

国際会議報告

石井亮 (NTT), 井上昂治 (京大), 千葉祐弥 (東北大),
角森唯子 (NTTドコモ), 成松宏美 (NTT),
福田悠人 (埼玉大), 増村亮 (NTT)

本報告の趣旨

1. 対話に関する国際会議の研究動向について共有
 - この会議ではこのテーマが流行っている！
2. 本シンポジウム参加者の国際会議への投稿・参加を促進
 - コレコレに関する研究はこの会議がベスト！
 - こうすればアクセプトされる（と思う）！
 - 研究以外にも面白いことがたくさん！

報告する会議

- SIGDIAL - 千葉 祐弥 (東北大学)
- SemDial - 角森 唯子 (NTTドコモ)
- ACL - 増村 亮 (NTT)
- COLING - 増村 亮 (NTT)
- EMNLP - 増村 亮 (NTT)
- IJCAI-ECAI - 成松 宏美 (NTT)
- IVA - 石井 亮 (NTT)
- ICMI - 福田 悠人 (埼玉大学)
- Interspeech - 井上 昂治 (京都大学)

国際会議報告 SIGDIAL 2018

千葉 祐弥
(東北大学)

基本情報

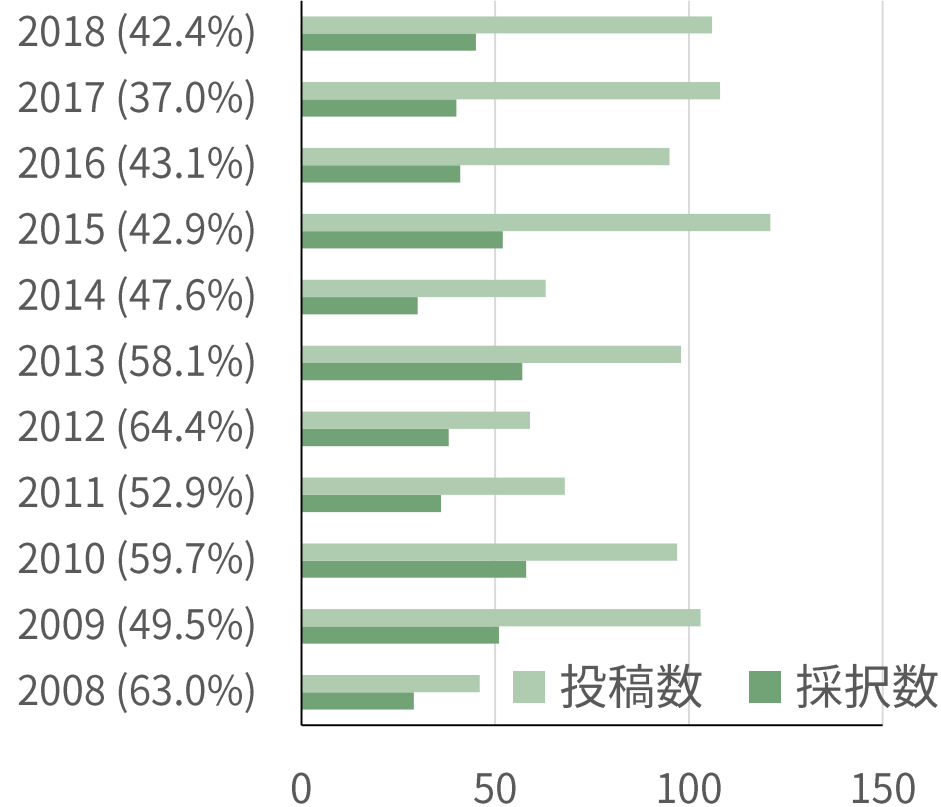
- **対話システム・対話分析に関する幅広い研究が対象**
 - 談話処理
 - 対話システム
 - 対話コーパス収集・分析
- **ACL 2018, YRRSDS 2018と併催**
 - YRRSDS: 対話システムの若手研究者が集い議論する会議
- **発表件数 49件 (Long: 2, Short: 1件のwithdraw)**
 - Long: 34, Short: 11, Demo: 4
- **全体の採択率 47%**
 - Long: 54% (36/67), Short: 31% (12/39), Demo: 80% (4/5)

採択率から見る傾向

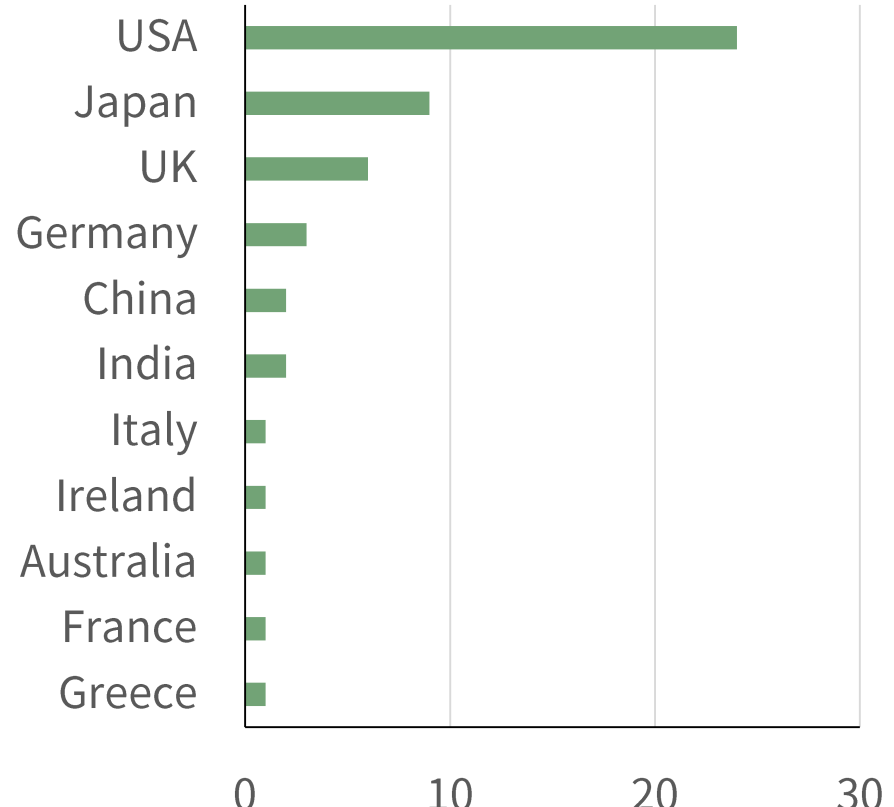
- 近年は大体40%前半くらいの採択率
- 2018年は日本が米国に次いで二位、プレゼンスは比較的高い
 - 日本からの発表では非タスク指向が多めな印象？

