EMNLP2013 参加報告（その１）
– 概要 –

河原 大輔[†]

1　はじめに

Conference on Empirical Methods in Natural Language Processing (EMNLP) は the Association for Computational Linguistics (ACL) の SIGDAT (Special Interest Group for Liguistic Data and Corpus-based Approaches to Natural Language Processing) によって毎年開催されている自然言語処理の経験的手法[1]に関する国際会議である。今年の EMNLP (EMNLP2013) は 2013 年 10 月 18 日から 21 日に米国・シアトルで開催された。10 月 18 日に 3 件のワークショップがあり、10 月 19 日から 21 日の 3 日間に本会議があった。本稿では、EMNLP2013 本会議の概要を報告する。

2　投稿論文

EMNLP2013 には 772 件[2]の論文投稿があり、これは過去の EMNLP において最多の件数であった。その内訳は、long paper が 539 件、short paper が 233 件であった。なお、short paper は、EMNLP においては今回初めて導入された。採択数は、long paper が 151 件、short paper が 56 件であり、採択率は、long paper が 28%、short paper が 24%となる。最近 5 年の EMNLP の投稿数、採択数、採択率を表 1 に示す。投稿数は徐々に増加傾向にあり、採択率は 25%程度で安定している。

投稿論文の分野は 15 個に分かれており、Machine Translation が最多の 98 件の投稿があった。その次に、Semantics が 87 件、NLP-related Machine Learning が 76 件と続いている。

また、論文査読プロセスにおいて、author response[3]の機能が強化された。これは、author

[†]京都大学, Kyoto University

[1]つまり、言語データやコーパスに基づく手法を意味する。

[2]他会議との double submission の論文で、EMNLP2013 への投稿が取り下げられたものは含まない。

[3]査読プロセスの途中に、査読者のスコア、コメントを一旦著者に返し、査読者に誤解などがあればそれを指摘する枠組み。

表 1　直近 5 年の EMNLP の投稿数、採択数、採択率

	投稿数	採択数	採択率
EMNLP2009	475	163	34.3%
EMNLP2010	500	125	25.0%
EMNLP2011	628	149	23.7%
EMNLP2012	570	139	24.4%
EMNLP2013	772	207	26.8%

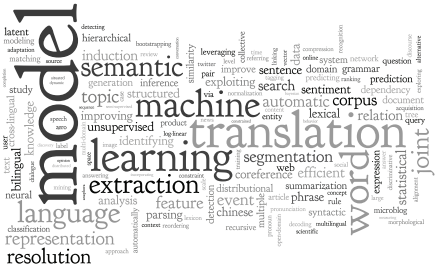

図 1　EMNLP2013 の論文タイトルの word cloud

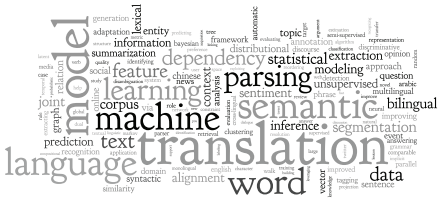

図 2　ACL2013 の論文タイトルの word cloud

response 終了後に、査読者に author response を考慮したかどうかを選択させることによって、author response を査読結果により反映させるというものであり、査読の正確性を上げるための良い試みだと思われる。

採択された論文のうち、74 件の long paper と 14 件の short paper が口頭発表となり、それ以外の論文はポスターでの発表となった。また、ACL2013 などの他の国際会議と同様に、ACL の論文誌 Transactions of the Association for Computational Linguistics (TACL) の採択論文 5 件の口頭発表も行われた。

3　採択論文の傾向

口頭発表のセッションの件数では、Machine Translation と Semantics が 3 セッションずつあり、近年の ACL、EMNLP と傾向は変わらず、機械翻訳、意味解析の論文が多い。また、NLP Applications が 3 セッションあり、エッセイや小説の自動評価のような応用研究も多かった。その他には、NLP for Social Media が 2 セッションあり、ACL2013 から引き続きソーシャルメディ

