

特許翻訳における長文の翻訳精度の改善

学術院情報学領域 情報科学系列 准教授 綱川 隆司

Keyword: 特許翻訳、請求項、箇条書き、分割統治法

まとめ

- 特許出願における「特許請求の範囲(請求項)」は専門用語や複雑な係り受けを含む長文が多く、大規模言語モデル(LLM)でも翻訳ミスが生じやすい。
- 本研究では、請求項をLLMで「単体で文として成立する箇条書き」に書き直してから翻訳し、翻訳後に1文へ再結合する日英翻訳手法を検討している。
- 化学分野の特許約400件で評価した結果、BLEU・COMETスコアはベースライン(Google翻訳)と比べやや低下した。意味内容や特許特有の文体は維持されているが、ベースラインに含まれる翻訳ミスは修正できなかった。

提案手法

- LLMを用いて原言語文を箇条書きに分割する(指示は右の表を参照)
 - 単体で文として成立させる(主語などは反復して書くことで、省略されたものが異なる訳になるといったことを防ぐ)
- 分割後の各項目をLLMで翻訳する
- 翻訳結果を再び1文にまとめる
- 「事例あり」の場合、箇条書きへの分割例や翻訳例を指示とともに複数与える

実験設定

- 実験データ: Google Patents Public Data上の化学に関する特許約400件から最初の2つの請求項の日本語、英語ペアを取得
- 使用したLLM: GPT-4o
- 比較対象: Google Cloud Translation APIによる翻訳結果

実験結果

- 評価指標BLEU、COMETでは、提案手法により評価値が有意に低下
- 人手評価では、提案手法により意味が大きく損なわれていないことが確認された
- Google翻訳の出力にある翻訳ミスは提案手法でも改善できていなかった

請求項の翻訳ミスの例

原文	翻訳結果
アルカリ金属およびアルカリ土類金属の少なくとも一方	at least one of alkali metal and at least one of an alkali earth metal (少なくとも1つのアルカリ金属および少なくとも1つのアルカリ土類金属)

提案手法による請求項の箇条書きへの変換例

原文(改行、空行は適宜追加したもの)	提案手法による変換結果
前記流路遮断バルブは、内部に前記閉鎖部材を収容し、カンチレバー形状からなる少なくとも1つの片持ちばりを備え、 円筒形に形成されて、 前記連通流路の流入口に挿入されるように設置されるボディー; ……	- 前記流路遮断バルブは、内部に前記閉鎖部材を収容する。 - 前記流路遮断バルブは、カンチレバー形状からなる少なくとも1つの片持ちばりを備える。 - 前記流路遮断バルブは、円筒形に形成されている。 - 前記流路遮断バルブは、前記連通流路の流入口に挿入されるように設置されるボディーである。……

LLMへの指示内容(プロンプト)

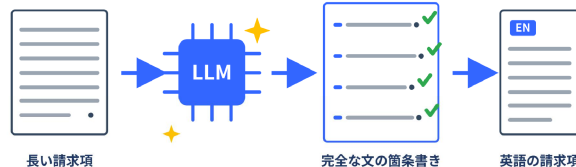
箇条書きへの分割指示	各箇条書きの翻訳および結合の指示
与えられた特許請求の範囲(claim)のテキストを、翻訳の難度を下げ精度向上を図るために文章を箇条書き形式に分割してください。原則として、一つの文には一つの情報のみが含まれるようにしてください。分割の際には係り受け関係を維持し、対象の文がもつ意味が損なわれることがなく、翻訳時に訳抜けが発生しないように注意してください。出力形式は各行の先頭を「-」(半角ハイフンとスペース)とした箇条書きとし、claimごとに記載してください。	与えられた分割済みの特許請求の範囲(claim)のテキストを英語に翻訳してください。また、翻訳後に翻訳結果を一文にまとめ、英語で書かれた特許請求項として提出できる形式にまとめてください。出力結果には一文にまとめた請求項のみを記載してください。また、claimごとに[claim1]のように分けて出力してください。

実験結果(翻訳性能評価指標の値)

手法	BLEU [%] ↑	COMET [%] ↑
提案手法 (事例なし)	23.82	75.85
提案手法 (事例あり)	23.61	75.78
Google翻訳	26.91	76.15

アピールポイント

- 原文を節で区切るだけの先行研究と異なり、各構成要素を独立した完全な文へ書き換えてから翻訳する点に新規性がある。
- 翻訳する前に箇条書きの各項目の正確さを人手でチェックすることが可能。
- 人手評価により、分割・併合しても意味内容と特許請求項としての文体が保たれることは確認できた。



応用先

- 海外特許出願時の請求項の下訳作成を支援し、知財部門の翻訳コスト低減と期間短縮を図る。
- 契約書・仕様書・マニュアルなど、長文で係り受けが複雑な技術/法律文書の機械翻訳の前処理として用いる。
- 他社特許の調査や侵害チェックの際、外国語特許の要点を箇条書き化して読解を支援するツールへ展開する。