採択率の推移

※デモペーパーを除いて計量



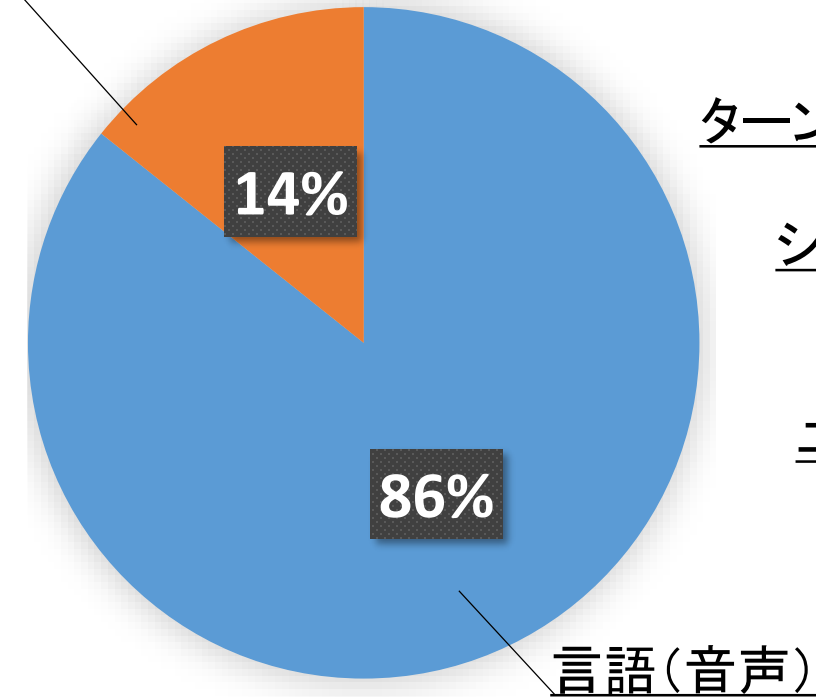
SIGDIAL 2018の国別の内訳



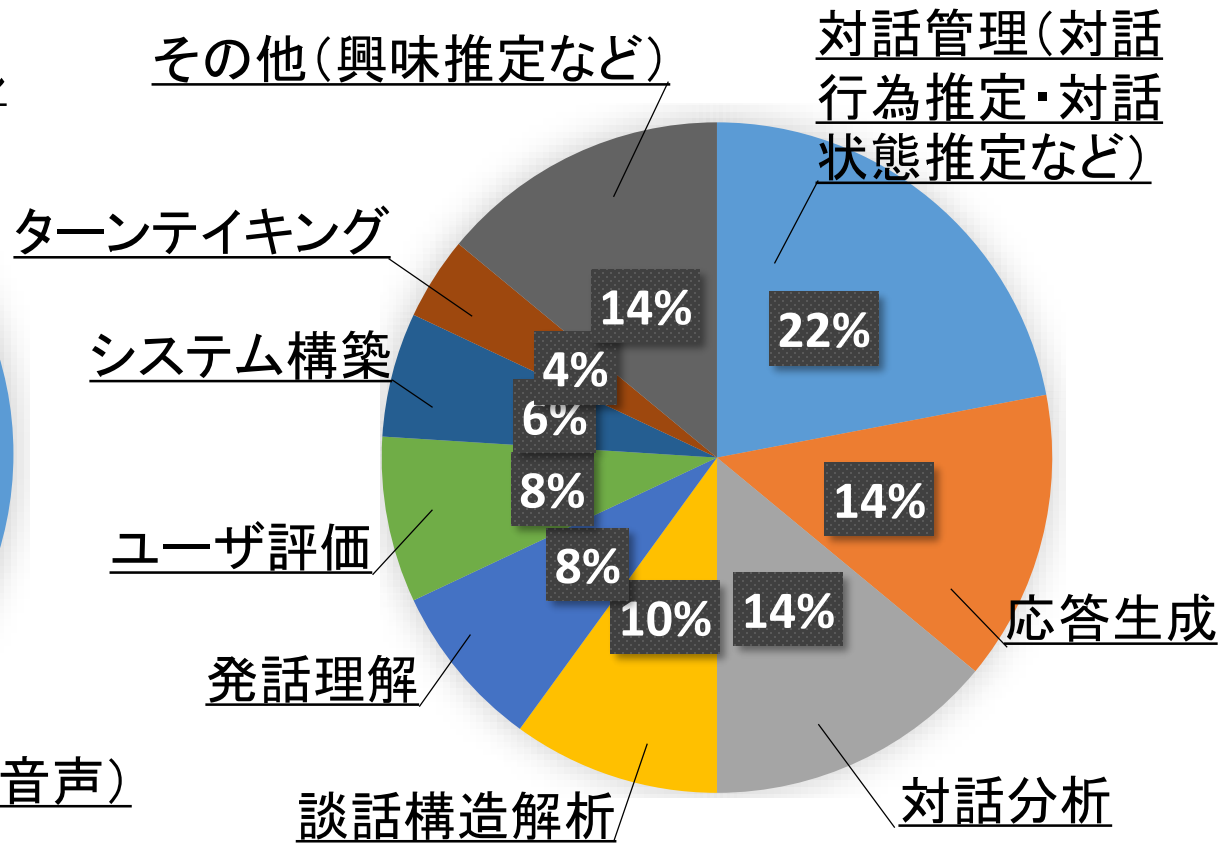
採択テーマから見る傾向

使用しているモダリティ
による分類

バイモーダル・マルチモーダル



テーマによる分類



- 分析系を含め、言語を対象とした論文が圧倒的多数
- ACLと併催だったがこちらでもニューラルベースの応答生成が多い
- 対話分析やユーザ評価、システム構築の論文が採択されるのもSIGDIALの特色

会議の様子

メルボルンの街並み



会議の様子

会場



Banquet



ピックアップ論文

- **R. Manuvinakurike et al., “Conversational Image Editing: Incremental Intent Identification in a New Dialogue Task” (ベストペーパー)**
 - 対話的に画像を編集するタスク
 - ユーザ発話の対話行為を推定
 - 対話行為の逐次推定も実施
- **T. Zhao and M. Eskenazi, “Zero-Shot Dialog Generation with Cross-Domain Latent Actions” (ベストペーパー)**
 - ニューラルベースの応答生成におけるドメイン適応
 - 少量用意した情報リソースを使ってタスク固有の知識を表現
- **Z. Hu et al., “Controlling Personality-Based Stylistic Variation with Neural Natural Language Generators”**
 - スタイルの制御が可能なニューラルベースの応答生成手法を提案
 - 入力とコンテキストにパーソナルを表現する情報を入れる

ピックアップ論文

- **L. Pragst and S. Ultes, “Changing the Level of Directness in Dialogue using Dialogue Vector Models and Recurrent Neural Networks”**
 - システムとの対話において発話の直接性を考慮する
 - 例) 質問: ‘What pizza do you want?’
 - 応答 (直接的): ‘I want a vegetarian pizza.’
 - 応答 (間接的): ‘I do not like meat.’
 - TDNNに基づく発話の直接性の推定手法を提案
- **A. Ravichander and A. Black, “An Empirical Study of Self-Disclosure in Spoken Dialogue Systems”**
 - 対話システムにおける自己開示の効果を評価実験で検証
 - 人間対機械でも自己開示に返報性があることを確認

所感

■ 対話システムに関連していればたいいていの研究が出せる

● 発表者の場合

- ◆ マルチモーダル情報を用いたユーザ状態の推定 [Chiba et al., 2014]
- ◆ 口調をだんだんとタメ口に変化させて親密度を表現する対話システムの評価 [Kageyama et al., 2018]
- ◆ 対話相手に協調して合成音声の感情を変える対話システムの評価 [Chiba et al., 2018]

