

様 式 F-7-1

科学研究費助成事業（学術研究助成基金助成金）実施状況報告書（研究実施状況報告書）（平成23年度）

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

若手研究(B)

4. 補助事業期間

平成23年度～平成24年度

5. 課題番号

2

3

7

0

0

1

7

7

6. 研究課題

自然言語処理における全体最適化のための大規模・超並列処理

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
60581329	コマチ マモル	情報科学研究科	助教
	小町 守		

8. 研究分担者

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

本年度は研究目的の(1)に示した MapReduce による超大規模多クラス意味カテゴリ獲得の研究に従事した。具体的にはヤフー株式会社の協力を得て、Yahoo! オークションの検索クエリログおよび検索クリックスルーログを使用することができたため、Google N グラムを用いるという当初の計画を変更し、オークション検索クリックスルーログを用いた属性獲得の研究を行った。競合するカテゴリのパターンやインスタンスを用いた知識獲得の理論的分析は達成できなかったものの、オーバーラップするカテゴリを用いた属性抽出の実験を行ない、抽出されたインスタンスに関して考察を行った。

研究目的(1)の後半部分に関して、多クラス知識獲得の問題点の検討のため、オークション検索クエリログにおけるカテゴリによるクエリの性質の違いについて定性的・定量的に分析し、カテゴリによって性質が大きく異なることを調査した。また、ウェブ検索クエリとオークション検索クエリの性質の違いについて考察した。オークション検索クエリはナビゲーションクエリが多いウェブ検索とは異なり、直接購入につながる検索クエリが多いということと、商品のカテゴリを選択してから検索するクエリが存在するため、多クラスである特徴を活かした分野適応手法が有効であることを示唆している。

研究成果はH23年度後半に言語処理学会第18回年次大会にて発表し、SIGIR などの情報検索系の国際会議に投稿するため、原稿の執筆を行った。

研究目的(2)に関して、本年度は単語・句クラスタリングに関する調査を行った。日本語ウェブテキストに対して Brown Clustering を実行し、品詞の制約を用いることで精度よくクラス N-gram を作成し、得られたクラスを用いて単語 N-gram モデルの改善に取り組んだ。

10. キーワード

(1) 自然言語処理	(2) 知識獲得	(3) 情報抽出	(4) クエリマイニング
(5) オークション	(6) 単語クラスタリング	(7)	(8)

11. 現在までの達成度

(区分) (3) やや遅れている。 (理由) 大規模多クラス知識獲得タスクに関しては、対象とするデータおよび従事する RA が異なったため、理論的背景の分析が中断しているものの、おおむね順調に進展している。 しかし、単語・句クラスタリング手法を依存構造解析に適用するタスクに関しては、担当できる RA が不足したため、本年度中にサーベイと実装を完了して手法の改善を検討する、という当初の計画から、やや遅れが生じており、サーベイが完了するのみとなっている。

12. 今後の研究の推進方策 等

(今後の推進方策) 大規模多クラス知識獲得タスクに関しては、研究計画書の通り進行する予定である。 やや遅れが生じている依存構造解析に関しては、昨年度末に研究代表者の所属する研究室に新たに着任した Kevin Duh 氏と協力することで、並列処理の高速化と依存構造解析器の多言語化に着手したいと考えている。特に意味役割付与や共参照解析などの上位レイヤー、あるいは形態素解析やチャンキングなどの下位レイヤーとの結合学習に関して、並列分散処理による高速化や大規模化に取り組みたい。
(次年度の研究費の使用計画) 未使用額が生じた要因は、研究の進捗状況に合わせ、予算執行計画を変更したことに伴うものである（本年度はリサーチ・アシスタントの謝金ではなく、旅費に使用した）。 次年度は研究計画書の通り、本年度の進展が遅れている依存構造解析に関してリサーチ・アシスタントを依頼し、研究の速度を早めたい。

13.研究発表(平成23年度の研究成果)

[雑誌論文] 計(0)件 うち査読付論文 計(0)件

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(3)件 うち招待講演 計(0)件

発 表 者 名	発 表 標 題	
Tetsuo Kiso, Masashi Shimbo, Mamoru Komachi and Yuji Matsumoto	HTS-based Seed Selection and Stop List Construction for Bootstrapping	
学 会 等 名	発表年月日	発 表 場 所
Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL HLT 2011): Short Papers, pp.30-36	June 21, 2011	Portland, USA

発 表 者 名	発 表 標 題	
Kei Uchiumi, Mamoru Komachi, Keigo Machinaga, Toshiyuki Maezawa, Toshinori Satou and Yoshinori Kobayashi	Japanese Abbreviation Expansion with Query and Clickthrough Logs	
学 会 等 名	発表年月日	発 表 場 所
Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), pp.410-419	November 10, 2011	Chiang Mai, Thailand

発 表 者 名	発 表 標 題	
水本智也, 坂口慶祐, 小町守, 内海慶, 河野洋志, 前澤敏之, 佐藤敏紀	オークション検索クリックスルーログからの属性値抽出	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第18回年次大会論文集, pp.1023-1026	March 16, 2012	広島県広島市, 日本

〔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

--