

平成22年度図書館情報メディア研究科プロジェクト研究 研究成果報告書

種 目	萌芽研究	500 (千円)	研究代表者 氏 名	関 洋 平
研究課題	CGM サイトから意見情報と体験情報を抽出するための ドメイン辞書の構築と情報抽出の自動化			
研究組織 (研究代表者及び研究分担者)				
氏 名	所属研究機関・部 局・職	現在の専門	役割分担	
関洋平	大学院図書館情 報メディア研究 科・助教	情報アクセス技術 自然言語処理	研究の取りまとめ, システムの実装	
研究目的				
<p>本研究を通じて、ドメインに応じて異なる意見情報・体験情報の傾向の知識を体系化することにより、情報アクセス技術、特に意見抽出と呼ばれる一連の応用に向けた言語学的な知見をまとめる。具体的には、英語のアプレイザル理論に基づく辞書から日本語の意見分析のための辞書を構築したのち、有用性を評価する。また、Yahoo!ブログを対象とした大規模コーパスを利用して、意見情報・体験情報をアノテートし、ドメインごとの傾向の違いを明らかにする。</p>				
研究成果				
<p>本研究では、アプレイザル理論に基づく英語の辞書から、日本語のアプレイザル辞書を構築し、日英の意見抽出における効果について検証を行った。アプレイザル理論に基づく辞書は、言語学の詳細な分析に基づいて構成されるため、語彙項目のカバー率は限定されたものになる。本研究では、汎用ソーラスに基づき同義語を展開することで辞書のカバー率を向上させると同時に、SentiWordNet のスコアに基づきフィルタリングを行うことで、辞書の精度を向上させることを試みた。NTCIR-8 多言語意見分析タスクのテストコレクションを用いた実験から、この辞書は意見分析を行う上で、高い再現率を実現できる十分なカバレッジを得ていることを確認した。さらに、Yahoo!ブログを対象として、意見情報と意見対象タイプを付与したアノテーションを分析した結果、頻出する態度評価と意見対象タイプの組み合わせについて、コンピュータ、エンターテイメント、政治などドメインに特徴的な意見を把握できることを確認した。</p>				
代表的な研究発表・特許等の成果一覧、特記事項等				
<p>A. 国際会議等発表論文</p> <ol style="list-style-type: none"> <li>1. Tetsuya Sakai, Daisuke Ishikawa, <u>Noriko Kando</u>, <u>Yohei Seki</u>, Kazuko Kuriyama, and Chin-Yew Lin: Using Graded-Relevance Metrics for Evaluating Community QA Answer Selection, In Proceedings of The Fourth ACM Conference on Web Search and Data Mining (WSDM 2011), Kowloon, Hong Kong, February 2011, pp.187-196.</li> <li>2. <u>Yohei Seki</u>, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, and <u>Noriko Kando</u>: Overview of Multilingual Opinion Analysis Task at NTCIR-8 - A Step Toward Cross-Lingual Opinion Analysis, Proc. of the Eighth NTCIR Workshop, NII, Japan, June 2010, pp.209-220.</li> </ol> <p>B. 国内シンポジウム (査読付)</p> <ol style="list-style-type: none"> <li>1. <u>関洋平</u>: 言語を横断したアプレイザル辞書の構築と多言語意見分析タスクにおける評価. 情報アクセスシンポジウム2010, 2010年9月.</li> </ol>				