会議レポート

COLING 2016 参加報告

COLING 2016 開催概要

International Conference on Computational Linguistics (COLING) は，国際組織「International Committee on Computational Linguistics (ICCL)」により，1965 年から 2 年に 1 度開催されている，計算言語学，自然言語処理（NLP）分野に関する査読付き難関国際会議である．今回の COLING 2016 は，通算 26 回目の会議であり，2016 年 12 月 11 日～ 12 月 16 日に大阪のグランキューブ大阪で開催された．日本での開催は，1980 年の東京，1994 年の京都に続き 3 回目である．

会期中，最初の 2 日間（11，12 日）で 18 件の併催ワークショップおよび 7 件のチュートリアルが開催され，残りの 4 日間（13 ～ 16 日）で本会議が開催された．本会議では，採択された 337 件の論文が口頭（135 件）あるいはポスター（202 件）形式で発表され，デモ展示が 65 件行われた．研究発表以外にも，12 日夜にレセプション，14 日の午後にエクスカーションとして奈良へのバスツアー，14 日の夜にバンケットといったイベントも開催され，研究者同士の懇親の場が提供された．

参加者と採択論文の傾向

今回の COLING の参加者は 1,143 人であり，前回の 800 人程度，前々回の 612 人を大きく上回り大盛況であった．近年，NLP 分野が注目を集めて拡大していることを反映しているのかもしれない．日本開催ということで，通常の国際会議と比べて日本人の割合が多く，学生ボランティアとして参加している日本の学生が散見されたのも印象的だった．会議前日（10 日）に，NLP 関連の若手研究者のための組織である「NLP 若手の会」によるシンポジウムが併設されたことも日本人学生が多く参加できた要因と考えられる．このように，本会議での発表がない学生も参加しやすくなっており，これからの NLP の研究を担う若者が国際会議を体感し，刺激を受けられる貴重な場となっていた．

表 -1 に最近の COLING の投稿数と採択率をまとめる．今回，採択率は 32% と平均的であったが，投稿数は 1,000 件を超え，採択数は近年の COLING の中で最多となった．**表 -2** に採択数の多かった上位 5 カ国をまとめる．1 位は中国，2 位はアメリカ，3 位はドイツであり，中国とアメリカが二強でその後にドイツが続くという例年通りの傾向となった．日本は，COLING 2012, 2014 ではドイツに続いて 4 番目に多かったが，今回はイギリスより少なく 5 番目となった．地元開催で参加しやすい会議であることから，多くの論文が投稿され採択数も多くなると思われたが，発表数において日本の存在感を示すことができず，個人的には残念であった．

	開催地	投稿数	採択数	採択率
COLING 2006	シドニー	630	272	43%
COLING 2008	マンチェスター	600	180	30%
COLING 2010	北京	815	334	41%
COLING 2012	ムンバイ	1,000+	333	~25%
COLING 2014	ダブリン	705	217	31%
COLING 2016	大阪	1,039	337	32%

表 -1　投稿数と採択数の推移

	COLING 2012	COLING 2014	COLING 2016
1 位	アメリカ 19%	中国 20%	中国 23%
2 位	中国 17%	アメリカ 17%	アメリカ 17%
3 位	ドイツ 11%	ドイツ 10%	ドイツ 12%
4 位	日本 8%	日本 8%	イギリス 8%
5 位	フランス 5%	カナダ 5%	日本 6%
		スペイン 5%	
		フランス 5%	

表 -2　採択数上位 5 カ国

表 -3 に，会議で設けられた分野ごとの採択および投稿論文数を示す．COLING 以外の NLP に関する難関国際会議（Annual Meeting of the Association for Computational Linguistics 等）と比べると，「Linguistic Issues in NLP」や「Computational Psycholinguistics」といった分野が設けられ，計算言語学の特徴を持つ理論よりの論文も受け入れられているのが特徴的である．採択数が最も多かった分野は「Machine Translation」であった．投稿数が最も多かった分野は「Information Retrieval, Information Extraction, Question Answering」であり，続いて「Machine Translation」であった．ただし，投稿数最多の分野は，3 つの研究分野をまとめたものである．一方，機械翻訳は世界的に盛んに研究されている研究分野であることが分かる．また，会議で設けられた分野ごとの採択率を算出すると，最も採択率が高かった分野は「Under-resourced Languages」（採択率 48%）であり，続いて，「Computational Psycholinguistics」（採択率 47%），「Speech Recognition, Text-to-Speech, Spoken Language Understanding」（採択率 46%）であった．これらの分野には，投稿数自体は少なかったが，質の高い論文が多く集まったと考えられる．

