

システム情報工学研究科修士論文概要

年 度	平成 26 年度	学位名	修士(工学)
専 攻	知能機能システム	専攻	著者氏名 龍 梓
指導教員氏名 宇津呂 武仁			
論文題目 日中対訳特許文を用いた同義対訳専門用語の同定			
論文概要 <p>特許文書の翻訳は、他国への特許申請や特許文書の言語横断検索などといったサービスにおいて不可欠である。機械翻訳においても、また、翻訳者による翻訳においても、大規模で正確な対訳辞書は、高い品質を保証するための必要な情報源である。しかし、人手によって対訳辞書を作成し、継続的に収録語数を増やし辞書を維持・管理していく作業は膨大な時間と労力を要する。そこで、自然言語処理分野においては、多様なテキストデータを情報源として、対訳辞書を自動もしくは半自動的に作成する技術に関する研究が行われてきた。</p> <p>例えば、日中パテントファミリーから抽出した 360 万件の対訳特許文を情報源として、句に基づく統計的機械翻訳モデルのフレーズテーブルを用いることにより、日中対訳専門用語を獲得する手法が提案されている。しかし、上述の手法では、ある日本語専門用語の訳語を推定する際に、その日本語専門用語が出現する一つの対訳文に出現する訳語のみを推定対象としていた。従って、他の対訳文に出現している同義の専門用語対訳対とは全く独立に訳語推定が行われており、本来同義関係にある複数の専門用語対訳対の関係の同定できない、という問題があった。そこで、本論文では、ある日本語専門用語およびその同義語候補が出現する複数の対訳文を入力として、同義の専門用語対訳対を同定する手法を提案する。</p> <p>実際の手順においては、まず、ある日本語専門用語を種として、同義関係にある専門用語対訳対の候補を生成・収集する。次に、生成した候補集合に対して SVM 分類器を適用することにより、同義・異義集合を同定する。そして、日中パテントファミリーから抽出した 360 万対の日中対訳対に対して提案手法を適用し、約 90%の適合率を達成した。さらに、比較対象として、日英同義対訳専門用語の同定を対象とした先行研究における素性と同等の素性のもとで日中同義対訳専門用語の同定を行った評価結果との比較を行い、本論文で提案する素性の組み合わせによって大幅に性能が改善されることを示した。</p>			
審査日	平成 27 年 1 月 28 日		
審査員	(大学名 職名)	(学位)	(氏名)
主査	筑波大学 教授	博士(工学)	宇津呂 武仁
副査	筑波大学 教授	工学博士	白川 友紀
副査	筑波大学 教授	工学博士	丸山 勉