

## 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	若手研究(A)																										
4. 研究期間	平成28年度～平成30年度																										
5. 課題番号	1 6 H 0 5 8 7 3																										
6. 研究課題名	ニューラルネットに基づく制御可能な自然言語生成																										
7. 研究代表者	<table border="1"> <tr> <th>研究者番号</th> <th>研究代表者名</th> <th>所属部局名</th> <th>職名</th> </tr> <tr> <td>7 0 6 3 3 4 2 8</td> <td>ニュービッグ グラム Neubig Graham</td> <td>情報科学研究科</td> <td>客員准教授</td> </tr> </table>			研究者番号	研究代表者名	所属部局名	職名	7 0 6 3 3 4 2 8	ニュービッグ グラム Neubig Graham	情報科学研究科	客員准教授																
研究者番号	研究代表者名	所属部局名	職名																								
7 0 6 3 3 4 2 8	ニュービッグ グラム Neubig Graham	情報科学研究科	客員准教授																								
8. 研究分担者	<table border="1"> <tr> <th>研究者番号</th> <th>研究分担者名</th> <th>所属研究機関名・部局名</th> <th>職名</th> </tr> <tr><td> </td><td> </td><td> </td><td> </td></tr> </table>			研究者番号	研究分担者名	所属研究機関名・部局名	職名																				
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	<p>本年度は、機械翻訳と系列変換モデルの制御性を向上させる上で、主に3つの研究成果があった。</p> <p>まず、本研究に必要な計算基盤として、ニューラルネットの柔軟な計算を可能とするニューラルネットツールキットDyNetの大幅な拡張と効率化を行った。他の既存のツールキットと比較し、計算効率や実装のしやすさに関して、その優位性を示した。また、本研究で実現された改良にもとづき、ユーザベースが大幅に増えて、現在様々な研究機関で利用されている。</p> <p>また、ニューラルネットに外部から情報を与える様々な手法を提案した。例えば、ニューラルネットを用いた言語モデルに頻度に基づく言語モデルの情報を取り入れて、低頻度語に対する頑健性の向上を図った。また、ニューラルネットに基づく翻訳では、外部辞書の情報を活用して、学習データに十分な学習事例が存在しない単語に対する翻訳性能の向上も実現できた。</p> <p>また、ニューラルネットに基づく言語生成モデルには大量な学習時間を必要とするため、効率的な実験を行うための工夫も行った。具体的には、どの単語を出力するかを確定する場合、語彙中のすべての単語を考慮せずに、単語をバイナリのビット列として扱い、それぞれのビットを推定することで計算・メモリともに大幅な削減を行うことができる。</p> <p>これらに加えて、来年度以降のための基盤になる統語情報を翻訳に取り組む手法や、言語資源が少ない場合における翻訳の研究などにも取り組んだ。</p>																										

## 10. キーワード

- |            |          |            |                 |
|------------|----------|------------|-----------------|
| (1) 自然言語処理 | (2) 機械翻訳 | (3) 自然言語生成 | (4) ニューラルネットワーク |
| (5) 制御性    | (6) 機械学習 | (7)        | (8)             |

## 11. 現在までの進捗状況

(区分)(1) 当初の計画以上に進展している。

## (理由)

当初の予定では、28年度には入力の素性追加、事前確率の指定、翻訳辞書の利用等を予定していた。これらをすべて実現する上で、計算の効率向上やニューラルネットツールキットの大幅な拡張など、他の研究実績も実現している。

## 12. 今後の研究の推進方策

## (今後の推進方策)

29年度には、大きく分けて2つの課題に取り込む。

まず、翻訳における柔軟な目標関数が指定可能な学習に取り込む予定である。これを行う上で、近年提案された、最小リスク学習や強化学習に基づく手法を適応し、翻訳結果を直接制御する目標関数を提案する。例えば、文の長さを制御する場合、その文の長さがユーザが指定した目標の長さに近い場合目標関数が高くなるような設計などが考えられる。  
また、別タスクに対するモデル化を行う予定である。例えば、文書要約や対話など、様々なタスクに対する柔軟な制御を実現する枠組みに取り組む。