今回発表された論文全体の特徴として，ニューラルネットワークを活用した研究がとても多かった．近年，

セッション	投稿数	採択数
Machine Translation	88	31
Applications	76	28
Information Retrieval, Information Extraction, Question Answering	126	26
Resource, Software and Tools	56	25
Natural Language Generation, Summarization	62	25
Sentiment Analysis, Computational Argumentation	85	22
Morphology, Segmentation, Tagging, Chunking	72	21
Semantic Processing, Distributional Semantics, Compositionality	74	21
Linguistic Issues in NLP	50	20
Machine Learning for NLP	74	18
Under-resourced Languages	33	16
Computational Psycholinguistics	30	14
Discourse Relations, Coreference, Pragmatics	34	13
Lexical Semantics, Ontologies	47	13
Dialog Processing and Dialog Systems, Multimodal Interfaces	29	12
Syntactic and Semantic Parsing Grammar Induction	47	12
Speech Recognition, Text-to-Speech, Spoken Language Understanding	24	11
Paraphrasing, Textual Entailment	32	9

表 -3　分野ごとの投稿および採択論文数

NLP のさまざまなタスクでニューラルネットワークが成果をあげており，ここ数年，NLP 業界全体がニューラルネットワークに着目し，爆発的なブームとなっている．今回の COLING においてもそのブームが反映されていた．タイトルに「neural」を含む論文は，COLING 2012 では 1 本，COLING 2014 では 6 本であったのが，今回は 49 本へと急増していることがそれを物語っている．また，筆者の専門である機械翻訳の分野においては，半分以上の発表（31 本中 17 本）がニューラルネットワークを活用した手法に関する発表であった．

ベストペーパー

会議の最後には，アジアの NLP 組織「Asian Federation of Natural Language Processing (AFNLP)」と日本の NLP 組織「The Association for Natural Language Processing (ANLP)」が合同で 2 つの論文（Deyne ら [1] と Eskander ら [2]）をベストペーパーとして表彰した．

Deyne らの論文は，人間が判定することにより単語間の関連性データを新たに作成し，そのデータに基づいて単語間の類似度を算出する手法を提案している．そして，提案手法は，word2vec[3] 等，大量のテキストデータにおける単語分布情報から導出した単語表現に基づく類似度算出手法よりも優れていることを示している．本論文は，単語類似度を，単語間関係を明示的に示していないテキストデータから求めることが通常になっている現状に一石を投じると同時に，大規模テキストデータと小規模な心理言語学的データの補完的な活用の可能性を示唆している点が評価されたと思われる．

Eskander らの論文は，Adaptor Grammar（AG）という木構造上のノンパラメトリックベイズモデルに基づく教師なし形態素分割モデルを提案している．解析対象の言語を学習したことがない学者が短時間で Web から収集した接辞情報をモデルに導入することで性能を改善させている．また，この学者が与える知識を使った場合と同等の性能改善が，適合率重視の AG モデルで学習した文法により実現可能であることを示している．提案モデルの面白さだけではなく，人間を介した手法との比較という興味深い視点の内容を含んでいる点が評価されたと思われる．

次回の COLING

次回の COLING は，アメリカのサンタフェ（ニューメキシコ州）で 2018 年 8 月下旬に開催される予定である．前述の通り，COLING は計算言語学の特色が強い論文も受け入れられる懐が深い会議であり，NLP の国際会議の中で発表論文の多様性が高い会議であるので，興味のある読者はぜひ次回の COLING への参加を検討してほしい．筆者は，今回の COLING には会議冊子の制作担当として運営側の立場で参加したが，次回は，発表者として会議を盛り上げられるように研究に邁進する所存である．

参考文献

1）Deyne, S. D., Perfors, A. and Navarro, D. J. : Predicting Human Similarity Judgments with Distributional Models : The Value of Word Associations. In *Proceedings of COLING 2016*, pp.1861-1870 (2016).

2）Eskander, R., Rambow, O. and Yang, T. : Extending the Use of Adaptor Grammars for Unsupervised Morphological Segmentation of Unseen Languages. In *Proceedings of COLING 2016*, pp.900-910 (2016).

3）Mikolov, T., Chen, K., Corrado, G. and Dean, J. : Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781* (2013).

（田村晃裕／国立研究開発法人 情報通信研究機構）