アを対象にした論文が多いという傾向がある。逆に、構文解析に関する論文は少なく、1 セッションしかなかった。これは、近年の ACL や EMNLP とは異なる傾向である。

参考までに、EMNLP2013 の論文タイトルから作成した word cloud を図 1 に、ACL2013 のものを 2 に示す。ACL2013 において目立っている構文解析 (parsing) が EMNLP2013 では小さいことなどが見てとれる。

もう少し細かく見て行くと、共参照・ゼロ照応解析や、情報抽出・関係抽出に関する論文がそれぞれ 10 件程度あった。Deep learning あるいは Word representation に関する論文も目立った。このあたりからも、研究の中心は構文から意味表現・解析にシフトしつつあることが伺える。

4 Best Paper Award

5 件の論文が best paper の候補にノミネートされ、会議最終日 (10 月 21 日) 午後のシングルセッションでそれぞれの発表が行われた。機械翻訳、機械学習、意味解析と幅広い分野から選ばれた質の高い論文であり、このセッションは非常に面白かった。

- Regularized Minimum Error Rate Training
  (Michel Galley, Chris Quirk, Colin Cherry and Kristina Toutanova)
- Of Words, Eyes and Brains: Correlating Image-based Distributional Semantic Models with Neural Representations of Concepts
  (Andrew J. Anderson, Elia Bruni, Ulisse Bordignon, Massimo Poesio and Marco Baroni)
- Easy Victories and Uphill Battles in Coreference Resolution
  (Greg Durrett and Dan Klein)
- Breaking Out of Local Optima with Count Transforms and Model Recombination: A Study in Grammar Induction
  (Valentin Spitkovsky, Hiyan Alshawi and Daniel Jurafsky)
- Cross-Lingual Discriminative Learning of Sequence Models with Posterior Regularization
  (Kuzman Ganchev and Dipanjan Das)

これらの論文から、上から 4 件目の Spitkovsky らの論文 (Spitkovsky, Alshawi, and Jurafsky 2013) が best paper に選ばれた。これは、教師なし文法導出というタスクにおいて、transform と join という二つの操作を用いて、教師なし学習手法において大きな問題となる局所解の問題に対処した論文であり興味深い。

5　おわりに

本稿では EMNLP2013 本会議の概要を紹介した。次回の EMNLP (EMNLP2014) は、2014 年 10 月 25 日から 29 日にカタール国・ドーハで行われる予定である[4]。自然言語処理分野のトップカンファレンスが中東で行われるのは初めてのことだと思われる。カタールには、カーネギーメロン大学カタール校や、Qatar Computing Research Institute (QCRI) があり、この地域でコンピュータ科学の研究が近年活発に行われている。EMNLP2013 では過去最多の論文投稿数を記録したが、EMNLP2014 でも世界中から多くの論文が投稿されると思われる。その中で、日本の研究者のさらなる活躍を期待したい。

謝辞

本稿の執筆に当たり、さまざまな情報を提供いただいたプログラム Co-chair の Tim Baldwin 氏に感謝の意を表す。

参考文献

Spitkovsky, V. I., Alshawi, H., and Jurafsky, D. (2013). "Breaking Out of Local Optima with Count Transforms and Model Recombination: A Study in Grammar Induction." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1983–1995 Seattle, Washington, USA. Association for Computational Linguistics.

略歴

河原 大輔（正会員）：1997 年京都大学工学部電気工学第二学科卒業．1999 年同大学院修士課程修了．2002 年同大学院博士課程単位取得認定退学．東京大学大学院情報理工学系研究科学術研究支援員，独立行政法人情報通信研究機構研究員，同主任研究員を経て，2010 年より京都大学大学院情報学研究科准教授．自然言語処理，知識処理の研究に従事．博士 (情報学)．言語処理学会，情報処理学会，人工知能学会，電子情報通信学会，ACL，各会員．

（2013 年 11 月 13 日依頼）
（2014 年 1 月 21 日受付）

[4] http://emnlp2014.org/