● 査読者を納得させるテクニックは必要

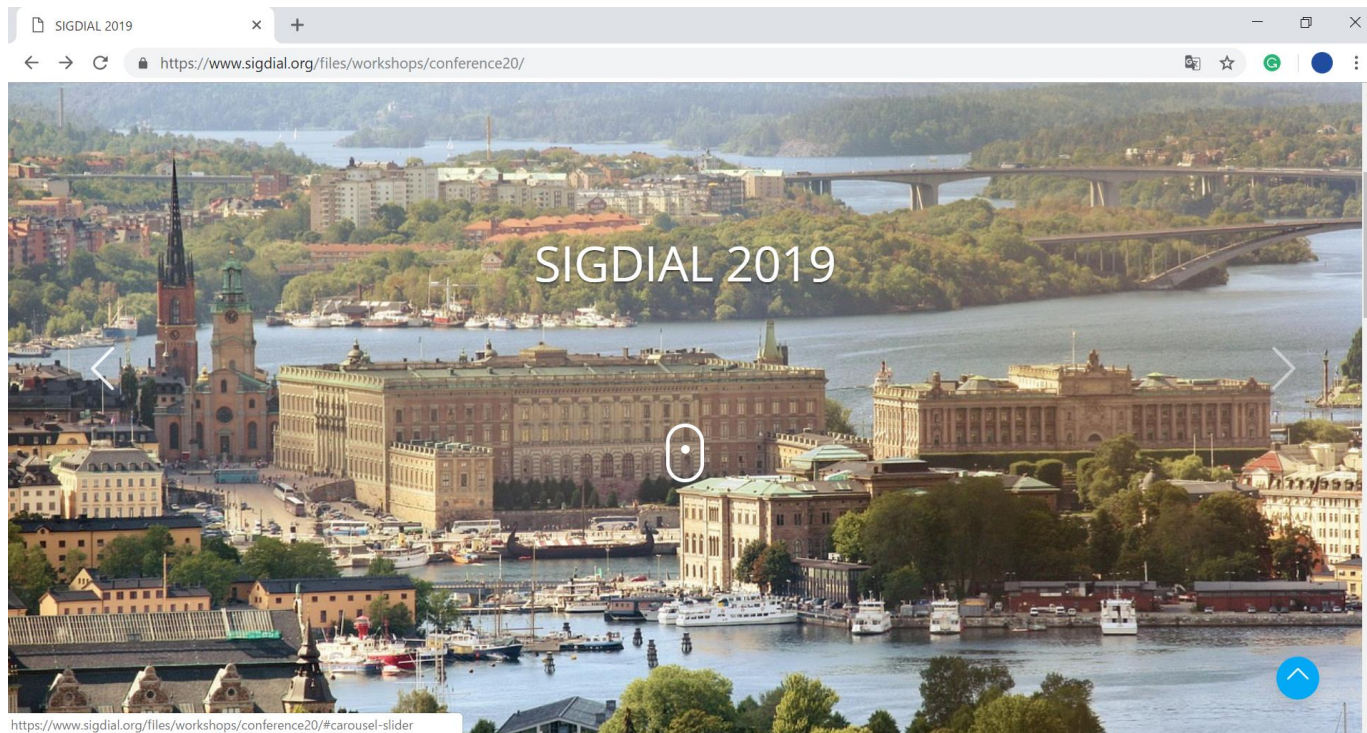
- ◆ 推定・生成系のタスクでは適切な評価尺度で性能改善を示す
- ◆ システム構築・ユーザ評価などは文献調査が重要

- ✓ 王道的なタスクだけでなく、変わったタスクでも挑戦可能
- ✓ シングルトラックなので色々な人に発表を聞いてもらえる
- ✓ LongでもOn-goingな研究が意外と多い

次回のSIGDIAL

■ SIGDIAL 2019

- 開催日 9/11-9/13
- 投稿締め切り日はまだ未定 (例年の傾向だと5, 6月くらい?)
- スtockホルムで開催
- INTERSPEECHと併催



国際会議報告 SemDial 2018



角森唯子
(NTTドコモ)

SemDial 2018 とは

■ 開催地

- プロヴァンス (フランス)

■ 期間

- 2018年11月8日～10日 (3日間)
- 初日はワークショップのみ

■ 参加人数

- 約 50 人
- うち日本人 3人



SemDial 2018 とは

■ 採択率

- 60 %

Oral: 17採択 / 28 投稿, Poster: 18採択 (投稿数は不明)

- 日本からの採択は2件 (Oral: 1, Poster: 1)
- Oral採択の約7割がヨーロッパ, 2割がアメリカ

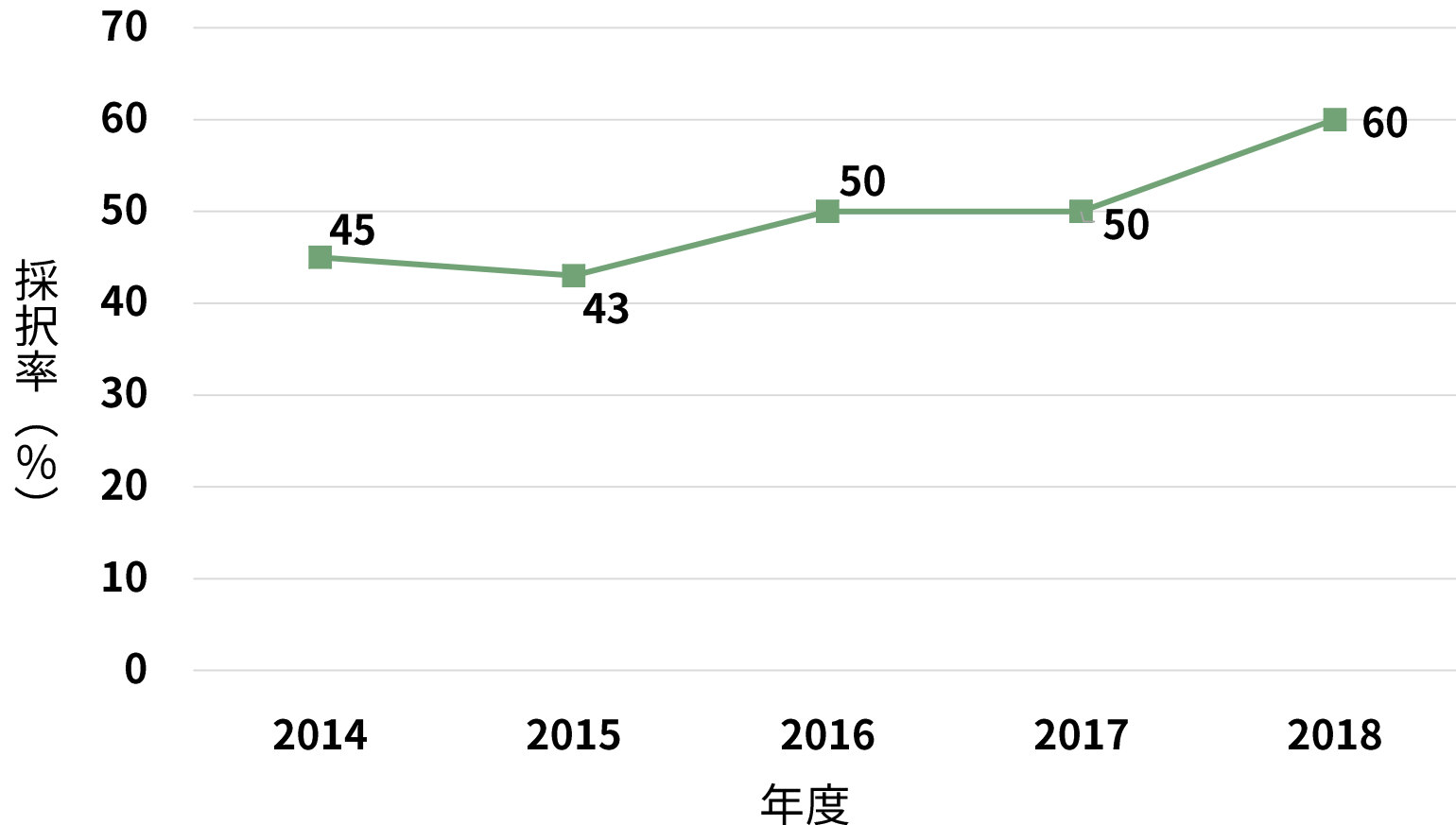
■ 発表の形態

- オーラルと ポスター
- シングルトラック
 - ◆ ポスターは小会場 (受付をした部屋)

■ チュートリアルはなし

■ 22回目の開催

5年間の採択率の推移



採択率が上昇傾向にあるので狙い目…？

対話系の発表

■ 会議全体に対する割合

- 基本的には、ほぼすべての発表が対話関連
- 対話システム、モダリティ、対話分析、談話解析など

■ 採択傾向 (Oral)

- モダリティ
 - ◆ 言語、音声、画像、ジェスチャ
- タスク
 - ◆ 対話システムを構築している発表のほとんどがタスク指向
- ニューラルを使用した発表は少ない
- モダリティや対話分析系の発表が多い

