ACL2013 参加報告（その 1）
– 概要 –

西川 仁[†]

1　はじめに

本稿では，2013 年 8 月 4 日から 9 日にかけてブルガリア共和国の首都，ソフィア市で開催された国際会議 The 51st annual meeting of the Association for Computational Linguistics（以下，ACL 2013 とする) について報告する．ACL 2013 は計算言語学や自然言語処理に関する国際的な学会である The Association for Computational Linguistics[1] が主催する会議であり，これらの分野を代表する会議の一つである．本稿では特に，採択率や参加者数，論文のトピックの分布など，会議の概要を中心に報告する[2]．

2　投稿数と採択率

ACL 2013 には 1,286 件の投稿があった．うち，採択されたものは 328 件であり，採択率は約 25%である．ロングペーパーとショートペーパーを個別に見ると，前者は 662 件の投稿に対し 174 件の論文が採択され，後者は 624 件の論文に対し 154 件の論文が採択された[3]．ここ 5 年の ACL の投稿数の推移を表 1 に示す．各ます目の数字は，左からそれぞれ，合計の投稿数（採択率)，ロングペーパーの投稿数（採択率)，ショートペーパーの投稿数（採択数）である．ACL 2013 への論文の投稿数は過去最高である．表 1 を見る限り，ロングペーパーの投稿数は大きく変化していないが，ショートペーパーの投稿数が大きく増加しており，これが過去最高の投稿数の原因となっている．

参加者数は 990 人であった．国別の参加者数に関する情報は公開されていないが，筆者の印象では，本分野を代表する会議にふさわしく，世界の様々な国から広く参加者が集い，国際色豊かな会議となっていた．

今年からの新しい試みとして， ACL による新しい論文誌 Transactions of the Association for Computational Linguistics[4]（以下， TACL とする）に採択された論文に関する口頭発表が行え

[†]日本電信電話株式会社 NTT メディアインテリジェンス研究所, NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation

[1] http://www.aclweb.org/

[2] 本稿の体裁は飯田による COLING 2012 参加報告 (飯田 2013) によった．

[3] これらの情報は ACL Wiki http://www.aclweb.org/aclwiki/ および ACL Admin Wiki http://www.aclweb.org/adminwiki/ にまとめられている．

[4] http://www.transacl.org/

表 1　ここ 5 年の ACL の投稿数の推移.

	Submitted	Accepted	Rate
ACL 2009	925 (569/356)	214 (121/93)	23.1% (21.3%/26.1%)
ACL 2010	970 (638/332)	223 (160/63)	23.0% (25.1%/19.0%)
ACL 2011	1146 (634/512)	292 (164/128)	25.5% (25.9%/25.0%)
ACL 2012	940 (571/369)	187 (111/76)	19.9% (19.4%/20.6%)
ACL 2013	1288 (664/624)	328 (174/154)	25.5% (26.2%/24.7%)

るセッションが設けられた．TACL に採録された論文は ACL， NAACL， EACL， EMNLP のいずれかの会議で口頭発表を行うことができ，今回の ACL は TACL のセッションが設けられた初の会議となった．TACL からは 16 件の論文の口頭発表があり，うち 9 件がオーラルセッション，7 件がポスターセッションで発表された．

3　投稿論文の分野

会議に採択された論文の分野について表 2 にまとめる[5]．実際には投稿論文は 27 の分野に分類されており，ここでは上位 10 分野のみを示す．詳細については脚注の URI を確認されたい[6]．ACL の例年の傾向と大きな変化はなく，構文解析や意味解析，統計的機械翻訳に関する研究が大きな割合を占めている．例年と大きく異なる点は，評判分析が大きな割合を占めている点である．特にツイッターなどのマイクロブログを対象として評判分析を行う研究が多く見られ，新しい種類のテキストメディアの登場が本分野に与える影響の大きさが如実に表れている．また，言語処理と画像処理を横断するような研究が多く見られたことも特筆すべき点であろう．次に述べるが，ベストペーパーは映像から映像を説明する文を生成するものであり，今後の言語処理研究の広がりを予感させる．

4　Best Paper Awards

ACL 2013 では “Grounded Language Learning from Video Described with Sentences” (Yu and Siskind 2013) がベストペーパーとして選出された．この論文では，短いビデオクリップからそれを説明する文を生成するタスクを扱っている．訓練事例としてビデオクリップとそれを説明

[5]分類については ACL Admin Wiki 記載の Program Chairs による報告 http://aclweb.org/adminwiki/index.php?title=2013Q3_Reports:_Program_Chairs によった．

[6]2013 年のロングペーパーの投稿数が，ACL Wiki では 664 件となっているが，ACL Admin Wiki では 662 件となっている．この表は主に ACL Wiki に基づいて作成したものであるため，そちらの数字とした．

表 2　採択された論文の分野

分野	投稿数（割合）
Machine Translation: Statistical Models	31 (9.4%)
Semantics	27 (8.2%)
Sentiment Analysis, Opinion Mining and Text Classification	27 (8.2%)
Syntax and Parsing	26 (7.9%)
Machine Translation: Methods, Applications and Evaluations	25 (7.6%)
NLP Applications	25 (7.6%)
Summarization and Generation	17 (5.2%)
Statistical and Machine Learning Methods in NLP	16 (4.9%)
Text Mining and Information Extraction	16 (4.9%)
Language Resources	14 (4.2%)

