

平成23年度科学研究費助成事業（科学研究費補助金）実績報告書（研究実績報告書）

1. 機関番号 14603 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 基盤研究(C) 4. 研究期間 平成21年度～平成23年度
5. 課題番号 21500141
6. 研究課題名 リンク解析に基づく自然言語処理・テキストマイニング技術の開発

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
90311589	シンボ 新保 マサシ 仁	情報科学研究科	准教授

8. 研究分担者(所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。)

研究者番号	研究分担者名	所属研究機関名・部局名	職名

9. 研究実績の概要

下欄には、当該年度に実施した研究の成果について、その具体的内容、意義、重要性等を、交付申請書に記載した「研究の目的」、「研究実施計画」に照らし、600字～800字で、できるだけ分かりやすく記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

本年度は、高次元データのハブ節点に着目し研究を行った。ハブとは、データ集合中で、他の数多くのオブジェクトと類似したオブジェクトのことである。自然言語データをはじめとする高次元データでは、本質的にハブが存在する可能性が高いこと（「次元の呪い」と総称される現象の一つである）が報告されている（Radovanovic et al. 2010）。グラフに基づく手法もその影響を大きく受けるため、これに関する更なる調査を行った。まず、一般のグラフに基づく半教師あり法（Zhu et al. および Zhou et al. の各種ラベル伝搬法など）は、データがグラフとして表現されていることを仮定しており、その際にはk近傍グラフを用いることがほとんどである。われわれは、k近傍グラフがハブを多く出現させることを指摘し、一方、相互k近傍グラフと呼ばれるグラフが相対的にハブを軽減することを実験によって示した。さらに、Kleinbergによる HITS と呼ばれるグラフ節点の重要度算出（ランキング）手法と、自然言語処理分野で用いられる知識獲得法の一つである Espresso ブートストラッピング法との類似性（これは前年度・前前年度の本研究の一環として指摘し・発表した）に基づき、HITS 上位オブジェクトをトピックドリフト（ブートストラッピングをはじめとする self-training 法で、本来獲得したいクラス以外のオブジェクトばかりが獲得されてしまう問題）の予防に積極的に活用する方法を提案し、その有効性を実験によって示した。なお、HITS 法で上位にランキングされる事例は、ハブの一種とみなすことができ、従来その負の側面が指摘されてきたが、本研究では、ハブを、一種の「負例」として活用した。

10. キーワード

- (1) リンク解析 (2) テキストマイニング (3) 自然言語処理 (4) ハブ
- (5) (6) (7) (8)

11. 現在までの達成度

下欄には、交付申請書に記載した「研究の目的」の達成度について、以下の区分により自己点検による評価を行い、その理由を簡潔に記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

<区分>①当初の計画以上に進展している。②おおむね順調に進展している。③やや遅れている。④遅れている。

(区分)	
(理由)	

12. 今後の研究の推進方策

本研究課題の今後の推進方策について簡潔に記述すること。研究計画の変更あるいは研究を遂行する上での問題点があれば、その対応策なども記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

13. 研究発表（平成23年度の研究成果）

※ 「13. 研究発表」欄及び「14. 研究成果による産業財産権の出願・取得状況」欄において記入欄が不足する場合には、適宜記入欄を挿入し、それによりページ数が増加した場合は、左端を糊付けすること。

【雑誌論文】 計（2）件 うち査読付論文 計（2）件

著者名	論文標題				
A. Mantrach / N. van Zeebroeck / P. Francq / M. Shimbo / H. Bersini / M. Saerens	Semi-supervised classification and betweenness computation on large, sparse, directed graphs.				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	
Pattern Recognition	有	44	2 0 1 1	1212-1224	
掲載論文のDOI（デジタルオブジェクト識別子）					
なし					

著者名	論文標題				
S. Garcia-Diez / F. Fouss / M. Shimbo / M. Saerens	A sum-over-paths extension of edit distances accounting for all sequence alignments				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	
Pattern Recognition	有	44	2 0 1 1	1172-1182	
掲載論文のDOI（デジタルオブジェクト識別子）					
なし					

著者名	論文標題				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	
掲載論文のDOI（デジタルオブジェクト識別子）					

【学会発表】計(2)件 うち招待講演 計(0)件

発表者名	発表標 題	
K. Ozaki / M. Shimbo / M. Komachi / Y. Matsumoto	Using the mutual k-nearest neighbor graphs for semi-supervised classification of natural language data	
学会等名	発表年月日	発表場所
15th CoNLL	2011年6月23日	Portland, OR, USA

発表者名	発表標 題	
T. Kiso / M. Shimbo / M. Komachi / Y. Matsumoto	HITS-based seed selection and stop-list construction for bootstrapping	
学会等名	発表年月日	発表場所
49th ACL: Short Papers	2011年6月21日	Portland, OR, USA

【図 書】 計(0)件

著者名	出版 社		
	書 名	発行年	総ページ数

14. 研究成果による産業財産権の出願・取得状況

【出 願】 計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

【取 得】 計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別
				出願年月日	

15. 備考

※ 研究者又は所属研究機関が作成した研究内容又は研究成果に関するwebページがある場合は、URLを記載すること。

--