

ACL 2017 参加報告

～会議概要およびソーシャルメディア分析技術～

三浦康秀^{†,‡}

2017/9/7

[†]富士ゼロックス株式会社

[‡]東京工業大学

Outline

- ACL 2017について
 - ACLとは
 - ACL 2017概要
 - Chairs/Presidentsからの話題
- 招待講演紹介
- 論文紹介
 - ニューラルモデルの流行
 - ソーシャルメディア分析

ACLとは

- ACL

- The annual meeting of the Association for Computational Linguistics

- 計算言語学(Computational Linguistics), 自然言語処理(Natural Language Processing)のトップカンファレンス

- 開催地

- 北米→アジア/オセアニア→ヨーロッパで巡回

- Beijing (2015) → Berlin (2016) → Vancouver (2017) → Melbourne (2018) → Florence (2019)

ACL 2017概要 (1)

- 期間

- 7/30~8/4

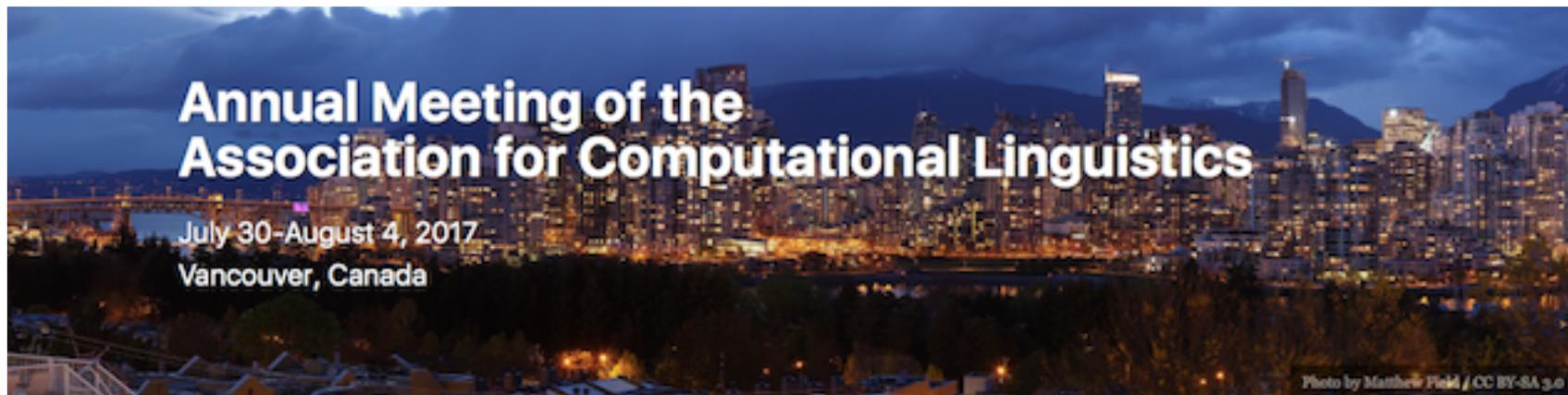
- 7/30 チュートリアル+ワークショップ

- 7/31~8/2 本会議

- 学生ワークショップ

- 8/3~8/4 ワークショップ

© 2017 ACL
<http://acl2017.org/>



ACL 2017概要 (2)

