

様 式 C - 7 - 1

平成 28 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

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

研究活動スタート支援

4. 研究 期 間

平成 28 年度～平成 29 年度

5. 課 題 番 号

1	6	H	0	6	9	8	1
---	---	---	---	---	---	---	---

6. 研究課題名

単語のベクトル表現に基づく分野の変化に頑健な構文解析器に関する研究

7. 研究代表者

研究 者 番 号	研究 代 表 者 名	所 属 部 局 名	職 名
0 0 7 8 2 5 4 1	ノウジ ヒロシ 能地 宏	情報科学研究科	助教

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

本年度は様々な観点で分野に依存しない頑健な構文解析器を構築するための研究を進め、以下の2つの成果があった。

1つ目は、文法として依存構造文法を仮定した場合の適切な表現に関する研究である。近年 Universal Dependencies (UD) という様々な言語にわたる依存構造のアノテーションを統一化しようとする試みがなされており、これが将来標準になっていくものと思われる。しかしながら、UD で仮定されている文法は多言語間の統一性を測る点が重視された結果、構文解析の点からは解析しにくいものになっているという意見も挙がっていた。本研究はまずこの点を調査するため、UD の構文木を別の、従来の構文解析の研究で標準的に用いられる形に似たものに変換する手法を開発し、変換後の木の方が解析が容易であることを確認した。更に、変換した木を逆に変換元の UD の木に戻す可逆変換の手法を開発した。二つの手法を組み合わせ、実際の解析は変換した解析しやすい木で行い、解析後に UD の形に戻す研究を行ったところ、多言語にわたり精度が向上し、良好な結果を示した。

2つ目は組み合わせ範疇文法 (CCG) に基づく品詞情報を用いない構文解析手法の研究である。近年単語の分散表現から、CCG の各単語に割り当てられるカテゴリを予測することで、品詞を用いなくとも高精度の解析が行えることが示されつつあったが、まだ問題も残る状態であった。本研究では、この既存手法を拡張し、単語のカテゴリと別に各単語の主辞を予測することで、同じく品詞を用いずに精度を向上させることに成功した。現在この手法は英語及び日本語の CCG 構文解析における最高精度を達成している。

10. キーワード

(1) 計算言語学	(2) 自然言語処理	(3) 構文解析	(4)
(5)	(6)	(7)	(8)

11. 現在までの進捗状況

(区分)(2) おおむね順調に進展している。

(理由)

特に CCG を用いた品詞を用いない高精度な解析法は申請時では想定していなかった成果であるが、これがうまくいったことで、CCG に基づいて、もともとの目標である分野に頑健な構文解析器を構築できる可能性が出てきた。当初予定していた、既存の依存構造解析器の各分野に対するエラー分析はまだまとまった成果が得られていないが、今後は構築した CCG 解析器の異なる分野における性能を評価し、これをもとに研究を進めていけると考えている。

12. 今後の研究の推進方策

(今後の推進方策)

本研究は以下の二つの研究に取り組む予定である。

まず、CCG に基づいて高精度が解析が行えることが明らかになったが、一つの問題点として、CCG は応用面では必ずしも必要とされず、より単純な依存構造の方が扱いやすい場合が多い点が挙げられる。従って、もし CCG の導出木を UD などの依存構造木に変換する手法が得られれば、この問題点を解決できるため、これに取り組む予定である。二つ目は、構築した CCG 構文解析器を用いた分野適応の研究である。構築した手法は品詞を用いないという点でエラー伝搬の可能性が低く、また非常に単純で拡張性が高いため、高い分野適応性も持っていると考えられる。

13. 研究発表（平成 28 年度の研究成果）

〔雑誌論文〕 計（0）件／うち査読付論文 計（0）件／うち国際共著論文 計（0）件／うちオープンアクセス 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	国際共著	
掲載論文のDOI（デジタルオブジェクト識別子）						
オープンアクセス						

〔学会発表〕 計（2）件／うち招待講演 計（0）件／うち国際学会 計（0）件

発 表 者 名	発 表 標 題	
小比田涼介、能地宏、松本裕治	依存構造解析のための内容語と機能語の多言語可逆変換	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第23回年次大会	2017年03月13日～ 2017年03月16日	筑波大学（茨城県つくば市）

発 表 者 名	発 表 標 題	
吉川将司、能地宏、松本裕治	係り受け構造との同時予測による A* CCG 解析	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第23回年次大会	2017年03月13日～ 2017年03月16日	筑波大学（茨城県つくば市）

〔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. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計（ 0 ）件

国際研究集会名	開催年月日	開催場所

16. 本研究に関連して実施した国際共同研究の実施状況

（ 1 ）国際共同研究： -

17. 備考

依存構造変換ソフトウェアの公開ページ
<https://github.com/kohilin/MultiBFConv>