## 13.研究発表(平成28年度の研究成果)

〔雑誌論文〕 計(4)件 / うち査読付論文 計(4)件 / うち国際共著論文 計(1)件 / うちオープンアクセス 計(4)件

著者名	論文標題				
Graham Neubig, Chris Dyer	Generalizing and Hybridizing Count-based and Neural Language Models				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	国際共著
Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing	有	2016	2   0   1   6	1163-1172	該当する
掲載論文のDOI(デジタルオブジェクト識別子)					
なし					
オープンアクセス					
オープンアクセスとしている(また、その予定である)					

著者名	論文標題				
Philip Arthur, Graham Neubig, Satoshi Nakamura	Incorporating Discrete Translation Lexicons into Neural Machine Translation				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	国際共著
Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing	有	2016	2   0   1   6	1557-1567	-
掲載論文のDOI(デジタルオブジェクト識別子)					
なし					
オープンアクセス					
オープンアクセスとしている(また、その予定である)					

著者名	論文標題				
三浦 明波, Graham Neubig, Sakriani Sakti, 戸田 智基, 中村 哲	中間言語情報を記憶するピボット翻訳手法				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	国際共著
自然言語処理	有	23-5	2   0   1   6	499-528	-
掲載論文のDOI(デジタルオブジェクト識別子)					
<a href="http://doi.org/10.5715/jnlp.23.499">http://doi.org/10.5715/jnlp.23.499</a>					
オープンアクセス					
オープンアクセスとしている(また、その予定である)					

著者名		論文標題【掲載確定】					
		Neural Machine Translation via Binary Code Prediction					
雑誌名		査読の有無	巻	発行年	最初と最後の頁	国際共著	
Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)		有	55	2   0   1   7		印刷中	-
掲載論文のDOI(デジタルオブジェクト識別子)							
なし							
オープンアクセス							
オープンアクセスとしている(また、その予定である)							

〔学会発表〕 計(7)件 / うち招待講演 計(1)件 / うち国際学会 計(2)件

発表者名	発表標題	
森下 睦, 小田 悠介, Graham Neubig, 吉野 幸一郎, 中村 哲	様々な分野における対訳コーパスを用いた構文解析器の自己学習効果の検証	
学会等名	発表年月日	発表場所
情報処理学会 第226回自然言語処理研究会	2016年05月15日 ~ 2016年05月16日	東京工業大学(東京都大田区)

発表者名	発表標題	
三浦 明波, Graham Neubig, 中村 哲	木構造を中間表現とするピボット翻訳手法	
学会等名	発表年月日	発表場所
情報処理学会 第227回自然言語処理研究会	2016年07月29日 ~ 2016年07月30日	岡山県立大学(岡山県総社市)

発表者名	発表標題	
Graham Neubig	Structured Neural Networks for NLP: From Idea to Code	
学会等名	発表年月日	発表場所
Young Researcher's Symposium on Natural Language Processing(招待講演)(国際学会)	2016年12月10日	Osaka International Convention Center(大阪府大阪市)

発表者名	発表標題	
Graham Neubig	Lexicons and Minimum Risk Training for Neural Machine Translation: NAIST-CMU at WAT2016	
学会等名	発表年月日	発表場所
Workshop on Asian Translation (国際学会)	2016年12月12日	Osaka International Convention Center (大阪府大阪市)

発表者名	発表標題	
札場 寛之, 小田 悠介, 吉野 幸一郎, Graham Neubig, 中村 哲	目的言語側の構造を考慮した自然言語からの構文的に正しいソースコード生成	
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17日	筑波大学 (茨城県つくば市)

発表者名	発表標題	
森下 瞳, 小田 悠介, Graham Neubig, 吉野 幸一郎, 須藤 克仁, 中村 哲	ニューラル機械翻訳におけるミニバッチ構成法の違いによる影響の調査	
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17日	筑波大学 (茨城県つくば市)

発表者名	発表標題	
小田 悠介, Philip Arthur, Graham Neubig, 吉野 幸一郎, 中村 哲	二値符号予測と誤り訂正に基づくコンパクトなニューラルネットワーク翻訳モデル	
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17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

## 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

DyNet: The Dynamic Neural Network Toolkit

<https://github.com/clab/dynet/>

本研究で大幅に拡張・効率化されたニューラルネットツールキットDyNet