

平成22年度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 1 4 6 0 3 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)(一般) 4. 研究期間 平成21年度～平成23年度
5. 課題番号 2 1 5 0 0 1 4 1
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
9 0 3 1 1 5 8 9	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。

本年度の主要な成果は以下の通り。

(1) 前年度開発した系列アラインメントの重み学習法に文法制約を組み合わせる並列構造の推定法を発展させ、その有効性を検証するとともに誤り傾向の分析を行った。文法的な制約(並列句どうしはオーバーラップしない, 等)を文法規則によって強制したうえで、並列句の範囲同定には並列句間の単語列の編集距離を素性として用い、パーセプトロン学習によって素性重みを最適化する。

(2) リンク解析手法を非グラフデータに適用する際には、一旦データをグラフに変換しなければならない。そのための手法としては一般に、k-近傍グラフ(データ点各々について、最も類似するk個のデータ点を辺で結ぶ)が用いられる。我々はこのグラフ変換法を自然言語データ(語義曖昧性解消用のベンチマークデータ)に適用し、評価・分析を行った。その結果、少数のハブと呼ばれる数多くの節点と接続された節点が生成され、後続のリンク解析に基づく半教師あり学習の精度に悪影響を与えることがわかった。その原因はk-近傍関係が非対称であることにより、これを解消するための簡単な手法を考案した。(3) 自然言語処理でリソース拡張のためにしばしば用いられる、ブートストラップ(自己トーレニング)法において、初期シード選択およびストップリスト構築にリンク解析を用いる手法を考案し、予備調査によって有効性を確認した。

(2), (3)については、平成23年度に、より詳細な評価を多くの自然言語データを用いて分析し、公表することを予定している。

10. キーワード

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

(裏面に続く)

11. 研究発表（平成22年度の研究成果）

〔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, L. Yen, J. Callut, K. Francoisse, M. Shimbo, M. Saerens	The sum-over-paths covariance kernel: a novel covariance measure between nodes of a directed graph			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
IEEE Transactions on Pattern Analysis and Machine Intelligence	有	32	2010	1112-1126

著者名	論文標題			
原一夫, 新保仁, 松本裕治	文法制約と系列アラインメントによる並列構造の解析			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
人工知能学会論文誌	有	25	2010	560-569

〔学会発表〕 計（1）件 うち招待講演 計（0）件

発表者名	発表標題	
Silvia Garcia Diez, François Fouss, Masashi Shimbo, Marco Saerens	Normalized sum-over-paths edit distances	
学会等名	発表年月日	発表場所
International Conference on Pattern Recognition	2010.8.24	Istanbul, Turkey

〔図書〕 計（0）件

著者名	出版社		
書名	発行年	総ページ数	

12. 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

13. 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。

--