

# ACL2014 参加報告 (その 1)

## – 会議の概要と Best Paper の紹介 –

小田 悠介<sup>†</sup>

### 1 はじめに

本稿では、2014 年 6 月 22 日から 6 月 27 日にかけてアメリカ合衆国のボルチモア市で開催された国際会議 The 52nd annual meeting of the Association for Computational Linguistics (以下、ACL 2014 とする) について報告する。ACL 2014 は計算言語学、自然言語処理分野の国際的な学会である The Association for Computational Linguistics<sup>1</sup> が毎年開催している会議であり、この分野では実質上のトップカンファレンスと見なされている。

本稿では ACL 2014 の全体に関する概要と統計を述べた後、採択論文のうち best paper award を受賞した論文を紹介する。なお、本稿の統計情報は基本的に全て ACL Admin Wiki<sup>2</sup> に公開されている情報、および過去の参加報告 (西川 2013) に基づく。更なる詳細については ACL Admin Wiki 上のレポート等を参照されたい。

### 2 会議の日程と構成、参加者

ACL 2014 は 3 つの日程に分けられる。初日 (6 月 22 日) は最近の研究に関するチュートリアルがあり、事前募集により集まった 19 講義から 8 講義が選抜され、並列セッションとして開催された。23 日から 25 日の本会議では、ACL により採択された論文の口頭発表およびポスター発表が行われた。本会議のセッションの間には招待講演、best paper awards や学会への貢献者に対する表彰も行われた。26 日および 27 日は ACL に付随して各分野のワークショップが開催された。ワークショップの応募は全部で 48 企画あり、このうち ACL では 17 企画 (1 日開催が 15 企画、2 日開催が 2 企画) が開催された。他の企画については EACL や EMNLP といった他の会議に持ち越されることになっている。

<sup>†</sup>奈良先端科学技術大学院大学, Nara Institute of Science and Technology

<sup>1</sup><http://acl2014.org/>

<sup>2</sup><http://www.aclweb.org/adminwiki/>

表 1 ACL の論文投稿数と採択数・採択率の推移

年	投稿数			採択数			採択率 %		
	Long	Short	計	Long	Short	計	Long	Short	計
2010	638	332	970	160	63	223	25.1	19.0	23.0
2011	634	512	1146	164	128	292	25.9	25.0	25.5
2012	571	369	940	111	76	187	19.4	20.6	19.9
2013	664	624	1288	174	154	328	26.2	24.7	25.5
2014	557	532	1089	146	139	285	26.2	26.1	26.1

参加者数は当日の変動もあり正確な値は不明だが、事前登録者は 1302 人であった。各セッションの間には通行が困難になるほど通路に人が溢れ、近年の言語処理分野の活発さを印象付ける光景であると感じられた。

### 3 投稿数と採択率

最近 5 年間の投稿数と採択数・採択率を表 1 に示す。本会議へは合計 1089 件の論文が投稿された。このうち採択された論文は 285 件で、採択率は 26.1% であった。投稿数が過去最大であった昨年の ACL 2013 と比較すると若干減少しているものの、今年も比較的投稿数の多い開催であったと言える。採択率に関しては、わずかに高いがおおむね例年通りと考えることができる。

論文の種別ごとでは、long paper (9 ページ程度) は 557 件が投稿され、このうち 146 件が採択された。採択されたもののうち 95 件が口頭発表、51 件がポスター発表であった。一方、short paper (5 ページ程度) は 532 件が投稿され、139 件が採択された。このうち口頭発表は 51 件、ポスター発表は 88 件であった。ショートペーパーは近年になって投稿数が顕著に増加しており、今年の投稿数および採択数は昨年に引き続きロングペーパーに匹敵する数である。この傾向は会議の日程にも影響を与えているようで、従来は 1 日で行われているポスター発表のセッションが ACL 2014 では 2 日に分られ、5 時間強の時間を割いて行われた。General Chair によるレポートによれば、この日程には参加者同士の交流を促進させるという主催者側の意図もあったようである。