会場の様子（その1）



会場の案内板



↑ メイン会場入口

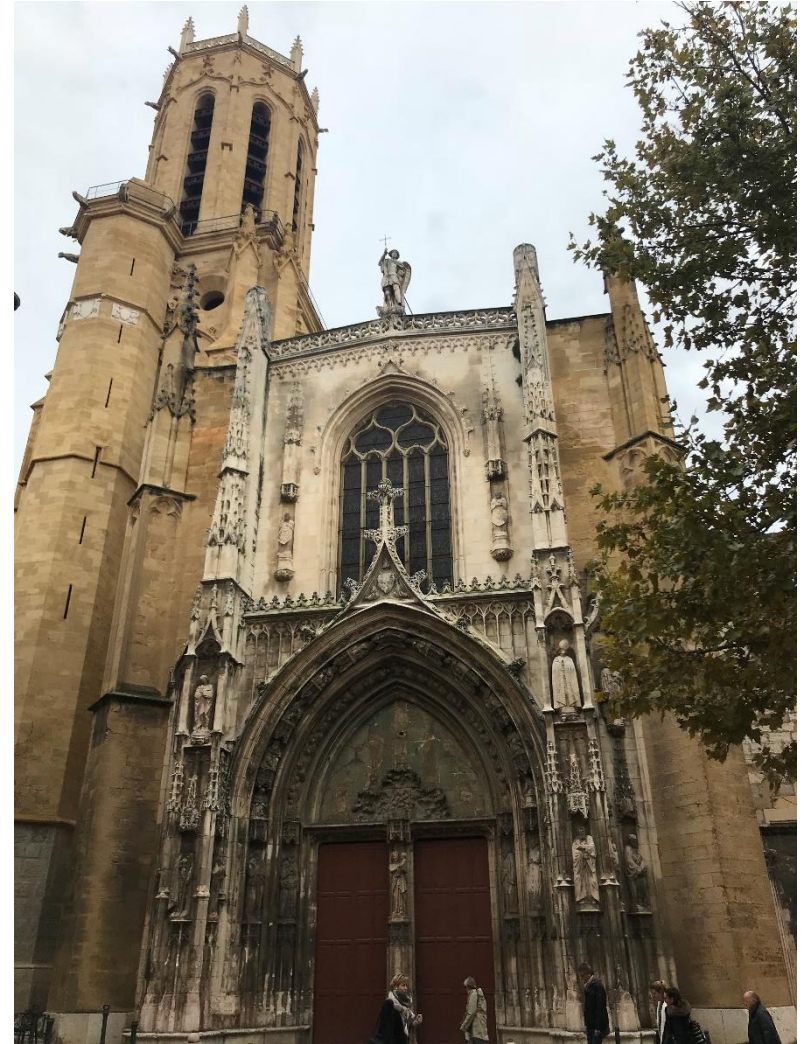
会場の様子（その2）



会場周辺の様子（その1）



会場周辺の様子（その2）



研究紹介（その1）

■ 対話システムの評価

- ✓ 主観評価における対話システムの重要な特徴を分析
Evaluating Subjective User Feedback for Internet of Things Dialogues

■ タスク指向対話システム

- ✓ タスクの回答にかかる時間におけるPOMDPを使用した対話制御
Learning to Buy Time: A Data-Driven Model For Avoiding Silence While Task-Related Information Cannot Yet Be Presented
- ✓ LSTMを用いたユーザ意図の認識．DSTCのデータセットで評価
A Multi-Task Approach to Incremental Dialogue State Tracking
- ✓ ユーザ入力の言いよどみ検出を備えたマルチタスクLSTMモデル
Multi-Task Learning for Domain-General Spoken Disfluency Detection in Dialogue Systems

研究紹介（その2）

■ モダリティ

- ✓ 画像に対して，検索ベースの手法でスケッチと短い説明を出力
Learning to describe multimodally from parallel unimodal data? A pilot study on verbal and sketched object descriptions

■ 対話の分析

- ✓ ユーザ意図を対話システムが誤解したとき，ユーザがどのように解決するかwozを使って収集したコーパスを分析
Investigating Strategies for Resolving Misunderstood Utterances with Multiple Intents
- ✓ 交渉を行うゲームにおける対話を収集し，シーンごとに分類
Categorisation of conversational games in free dialogue over spatial scenes

来年度の情報

■ SemDial 2019

- 場所：未定
- 例年，7～11月あたりで開催
- 投稿締切：2018年は6月10日
 - ◆ ページ数は，Oral: 8ページ，Poster: 2ページ
 - ◆ 小規模でアットホームな雰囲気
 - ◆ 少しでも対話に関係していれば，割とどんなものでも受け入れられる

対話（に少しでも関係する）研究はSemDial

国際会議報告

ACL • COLING • EMNLP 2018



増村 亮
(NTT)

ACL2018

■ ACLとは

- 毎年開催の自然言語処理のフラグシップ会議
- Short paper(5ページ)かLong paper(9ページ)

■ ACL2018の概要

- 開催地：メルボルン (オーストラリア)
- 期間：2018年7月15日～20日
- 採択率：25% (Short: 23.7%, Long:25.1%)



メルボルンの街並み



ACLの講演会場

COLING2018

■ COLINGとは

- 2年おき開催の自然言語処理のトップ会議
 - ◆ ACLやEMNLPよりもレベルは少し落ちるが、会議の歴史は自然言語処理系の中で最も長い
- Long paperのみ(10ページ)

■ COLING2018の概要

- 開催地：サンタフェ (アメリカ)
- 期間：2018年8月20日～26日
- 採択率：33% (accept後のwithdraw論文含む)



サンタフェの街並み



COLINGの講演会場

EMNLP2018

■ EMNLPとは

- ACLと並ぶ自然言語処理のトップ会議
 - ◆ 歴史は比較的浅めだが、近年投稿が急激に増加中。
- Short paper(5ページ)かLong paper(9ページ)

■ EMNLP2018の概要

- 開催地：ブリュッセル(ベルギー)
- 期間：2018年10月31日～11月4日
- 採択率：24.6% (Short: 23.2%, Long: 25.5%)