する文が与えられ，それらからある種の文法を学習することで，新しいビデオクリップに対しても説明を付与することができる．詳しくは高瀬氏の「 ACL 2013 参加報告（その 2)」を参照されたい．

5　論文の紹介

ここでは，2 本ほど論文を紹介する．

5.1　Unsupervised Transcription of Historical Documents

Berg-Kirkpatrick らによる論文 (Berg-Kirkpatrick, Durrett, and Klein 2013) は，18 世紀から 19 世紀頃に印刷された文書の文字認識を扱っている．タスクとしてはいわゆる光学文字認識であるが，大きく分けて 3 つの問題が存在し，既存の手法では太刀打ちできない．1 つめの問題は字形に関する問題である．200 年前と現在では同じ文字でも字形が異なり，そのため現在の字形に基づいて訓練されていた認識器では古い文字を認識できない．2 つめの問題は文字が正しい位置に存在しないことである．当時の文書は活版印刷によって印刷されているため，文字の位置が上下にずれていることがままあり，これが認識を攪乱する．3 つめの問題はインクの染みの問題である．印刷の際に生じた不注意などにより，当時の文書にはインクが大きく染みている箇所があり，正しく文字を読み取ることが難しくなっている．

Berg-Kirkpatrick らは，これらの問題に対処するため，教師なしの生成モデルを新しく考案した．彼らは，字形，文字の位置のずれ，インクの染みなどを一つ一つ確率的な現象としてモデ

リングし，背後に存在する文字 n-gram に対してこれらの要因が影響を与えた結果として，認識の対象となる文書が生成されたという仮定を置いた．このアプローチにより，教師ありのモデルに勝る，高精度の文字認識を達成している．

本論文は，特段難しい手法を利用しているわけではない．あくまで，一つ一つの問題を丁寧にモデリングすることで，古い文書の光学文字認識という難しいタスクにおいて高精度を達成している．その丁寧なアプローチから学べるものは多い．

5.2 HEADY: News headline abstraction through event pattern clustering

Alfonseca らによる論文 (Alfonseca, Pighin, and Garrido 2013) は，ある特定の話題に関する複数の新聞記事が与えられた際に，それらから 1 つの見出しを生成する課題を扱っている．典型的には Google News[7] のようなニュース・アグリゲーターにおいて収集された 1 つの記事集合に 1 つの見出しを付与する課題となる．

著者らはこの課題に対して，事前に見出しの元となるパタンを用意しておき，新しい新聞記事集合が与えられた際にはそれらのパタンの中から適切なものを選び出し，パタン中のスロットに必要な情報を埋め込むことによって見出しを生成するという方法で対処している．まず，事前に大規模な新聞記事集合を収集し，それらから見出しとなりうるパタンを事前に抽出しておく．パタンを抽出する際には，パタンはある潜在的なイベントから生成されたという仮定を置き，ある種の生成モデルを仮定する．パタンを用意したら，新しい記事集合が与えられたとき，当該記事集合の中から顕著な情報を抽出し，それらの情報をパタンに埋め込むことによって見出しが生成される．

自動要約の分野では，長い間，与えられた文書集合の中から文書の一部を構成する表現を抽出し，それらをつなげることで要約を作成するという抽出型の要約が主流であった．一方，近年では，本論文のように，元の文書にない表現を生成する要約，生成型の要約を試みる研究が徐々に増えてきており，本論文もその潮流の一部を成すものと言える．今後はこのようなアプローチの比重が更に増すものと思われる．

6　おわりに

本稿では国際会議 ACL 2013 の概要について紹介した．次回の ACL 2014 は米国のボルチモア市で開催される[8]．昨年度の COLING 2012 においても投稿数が過去最高に達した (飯田 2013) ように本分野の隆盛は誰の目にも明らかである．引き続き日本を拠点に活躍する研究者の本分

[7] https://news.google.co.jp/
[8] http://www.cs.jhu.edu/ACL2014/

野での活発な活動を期待したい．

謝辞

本稿の一部は ALAGIN & NLP 若手の会合同シンポジウムおよび第 3 回テキストマイニングシンポジウムでの ACL 2013 参加報告に基づく．各シンポジウムの運営に携わられた方々，参加報告の際に議論をしてくださった方々，また ACL 2013 現地において様々な情報共有をしてくださった方々に記して感謝する．

参考文献

Alfonseca, E., Pighin, D., and Garrido, G. (2013). “HEADY: News headline abstraction through event pattern clustering.” In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 1243–1253.

Berg-Kirkpatrick, T., Durrett, G., and Klein, D. (2013). “Unsupervised Transcription of Historical Documents.” In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 207–217.

Yu, H. and Siskind, J. M. (2013). “Grounded Language Learning from Video Described with Sentences.” In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 53–63.

飯田龍 (2013). “COLING 2012 参加報告（その 1）.” 言語処理学会ニュースレター, **20** (1).

略歴

西川 仁（正会員）：2008 年慶應義塾大学大学院政策・メディア研究科修士課程修了．同年，日本電信電話株式会社入社．2013 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了．博士（工学）．現在 NTT メディアインテリジェンス研究所研究員．自然言語処理，特に自動要約の研究開発に従事．The Association for Computational Linguistics ，言語処理学会，人工知能学会，情報処理学会，各会員．

（2013 年 8 月 1 日依頼）
（2013 年 9 月 20 日受付）