- 開催場所

- Vancouver
 - Westin Bayshore Hotel

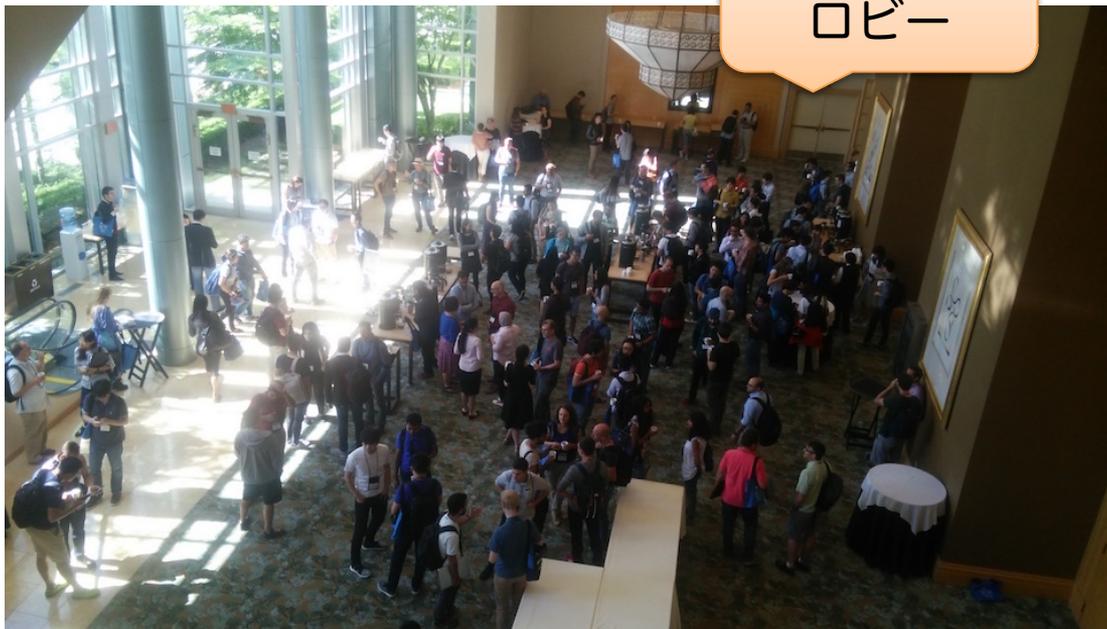
- 発表件数

- 本会議論文 302件
 - long: 195, short: 107
 - 採択率 22.9%
- TACL論文 21件
- ソフトウェアデモ 21件
- 各種ワークショップの論文も多数あり



会場の雰囲気

休息時間の
ロビー



ACL看板



Chairs/Presidentsからの話題 (1)

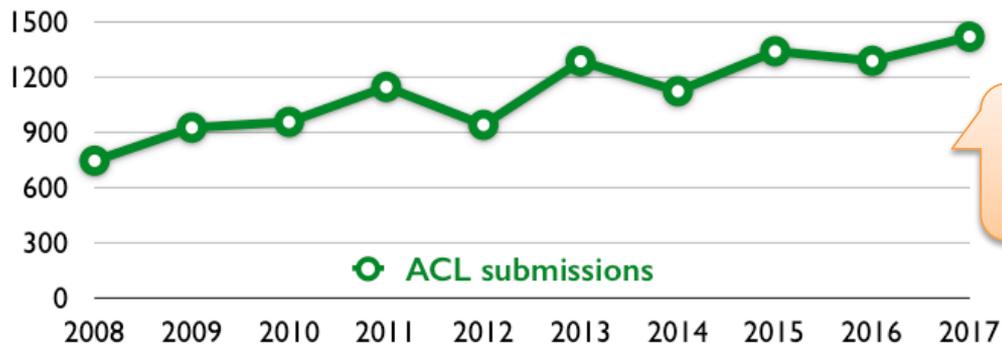
- Challenges for ACL
 - Computational Linguistic (CL) is booming!
 - Equity and Diversity
 - Publishing and Reviewing
 - Good Science

主な4つの話題のうち、3つを紹介

Chairs/Presidentsからの話題 (2)

- CL is booming!