ブリュッセルの街並み



EMNLPの講演会場

ACL・COLING・EMNLPのレベル

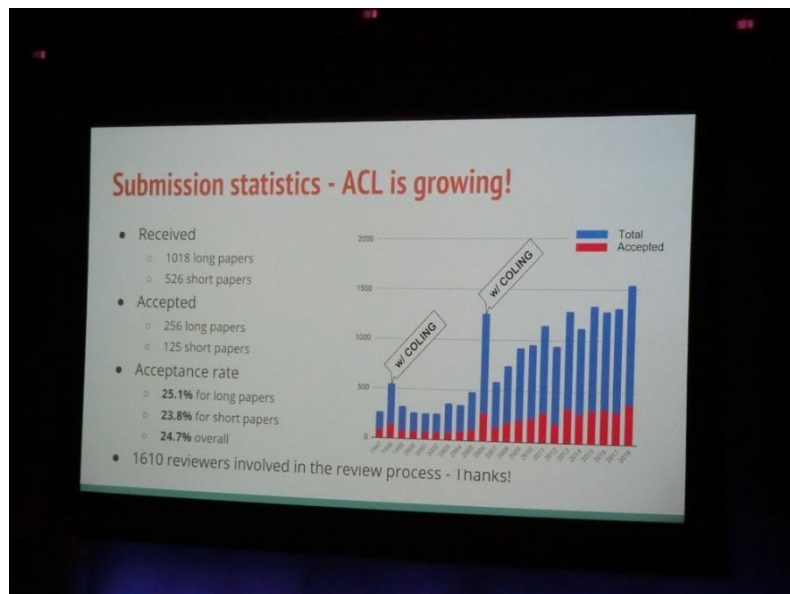
■ Google Scholarのh5-indexより

- ACL・EMNLP・NAACLが自然言語処理分野の3大トップ会議
- COLING・EACLなどはそれに続く

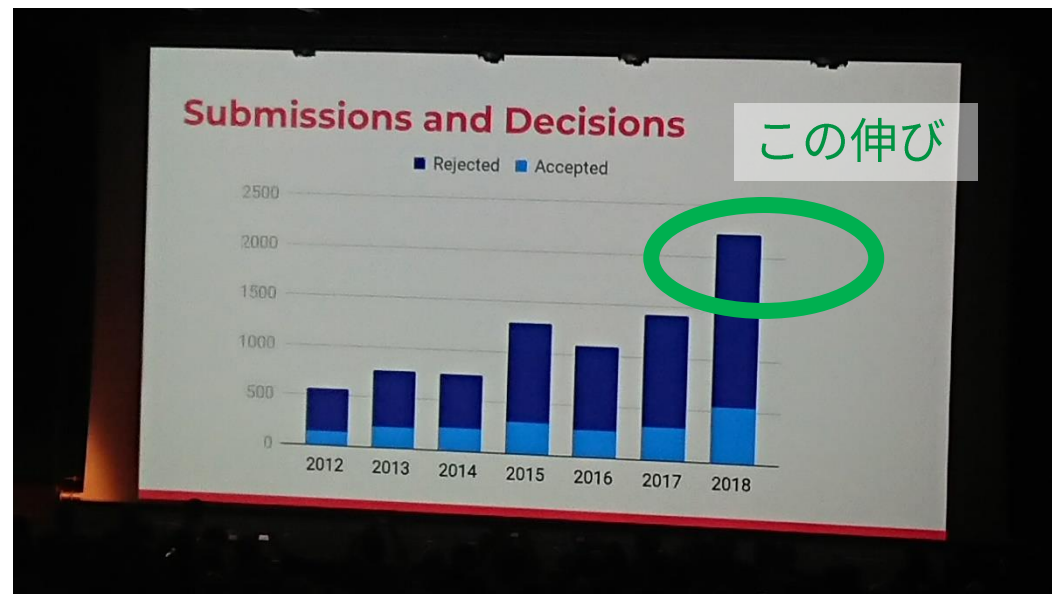
カテゴリ > Engineering & Computer Science > Computational Linguistics ▾			
	出版物	h5-指標	h5-中央値
1.	arXiv Computation and Language (cs.CL)	<u>79</u>	140
2.	Meeting of the Association for Computational Linguistics (ACL)	<u>67</u>	103
3.	Conference on Empirical Methods in Natural Language Processing (EMNLP)	<u>58</u>	108
4.	Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL)	<u>47</u>	79
5.	International Conference on Language Resources and Evaluation (LREC)	<u>43</u>	71
6.	International Workshop on Semantic Evaluation	<u>37</u>	62
7.	International Conference on Computational Linguistics (COLING)	<u>35</u>	43
8.	Transactions of the Association for Computational Linguistics	<u>32</u>	56
9.	Conference of the European Chapter of the Association for Computational Linguistics (EACL)	<u>31</u>	57
10.	Computational Linguistics	<u>31</u>	55

ACL・COLING・EMNLPの投稿数

- 他の人工知能関係の国際会議同様、右肩上がり
 - 特にEMNLPでは、2017年度よりも40%投稿が増加



ACLの統計情報



EMNLPの統計情報

EMNLPのレジストレーション待ちの列、約300m

長すぎる列を見守る人

列の流れ



- 各会議において、深層学習を用いた対話系技術に関するチュートリアルが実施
 - ACL: Neural Approaches to Conversational AI
 - ◆ Jianfeng Gao, Michel Galley and Lihong Li
 - ◆ <https://www.microsoft.com/en-us/research/uploads/prod/2018/07/neural-approaches-to-conversational-AI.pdf>
 - COLING: Deep Learning for Dialogue Systems
 - ◆ Yun-Nung Chen, Asli Celikyilmaz and Dilek Hakkani-Tür
 - ◆ https://www.csie.ntu.edu.tw/~yvchen/doc/DeepDialogue_Tutorial_ACL.pdf
 - EMNLP: Deep Chit-Chat: Deep Learning for ChatBots
 - ◆ Wei Wu and Rui Yan
 - ◆ <http://ruiyan.me/pubs/tutorial-emnlp18.pdf>

■ 会議全体に対する割合

● 10-15%程度 (COLINGは5%程度の印象)

- ◆ ニューラル会話モデルの台頭以降、対話の論文が増えた印象が強い、「生成」というトピックとコーパスの充実(Ubuntu Dialogue Corpus等)により、対話系以外の研究者にとっても研究トピックとなっている

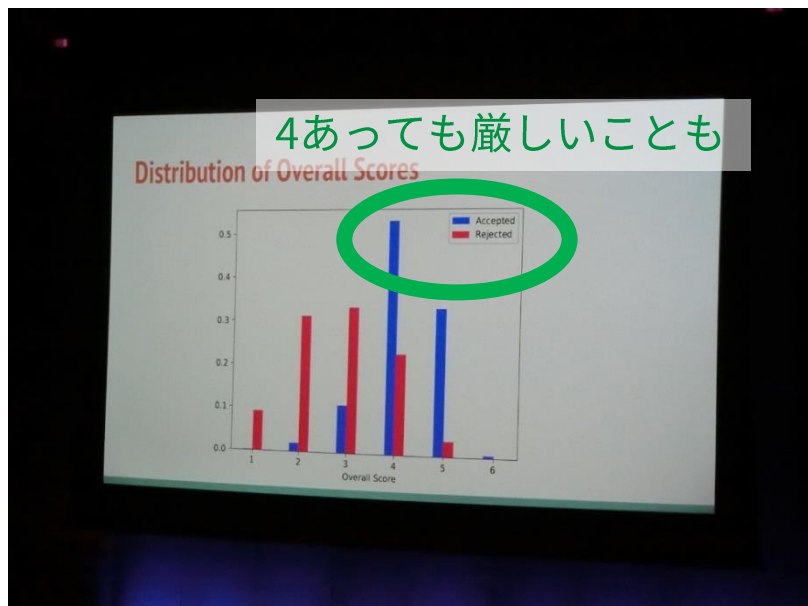
■ 今年度の採択傾向 (対話系に関して)

● ニューラル会話モデル(応答生成)の拡張が大部分

- ◆ 条件づける情報を増やしたモデル化(感情、対話行為等)
 - ◆ 応答にブレを入れるためのモデル化
 - ◆ ユーザがコントロール可能なモデル化
 - ◆ 外部知識を埋め込めるモデル化
 - ◆ 強化学習系
- ### ● 対話状態トラッキングや音声言語理解の検討も散見

ACL・COLING・EMNLPの査読

- Overall scoreが平均4近くでないとはacceptされない
 - どの査読者にも平均的によい印象を受けるようにする必要がある
 - ◆ 3はWeek reject, 4がAcceptのライン



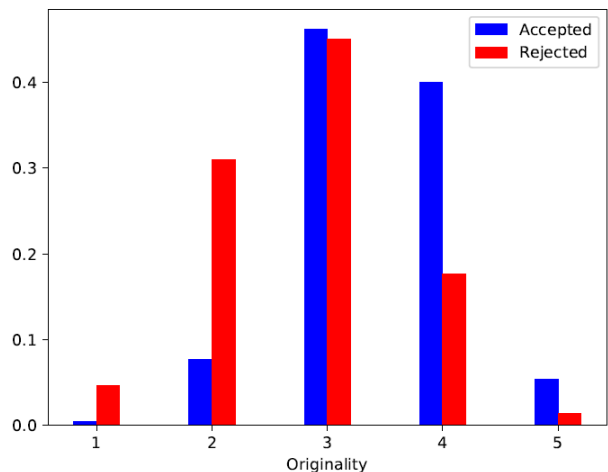
ACLの統計情報



EMNLPの統計情報

ACL・COLING・EMNLP に採録されるには？

- 査読はDouble Blindなので、著者は査読者に分からないので、ネームバリューは加味されない (arxiv・・・)
- Overall scoreで4をもらうためには独創性・新規性があることが最低ライン、そのうえでsolidな実験結果と、関連研究との明確なポジショニングが必要



ACLの統計情報

- これらの会議では、一時査読結果に対して、Rebuttalと呼ばれるコメントを返すことが可能、必須ではなくプラスにもマイナスにも
 - COLINGではRebuttalが有用だったという話もあった

