

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 董 麗娟
指導教員氏名	宇津呂 武仁		
論文題目	日中対訳特許文からの対訳専門用語の収集		
論文概要	<p>近年、中国国内における特許出願は大幅な伸びを見せている。ここで、中国特許文書の翻訳は、特許文書の言語横断検索等のサービスにおいて不可欠であるため、中国語の特許を日本語に翻訳する仕事が必要になっている。機械翻訳や人手による翻訳を行う場合、高い質を保つためには大規模で正確な対訳辞書が必要である。ここで、各国では、年々新しい技術開発が行われ、新しい専門用語が作られ、特許が申請されている。しかし、人手によって、対訳辞書を作成するためには、膨大な時間と労力を要するため、自動もしくは半自動的に日中専門用語対訳辞書を構築する手法が必要である。</p> <p>これまでに、日英対訳特許文を情報源として、専門用語対訳対を自動獲得する手法の研究が行われてきた。例えば、NTCIR-7 特許翻訳タスクにおいて配布された日英 180 万件の対訳特許文を用いて、対訳特許文から専門用語対訳対を獲得する手法が提案された。この手法では、句に基づく統計的機械翻訳モデルを用いて、対訳特許文から学習されたフレーズテーブルを用いることによって、専門用語対訳対獲得を行った。</p> <p>そこで、本論文では、上述の手法を適用することによって、日中パテントファミリーから抽出した 360 万件の日中対訳特許文を言語資源として、句に基づく統計的機械翻訳モデルにより学習されるフレーズテーブルを用いて、対訳専門用語を獲得する手法を提案する。具体的には、まず、専門用語対訳辞書獲得の情報源として用いる日中対訳文対に対して、句に基づく統計的機械翻訳モデルを適用することにより、フレーズテーブルを学習する。次に、このフレーズテーブルを用いて日本語専門用語の中国語訳語推定を行う。最後に、獲得した日中対訳専門用語に対して、複数の対訳文から得られる情報を素性として、Support Vector Machines (SVMs) を適用する。結果として、提案手法により、適合率最大の場合、90%程度の適合率を達成した。</p>		
審査日	平成 27 年 1 月 29 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	博士(工学)	中内 靖
副査	筑波大学 准教授	博士(情報科学)、博士(デザイン学)	星野 准一