© 2017 Joakim Nivre
<https://www.slideshare.net/aclanthology/joakim-nivre-2017-presidential-address-acl-2017-challenges-for-acl>



10年で約2倍



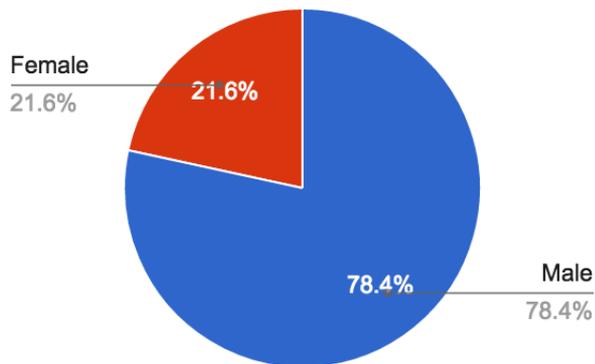
係り受け解析の
精度

Chairs/Presidentsからの話題 (3)

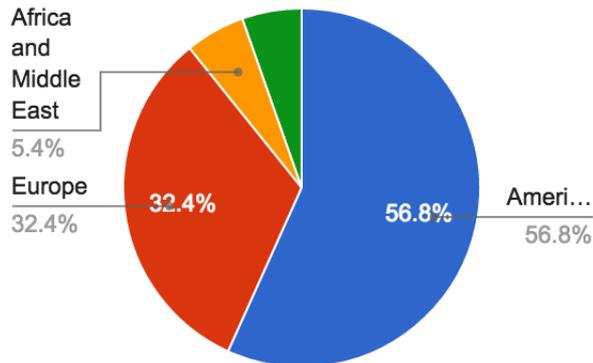
- Equity and Diversity
 - ACLにおける公平性と多様性の問題
 - 性別, 年齢, 地域
 - 選考委員・フェロー・Area Chair等での偏り

Area Chair
での例

Gender



Region



© 2017 knmnyn
<https://chairs-blog.acl2017.org/2016/11/11/last-call-for-area-chairs-a-call-for-diversity/>

Chairs/Presidentsからの話題 (4)

- WiNLP

- 第1回のワークショップをACL 2017で開催
- 今後もbig interest groupとして活動を継続

© 2017 WiNLP
<http://www.winlp.org/>



Chairs/Presidentsからの話題 (5)

- Publishing and Reviewing
 - 計算言語学/自然言語処理の研究者はカンファレンスが好き
 - 匿名による査読
 - 近年はarXivへの投稿が盛況
 - preprint(前刷り)であり，査読はなく匿名でもない

Chairs/Presidentsからの話題 (6)

- ACL 2017での事前アンケート
 - 22%がよく preprint をアップロード
 - 27%がよく preprint を引用
 - 88%が匿名の査読を支持
 - 87%は preprint が禁止されても ACL に投稿

Chairs/Presidentsからの話題 (7)

- まとめ

- 匿名の査読に対する強い支持
- 完全にpreprintを禁止することには否定的
- preprintサーバへの匿名機能の追加
- preprintを引用する際のガイドラインの設定

- 今後

- 引き続き長期的な課題として検討する

招待講演紹介

2人とも計算言語学・
自然言語処理分野

- 2件の招待講演

- Squashing Computational Linguistics

- 講演者: Noah A. Smith

- Translating from Multiple Modalities to Text and Back

- 講演者: Mirella Lapata

2件のうち
1件を紹介

Translating from Multiple Modalities to Text and Back (1)

- 講演者

- Mirella Lapata (University of Edinburgh)

- 概要

- Encoder-Decoderモデルによる複数モーダルのテキストへの変換(逆の場合も)について

Scream
(ムソクの叫び)



© 2017 Mirella Lapata
<https://www.slideshare.net/aclanthology/mirella-lapata-2017-translating-from-multiple-modalities-to-text-and-back>

Translating from Multiple Modalities to Text and Back (2)

- The Simplification Task
 - テキストの平易化
 - 例えばオバマ大統領のスピーチを分かりやすくする
 - タスク固有の制約を強化学習で学習
- Language to Code
 - 質問テキストの論理式への変換
 - Sequence-to-Treeモデルによる構造の学習
- Movie Summarization
 - 映画の作品情報の生成
 - 単語, キャラクターグラフ, シーン内のオブジェクト等の様々な情報を利用