ニューラル会話モデルに見る新規性を出すポイント

- 用例検索をしてから、それをシード情報にしてニューラル会話モデルで生成
Exemplar Encoder-Decoder for Neural Conversation Generation (ACL)
- 対話行為を推定してから、それをシードにしてニューラル会話モデルで生成
Dialogue-act-driven Conversation Model : An Experimental Study (COLING)
- Dialogue state trackingを行い、それをシードにしてニューラル会話モデルで生成
Sequence-to-Sequence Learning for Task-oriented Dialogue with Dialogue State Representation (COLING)
- 生成時の目的関数に話題等の制約を入れたニューラル会話モデル
Generating More Interesting Responses in Neural Conversation Models with Distributional Constraints (EMNLP)
- 生成時の目的関数に一貫性の制約を入れたニューラル会話モデル
Better Conversations by Modeling, Filtering, and Optimizing for Coherence and Diversity (EMNLP)

アイデアは単純で分かりやすいものも多い

来年度以降の情報

■ ACL 2019

- フィレンツェ（イタリア）
- 2019年7月28日～8月2日
- 投稿締切：3月4日

■ COLING 2020

- バルセロナ（スペイン）
- 未定

■ EMNLP-IJCNLP 2019

- 香港（中国）
- 2019年11月3日～7日
- 投稿締切：5-6月頃

ACL・COLING・EMNLP投稿のススメ

■ どんな研究を出すべき？

● 言語処理の範囲内での研究

- ◆ 特に機械学習によるモデル化に着眼した研究が理想的
 - 対話はState-of-the-artの性能争いにはなりにくいタスクなので、アイデアでも勝負可能
 - モデルかの検討をするなら、コーパスは、Ubuntu Dialogue CorpusやDialogue State Tracking Challengeのものがお手頃
- ◆ 分析系論文もあるが少なめなので、作戦を考えるべき

● 画像や音声情報も入力となるが、主眼が言語処理の研究

■ どんなことが嬉しい？

- 大規模な会議であり、対話に限らず言語処理の広い情報が得られる
- 採録されると技術分野での個人のプレゼンス向上にも大きく寄与

国際会議報告

IJCAI-ECAI 2018



成松 宏美
(NTT)

IJCAI-ECAI 2018 とは



AI全般を対象とするトップ会議

■ 会議名

the 27th International Joint Conference on Artificial Intelligence (IJCAI) と the 23rd European Conference on Artificial Intelligence (ECAI) の合同開催

■ 開催地：ストックホルム（スウェーデン）

■ 日程：2018年7月13日～19日（7日間）

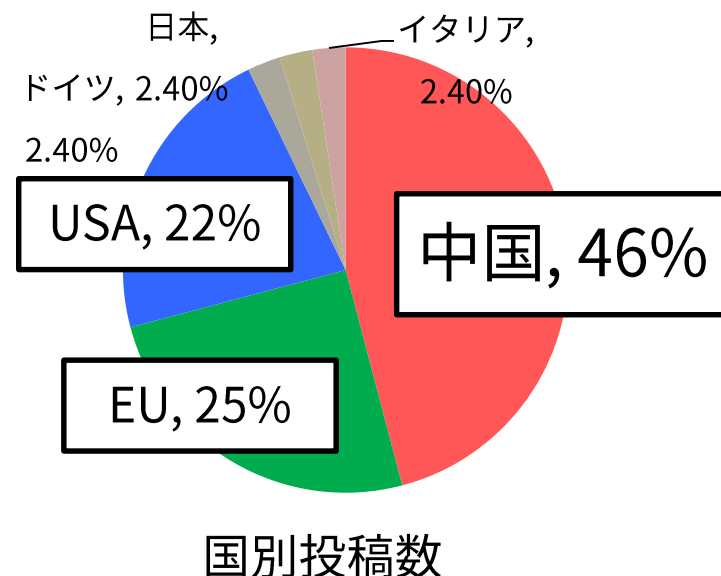
- 最初3日間は、ICML, AAMASと合同開催のワークショップ

■ 参加人数

- 約 2,500 名（日本からは約50-100名）

■ 採択率

- 20.5 % (710 採択 / 3470 投稿)



セッション構成

■ メイントラックの形態

- オーラル (口頭発表) と ポスター
- マルチトラック (11セッション並列)
 - ◆ 講演は大会場
 - ◆ 各セッションは50から100名規模の会場

■ チュートリアル：30件

- 対話関連は，エージェントベースの自動化系1件
Verifying Agent-Based Autonomous Systems

■ ワークショップ：66件

- 対話関連は2～5件

Linguistic and Cognitive Approaches To Dialog Agents (LaCATODA),
AI for Multimodal Human-Robot Interaction (AI-MHRI) etc.

ICML, AAMASと合同なこともあり、
メイントラックと同規模！！

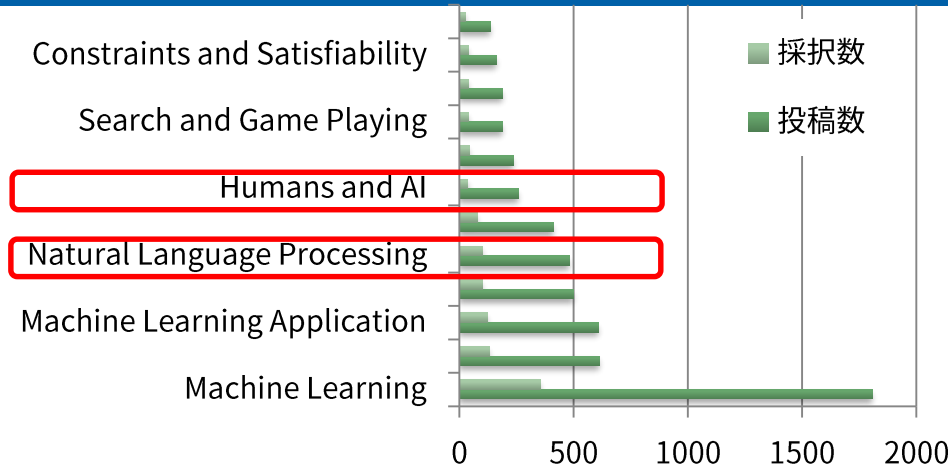
参加者同士の議論は活発！

対話系の発表

■ 会議全体に対する割合

- 機械学習系が主
- 対話は5%程度(発表20件)

意外に多い



■ 対話(言語・マルチモーダル・音声)と関連の発表数

招待講演数 (6)

セッション数 (6)

発表数 (41)

日本は
ほぼゼロ...

関連(概念)
(3)

対話系
(2)

関連(QA)
(3)

対話系
(2)

関連(言語)
(21)

対話(言語)
(15)

IJCAIも対話研究の投稿先として
視野に入れるべき対象になってきた。

関連(言語)
(1)

対話(その他)
(5)

会場の様子（講演会場）



会場の様子（セッション会場）



50-100名くらいの会場

会場の様子（懇親会）



会場周辺の様子（街中）



研究紹介（応答生成）

■ 雑談応答の精度向上

- ✓ マルチヘッドアテンションを用いた応答生成

Get The Point of My Utterance! Learning Towards Effective Responses with Multi-Head Attention Mechanism

- ✓ ナレッジグラフを用いた応答生成

Commonsense Knowledge Aware Conversation Generation with Graph Attention

- ✓ 抽出系モデルと生成系モデルの統合

An Ensemble of Retrieval-Based and Generation-Based Human-Computer Conversation Systems

■ 一貫性保証

- ✓ 複数人のソーシャルデータを用いた場合のパーソナルID推定

Assigning Personality/Profile to a Chatting Machine for Coherent Conversation Generation

- ✓ Seq2Seqの対話生成アプローチにおけるSemantic Matchingを機構を用いた応答生成

Reinforcing Coherence for Sequence to Sequence Model in Dialogue Generation

研究紹介（その他）

■ 検出

- ✓ 目的指向型対話に対する，強化学習アプローチによるスロットフィリング精度の向上
Submodularity-Inspired Data Selection for Goal-Oriented Chatbot Training Based on Sentence Embeddings
- ✓ 意味解析に基づくニューラルネットワークを用いた未知語検出
Learning Out-of-Vocabulary Words in Intelligent Personal Agents