上記以外の論文では、student research workshop へ 26 件の論文投稿があり、うち 13 件が採択 (採択率 50.0%) されている。また昨年に引き続き Transactions of the Association for Computational Linguistics (TACL)<sup>3</sup> 採択論文の講演があり、こちらは口頭発表が 13 件、ポスター発表が 6 件であった。

<sup>3</sup><http://www.transacl.org/>

表 2 投稿数の多い研究分野とその割合

分野	投稿数	全体に対する割合 %
Machine Translation	148	13.59
Information Extraction and Text Mining	103	9.46
Document Categorization, Sentiment Analysis, and Topic Models	101	9.27
Semantics	90	8.26
Tagging, Chunking, Syntax and Parsing	83	7.62
NLP Applications and NLP-enabled Technology	66	6.06
Language Resources and Evaluation	59	5.42
NLP for the Web and Social Media	58	5.33
Machine Learning for Language Processing	54	4.96
Lexical Semantics and Ontology	49	4.50

表 3 採択論文のタイトルに高頻度で含まれる単語 (分野名を除く)

単語	論文数	単語	論文数
dependency	18	detection	13
relation	16	unsupervised	11
Chinese	14	neural	11
automatic	13	supervised	11

#### 4 論文の傾向

投稿された論文は 20 種類の分野に分類されている。このうち上位 10 分野を表 2 に示す。例年の傾向と同じように、機械翻訳 (machine translation)、意味解析 (semantics)、構文解析 (parsing) といった分野の人气が高い。評判分析 (sentiment analysis) や Web に関する論文も昨年に続いて多数の投稿があり、近年の研究の大きな流れのひとつとなりつつあるようである。

表 3 に示すのは、ACL 2014 の採択論文のタイトルに含まれる単語のうち、出現回数が 11 回以上で分野名に含まれないものである。このようにして見ると、論文の着眼点としては係り受けや依存構造 (dependency) が高い人気を持っていることが分かる。特定言語に関する研究では、中国語が圧倒的に多いことも見て取れる。neural の語はニューラルネットを使用した研究が多く提案されている近年の傾向を反映していると思われる。

## 5 Best paper awards

ACL 2014 では, “Fast and Robust Neural Network Joint Models for Statistical Machine Translation” (Devlin, Zbib, Huang, Lamar, Schwartz, and Makhoul 2014) が best long paper award として選出された. この論文では従来の統計的機械翻訳の枠組みにニューラルネットから得られる素性を追加することで, 従来法よりも少ない素性で同等以上の翻訳精度を実現している. ニューラルネットの言語処理への応用に関する風潮は前述したとおりだが, この論文もそのような流れで出現したものであり, その中の優れた成果の一つであると考えられる.

Long paper からは他に student best paper が 1 本, honorable mention が 3 本選出されている. また, short paper からも honorable mention が 2 本選出されている. long, short とともに honorable mention は意味解析に関する論文が多く選出されているようである.

## 6 論文紹介

ここでは前述した ACL 2014 の best long paper を紹介する.

### 6.1 Fast and Robust Neural Network Joint Models for Statistical Machine Translation

統計的機械翻訳において, 原文に対して適切な目的言語の単語を選択するのは重要な問題である. 適切な単語と一口に言ってもその適切さを測る指標には様々なものが考えられ, ある単語が特定の単語に翻訳されやすいといった事実や, 出力文の流暢さなどがこれに当たる. これら複数の指標を機械翻訳で考慮するためには, 基本的にはデコード (出力文の生成処理) や候補文の再ランキングの際に用いる情報として, 各指標に対応する素性を追加することで対応される. 前述の指標の場合, 単語単位やフレーズ単位の翻訳確率, 言語モデルなどを素性として導入することで実現される.

従来法で使用される素性において主に考慮されているのは両言語の単語やフレーズ同士の直接の対応関係であったり, 出力文の目的言語としての整合性といったものであり, 原言語と目的言語双方の文脈による相互作用は基本的には考慮されてはいない. この論文の着眼点のひとつは, このような相互作用をニューラルネットを介して直接モデル化しようとする点にある.