バンクーバーでの滞在



名物のサーモンを
使ったバーガー

2010年オリン
ピックの聖火台



近隣の公園の
トーテムポール
(2009年製作)

Social Event: Vancouver Aquarium



水族館入り口

用意されていた
ACLロゴ



歓談中の様子



論文紹介

- 2つの観点での紹介
 - ニューラルモデルの流行
 - ここ数年の国際会議報告で挙がっているトピック
 - ソーシャルメディア分析
 - 私の研究テーマの1つで重点的に聴講したトピック

Neural Modelの流行 (1)

- 論文数

- 過去の調査と同じニューラル語リストを使用

- neural, lstm, recursive, rnn, recurrent, cnn, convolution, dnn, deep, embedding, distributed representation

会議	論文数 (割合)
EMNLP 2015 (増村2015)	70/312 (22.4%)
NAACL-HLT 2016 (鈴木2016)	71/182 (39.0%)
ACL 2016	97/328 (29.6%)
EMNLP 2016 (高瀬2016)	92/264 (34.8%)
ACL 2017	113/302 (37.4%)

Neural Modelの流行 (2)

- 多いのでキーワードを設定して絞ると
 - キーワード
 - 以前の報告(増村2015;鈴木2016;高瀬2016)で挙げたアプローチ
 - embedding, attention, encoder-decoder
 - 今年のチュートリアルで挙げた最近のアプローチ
 - variational, reinforce
 - 調査方法
 - 最初のページに現れる論文の数を確認
 - pdfgrepを利用

Neural Modelの流行 (3)

- 結果

- “embedding”以外は上昇傾向
 - 簡易的なものであり，より厳密な調査が望ましい

キーワード	ACL 2017	ACL 2016	NAACL 2016
neural	164	120	70
embedding	38 (23.2%)	45 (37.5%)	16 (22.9%)
attention	59 (36.0%)	35 (29.2%)	17 (24.3%)
encoder-decoder	19 (11.6%)	10 (8.3%)	4 (5.7%)
variational	4 (2.4%)	0 (0.0%)	0 (0.0%)
reinforce	6 (3.7%)	2 (1.7%)	1 (1.4%)

“neural”を含む
論文内で計算

Neural Modelの流行 (4)

- 2つのキーワードについて詳細を確認
 - Variational Auto-Encoder Approaches
 - 対話システム, Morphological Reinflection, 構文解析, 言語モデル
 - Reinforcement Learning Approaches
 - 質問応答, 対話システム, Semantic Parsing
 - いくつかピックアップして紹介

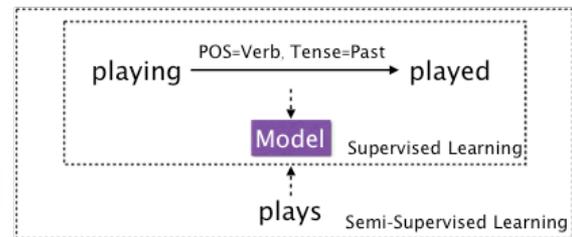
Multi-space Variational Encoder-Decoders for Semi-supervised Labeled Sequence Transduction

- 著者

- Chunting Zhou, Graham Neubig

- 概要

- ラベル付き系列変換のためのVariational Encoder-Decoderモデルの提案
 - 半教師あり学習が可能
 - ラベル付きの系列
 - ラベルなしの系列+ラベルのサンプリング
 - 変分推定による下限を最適化により学習可能
 - Morphological Reinflectionにおいて，従来手法を上回る性能を実現



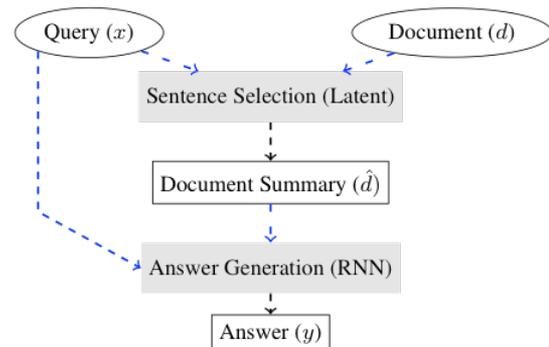
Coarse-to-Fine Question Answering for Long Documents

