機械翻訳における翻訳モデルに存在する二つの問題に対処するための手法を提案し、それらの手法が実際に効果的である事を実験によって示している。