

論文審査の結果の要旨

報告番号	甲 先 第 280 号	氏 名	Chao Li
審査委員	主査 北 研二 副査 獅々堀 正幹 副査 任 福継		
学位論文題目 <p style="text-align: center;">Text Classification Based on Background Knowledge (背景知識に基づく文書分類)</p>			
審査結果の要旨 <p>膨大な数の非構造化テキストがWeb ページ, 電子図書館, およびコミュニティに表れている。このような非構造化データマイニングについて, 現在人手による手作業に頼っているが, その分類や検索について自動化された方法が期待されている。</p> <p>本論文では, Baidu Baike の記事と未知のコーパスからの文字共起を背景知識として, 背景知識に基づくテキスト分類方法を提案している。その動機としては, 第一に, テキストからのより多くの情報がより合理的な分類を行うことができること, 第二に, 人々は基本概念と一般知識を持っているが, 従来のコーパスはいくつかの基本的な概念と一般知識が欠けている特殊なケースだと考えられる。</p> <p>最初に, Baidu Baike を使用することで, Baidu Baike のテキストと記事の類似点を持つ新しいテキスト表現手法を提案した。</p> <p>次に, 文字の共起を研究するために, テキストとラベル付けされていないコーパスとの間の文字共起の類似性を有するテキスト表現手法を開発した。</p> <p>さらに, 様々なデータを収集し, 提案した手法によって実験を行なった。その結果は, バックグラウンド知識を使用することでパフォーマンスが向上することが確認された。</p> <p>最後に, 提案された手法に基づいて, 実験システムを構築し, 様々な評価実験を行った。実験結果によって, 本論文で提案された手法の有効性を確かめることができた。</p> <p>以上本研究は, 当該分野の既存の問題を解決した貢献から価値のある研究であり, 本論文は学位論文としての水準を満たし, 博士(工学)の学位授与に値するものと判定する。</p>			