- 著者

- Eunsol Choi, Daniel Hewlett, Jakob Uszkoreit, Illia Polosukhin, Alexandre Lacoste, Jonathan Berant

- 概要

- 長い文書に対応できるQA手法の提案
- RNNモデルに文選択の機構を導入
 - 文選択は強化学習で行う
 - 正しい選択文はなく、報酬の高くなる文を選ぶ
- 文数が多いデータでは従来手法を上回る性能を実現
- 3.5~6.7倍の高速化を実現



Social Media分析

- 論文数

- Social Mediaエリア

- 10

- キーワードでの調査

- 35

- “social media”, “social network”, “twitter”, “tweet”のいずれかを含む

- ツイートを実験データに用いているような論文も含む

- いくつかピックアップして紹介

EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks

- 著者

- Muhammad Abdul-Mageed, Lyle Ungar

- 概要

- 24種類の感情を含む大規模コーパスを構築

- 約160万ツイート

- Distant Supervisionに基づく

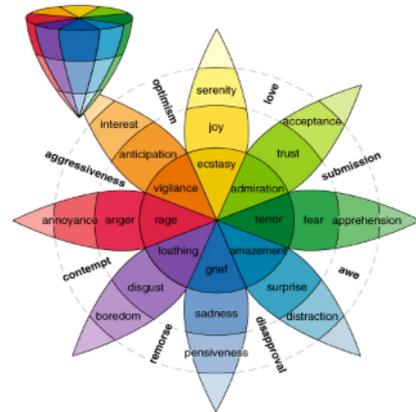
- GRNNsによりaccuracy=87.58%を達成

- GRNNsはRNNの一種

- Gated Recurrent Unit (GRU)という名称でよく知られる

- 従来のオンライン学習のモデルでは56.84~62.10%

- 8種類の基本的な感情に集約した場合はさらに高い95.68%



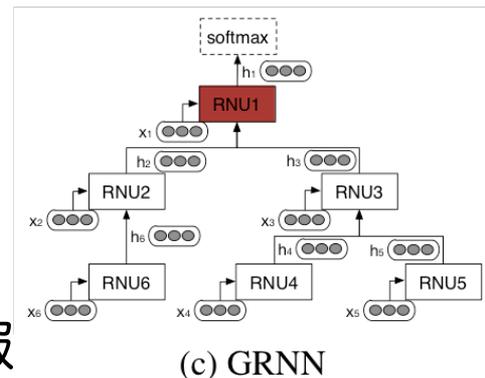
Demographic Inference on Twitter using Recursive Neural Networks

- 著者

- Sunghwan Mac Kim, Qiongkai Xu, Lizhen Qu, Stephen Wan, Cecile Paris

- 概要

- Twitterユーザのデモグラフィック属性の推定
 - 性別, 年齢, フォロー関係
- テキストに加えてソーシャルネットワークの情報を利用
- グラフ構造を対象としたRecursive Neural Networksを利用
 - 従来の語彙ベース, Bag-of-Words, ネットワークベースの手法を上回る性能を実現



Overcoming Language Variation in Sentiment Analysis with Social Attention

- 著者

- Yi Yang, Jacob Eisenstein

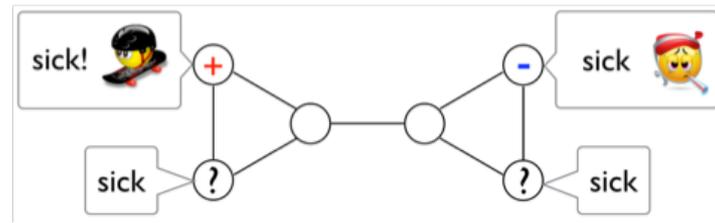
- 概要

- TACL論文

- ソーシャルネットワーク情報を用いSentiment Analysisの性能を向上

- k個のモデルをAuthor Embeddingに基づくattentionでアンサンブルする

- Twitterを対象としたSentiment Analysisで従来のモデルを有意に上回る性能を実現



まとめ

- 学会の方向性
 - 公平性と多様性
 - preprintの扱い
- 論文の傾向
 - Neural Modelの流行の継続
 - Neuralの中でのトレンドの変化はありそう
 - Social Media分析
 - 新しい感じはしないが依然として発表は多い