機械翻訳へのニューラルネットの応用としては, 言語モデルとしてリカレントニューラルネット (RNN) を用いてモデル化したもの (RNNLM) (Mikolov, Karafiát, Burget, Cernocký, and Khudanpur 2010) を用いることで再ランキング時の精度を向上させる手法がある. しかし, この手法は言語モデルの尤度を求めるために翻訳文全体を使用する必要があるため, より翻訳精度の向上を期待できるデコード時には適用できないという欠点がある. この論文ではより単純

なフィードフォワードニューラルネットによる言語モデル (Schwenk, Dchelotte, and Gauvain 2006) の枠組みを拡張し、ニューラルネットの入力として両言語の情報を使用するモデルを提案している。

この論文の提案するモデルは非常に単純だが強力であり、論文から導かれる素性のバリエーションだけで従来の様々な種類の素性を使用する翻訳器を上回る翻訳精度を実現している。またモデルの高速化のための考察も行われており、単純な実装に比べて数桁単位で高速であることなども評価された点だと考えられる。

論文では従来法との再ランキング性能のみの比較も行われている。ここで RNNLM を用いる従来法に対して、それに匹敵する再ランキング性能を提案法が実現しているのは印象的である。近年のニューラルネットを用いた手法では、非常に多い階層を持つものや RNN といった深い学習を行うものが注目されやすいように感じるが、この論文で用いられているネットワークは中間層が 1, 2 層とごく浅い代わりに、従来法よりも広範囲の情報を入力している点に特徴がある。結局のところ、問題の性質を正確に見極め、その問題に適切なモデルを提案することができた点がこの論文の最大の功績であるように思える。

## 7 おわりに

本稿では ACL 2014 の概要について述べ、best long paper を受賞した論文の紹介を行った。

ACL への日本からの一参加者として気になった点を挙げるならば、今回の開催ではスポンサーに日本企業が参加していなかった点を始め、日本の産業界からのアプローチが開催中を通してあまり見られなかった点である。言語処理分野が国際的に隆盛しているのは明らかだが、日本から遠方の開催では注目度が下がってしまうのかもしれない。次回の ACL 2015 は中華人民共和国の北京市で IJCNLP との合同開催として行われる。日本近隣での開催ということで、日本の研究者がこの分野で更に活躍することを期待するとともに、国内企業などの動向にも注目していきたい。

## 謝辞

本稿の一部は第 217 回情報処理学会自然言語処理研究会で行われた ACL 2014 参加報告に基づく。本稿作成の機会を与えて下さった編集委員の皆様、研究会運営に携わられた皆様、また ACL 2014 参加にあたって様々な助言を下された皆様にこの場で感謝を申し上げます。

## 参考文献

- Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., and Makhoul, J. (2014). “Fast and Robust Neural Network Joint Models for Statistical Machine Translation.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1370–1380 Baltimore, Maryland. Association for Computational Linguistics.
- Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., and Khudanpur, S. (2010). “Recurrent neural network based language model.” In *INTERSPEECH*, pp. 1045–1048.
- Schwenk, H., Dchelotte, D., and Gauvain, J.-L. (2006). “Continuous space language models for statistical machine translation.” In *Proceedings of the COLING/ACL on Main conference poster sessions*, pp. 723–730. Association for Computational Linguistics.
- 西川仁 (2013). “ACL2013 参加報告 (その 1).” 言語処理学会国際会議参加レポート No.3.

## 略歴

小田 悠介 :

2013 年神戸市立工業高等専門学校専攻科 (電気電子工学専攻) 卒業。同年, 奈良先端科学技術大学院大学情報科学研究科に進学, 現在に至る。統計的機械翻訳, 音声言語処理の研究に従事。電子情報通信学会学生員。

(2014 年 6 月 24 日依頼)

(2014 年 8 月 23 日受付)