■ 自動評価

- ✓ Adaboost手法の拡張による対話モデルの自動評価
Adaboost with Auto-Evaluation for Conversational Models

採択文献の傾向（主に言語）

■ 課題設定

- 応答生成系や強化学習を対象としたものが多い

■ 手法

- End-to-Endのニューラルネットワーク利用が多い

■ アプローチ面の特徴

- モデルの改良
- 外部知識との統合による新規性の創出

■ 評価

- 公開されているデータセットを利用
- 3~4アプローチを比較
- BLEU評価＋一部人手評価の併用

ほぼarXivに上がっている

来年度の情報

■ IJCAI2019

- マカオ（中国）
- 2019年8月10日～16日
- 投稿締切：2月19日アブスト，2月25日原稿締め切り
 - ◆ 卒論・修論の内容を投稿しやすい時期 (!?)
 - ◆ ページ数は6ページ
 - ◆ ワークショップも多数！（5月頃締め切り）

2020年度は
名古屋開催！

対話研究をぜひIJCAIへ！



IJCAI
2019
MACAO



国際会議報告 IVA 2018



石井 亮
(NTT)



■ 会議名称：

18th ACM International Conference on Intelligent Virtual Agents

#今年からACM SIGAIがスポンサーに入りACMの冠が付いた。

■ 会議概要： **バーチャルエージェント(VA)のトップ会議**

言語/非言語行動の理解・生成、ユーザ評価によるVAの行動、外見、インタラクションデザインの検討、アプリケーション構築などを広く扱う

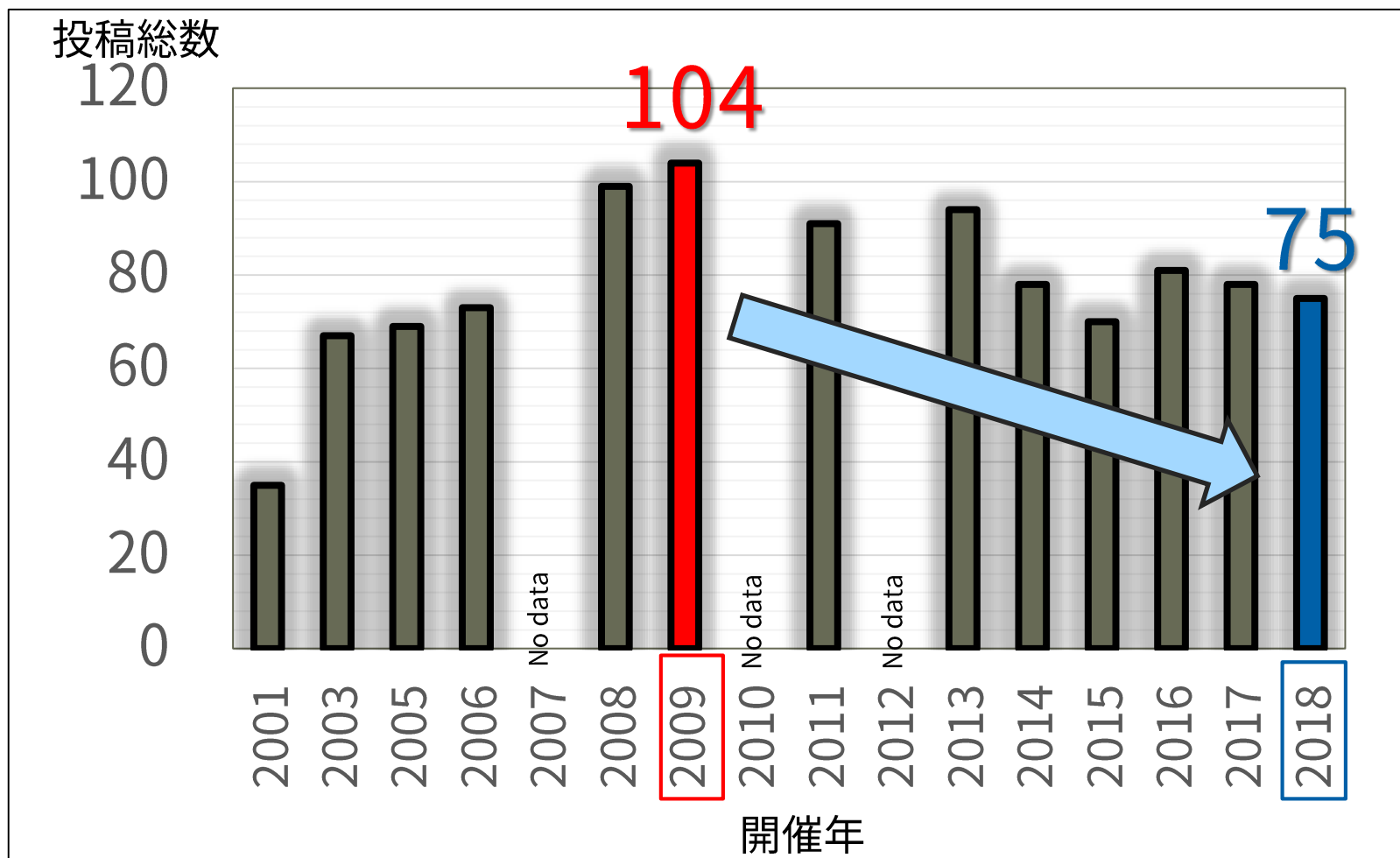
■ 日程： 2018年11月5-8日（4日間）

■ 開催地： Western Sydney University（オーストラリア・シドニー）

■ 参加人数： 90名程度（日本からは10名程度）

■ 会議構成： 1日目：WS×2、2-4日目：本会議（シングルトラック）

論文投稿数



10年間で3/4に減少
(ICMIなどに流れている模様)

査読システム・採択率

■ IVA査読システムの利点

査読評価結果に応じて

Full paper→Short paper→Poster

と段階的に採録してくれる。

■ 採択率

- Long paper: $17/49 = 34.7\%$ (+ Short採録: $20/49 = 40.8\%$)
- Short paper: $10/17 = 58.8\%$
- Poster/demo: $8 / 9 =$ ほぼ採択 (最終的に全部で18件)

以前 (10年前くらい) は Long での採録20%台だったが
投稿数が減少したために採択率が上昇
⇒ 現在は採録されやすいので狙い目

セッション構成

■ Keynote(4件):

- Hiroshi Ishiguro(阪大)
- Mark Sagar (Soul Machines/オークランド大) 他
⇒ Keynoteが充実。どれもとても面白かった。

■ Oral(27件):