参考文献

鈴木潤. 2016. NAACL-HLT 2016 参加報告 ~ 会議概要と研究傾向分析 ~. 第9回テキストマイニングシンポジウム.

高瀬翔. 2016. EMNLP2016参加報告.第3回自然言語処理シンポジウム.

増村亮. 2015. EMNLP 2015 (2) ~ 深層学習の観点から ~. 第2回自然言語処理シンポジウム.

Muhammad Abdul-Mageed, Lyle Ungar. 2017. EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks. In Proc. of ACL 2017.

Eunsol Choi, Daniel Hewlett, Jakob Uszkoreit, Illia Polosukhin, Alexandre Lacoste, Jonathan Berant. 2017. Coarse-to-Fine Question Answering for Long Documents. In Proc. of ACL 2017.

Sunghwan Mac Kim, Qiongkai Xu, Lizhen Qu, Stephen Wan, Cecile Paris. 2017. Demographic Inference on Twitter using Recursive Neural Networks. In Proc. ACL 2017.

Yi Yang, Jacob Eisenstein. 2017. Overcoming Language Variation in Sentiment Analysis with Social Attention. TACL.

Chunting Zhou, Graham Neubig. 2017. Multi-space Variational Encoder-Decoders for Semi-supervised Labeled Sequence Transduction. In Proc. of ACL 2017.

Appendix

スケジュール (1)

Monday, 31 July

09:00-10:00	Plenary Session. Welcome to ACL 2017 (brief statements from general chair, program co-chairs, and presidential address)
10:00-10:30	Coffee break
10:30-12:00	Session 1 (5 parallel sessions)
12:00-1:40	Lunch Break
1:40-3:15	Session 2 (5 parallel sessions)
3:15-3:45	Coffee Break
3:45-5:20	Session 3 (5 parallel sessions)
6:20-9:30	Evening Poster Session 1 with Dinner

Oralは5並列
(昨年は5~7並列)

Poster 3時間10分
(夕食含む)

スケジュール (2)

Tuesday, 1 August

09:00-10:10	Plenary Session. Invited Talk – <i>Squashing Computational Linguistics</i> Noah Smith
10:10-10:30	Coffee break
10:30-12:05	Session 4 (5 parallel sessions)
12:05-1:30	Lunch Break (<i>Salons E/F</i>)
1:30-3:05	Session 5 (5 parallel sessions)
3:05-3:25	Coffee Break
3:25-4:45	Session 6 (5 parallel sessions)
5:45-7:40	Evening Poster Session 2
7:00-10:00	Social event at the Aquarium

Poster 1時間55分
水族館 3時間
(時間の重なりあり)

スケジュール (3)

Wednesday, 2 August

09:00-10:10	Plenary Session. Invited Talk – <i>Translating from Multiple Modalities to Text and Back</i> Mirella Lapata
10:10-10:40	Coffee break
10:40-12:30	Session 7 (Outstanding papers; 2 parallel sessions)
12:30-1:30	Lunch Break
1:00-2:30 (Note: overlap)	ACL Business Meeting (Salons D/E/F)
2:30-3:00	Coffee break
3:00-4:45	Session 8 (Outstanding papers; 2 parallel sessions)
4:45-5:45	Lifetime Achievement Award session
5:45-6:00	Closing session (with award announcements)

Outstanding Papersは2並列
(昨年と同じ)

arXivについての
議論