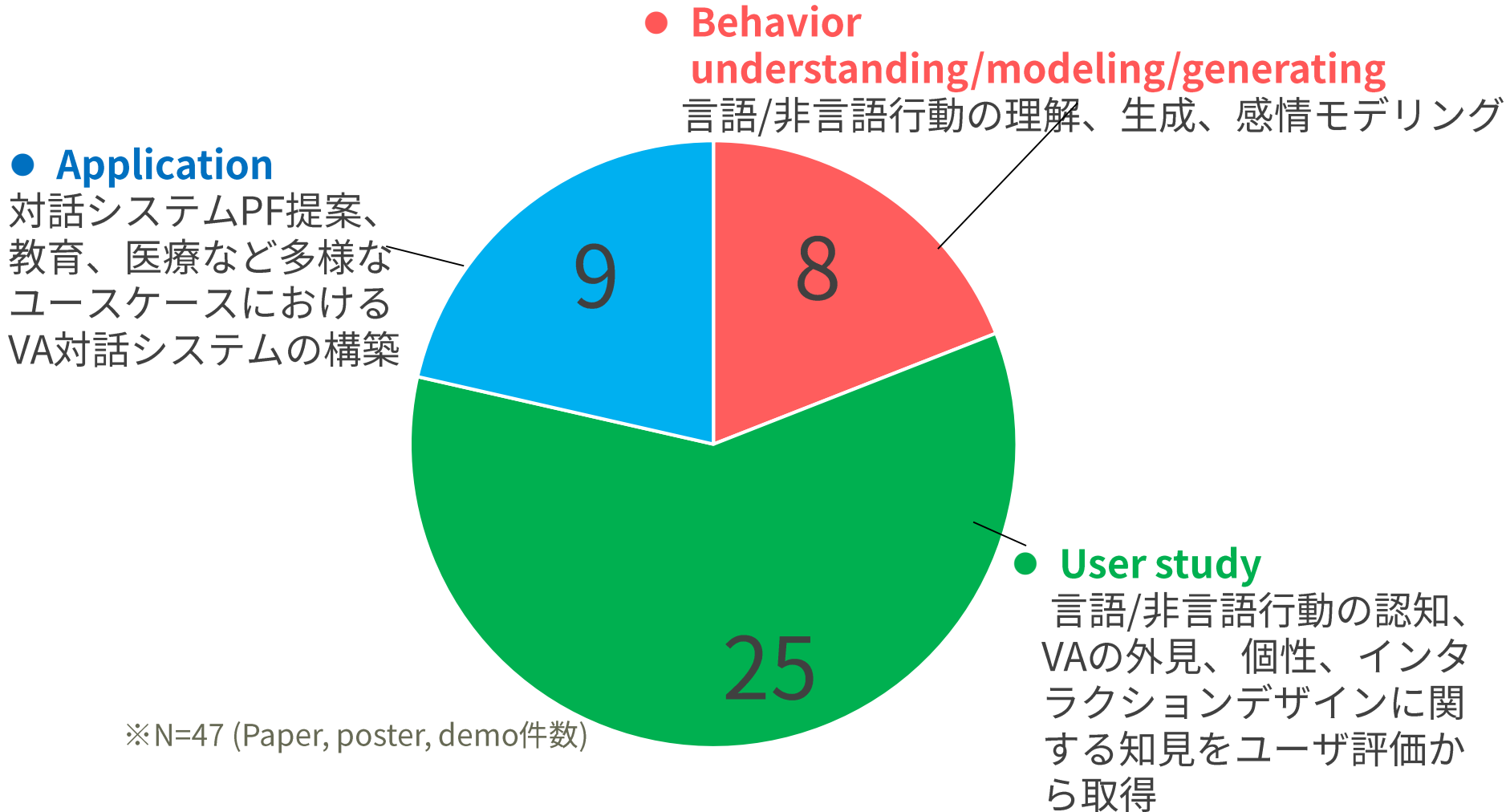
Evaluation, **Dialogue**×2, Generation, Health, **System**,
Perception×1.5, Personality, Simulation & curiosity
⇒ 対話セッション、対話システムセッションもあり

■ Poster(15件)、Demo(5件)、Gala(4件)



Human Computing Engine™
produced by Soul Machines

研究内容の傾向



■ 過去の15年間の研究のまとめの発表がありました

A Systematic Survey of 15 Years of User Studies Published in the Intelligent Virtual Agents Conference (Nahal Norouzi, etc.) , IVA2018.

研究紹介

- **Generating Body Motions using Spoken Language in Dialogue (Ishii+)**

発話言語から身体モーションを自動生成。対話時の人間の発話言語と頷き，頭部，表情，ハンドジェスチャ，上半身の姿勢の共起関係を学習し生成モデルを構築。

- **Culture, Errors, and Rapport-building Dialogue in Social Agents (Lucas+)**

エージェントが会話の誤りを起こしたときに，社会的対話が信頼の喪失を軽減・悪化することに，文化差が影響を与えるかどうかをユーザ実験から検証。

- **Flipper2.0 A Pragmatic Dialogue Engine for Embodied Conversational Agents (Heylen+) [Best student paper award]**

マルチモーダル情報を用いたVA対話システム開発PFの提案。（評価無し）

会場



講演会場



ポスター・デモ



Banquet



Banquet



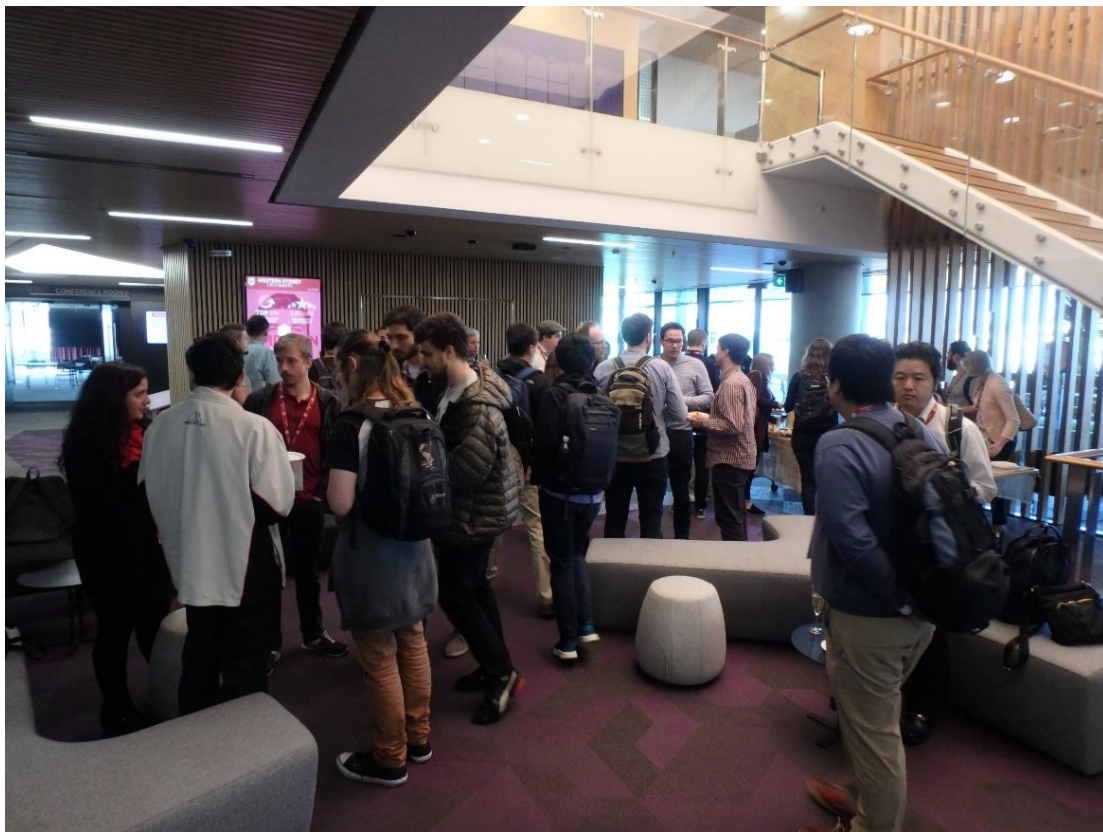
Banquet



Banquet



Closing session



集合写真



<会議の雰囲気>

- シングルトラックかつ小規模なため、密なディスカッションが可能。
- 雰囲気もかなりアットホームで、参加者間の交流も取りやすい。
- 若手研究者のヒューマンネットワークの構築、完成しきっていない研究成果の議論にも最適です。

投稿のすすめ

以前よりも現在は採録されやすいので狙い目。

- Behavior understanding/modeling/generating

数%の精度向上を争う世界ではなく、これまでにない新しい取り組み、アプローチが評価される。

- User study

ユーザ評価から面白い知見が見出せられればOK。評価品質については厳しくは問われない。

- Application

詳細な評価が無くても、新しいVAの設計方法、システム・アプリケーション提案でOK。特に開発PFに関しては評価は無くてもよい。



IVA2019 の情報

- 日程： 2019年7月2～5日
⇒今年よりも4ヶ月前倒し
- ペーパー投稿締め切り：**2019年3月1日**
- 開催地： CNRS(フランス国立研究センター) フランス・パリ
- CASAとの併設開催という新たな試み (KeynoteやBanquetが共催)
#CASA: International Conference on Computer Animation and Social Agents

A banner for IVA 2019 featuring a panoramic view of Paris with the Eiffel Tower in the center. The text 'IVA 2019' is overlaid on the left, and '2-5 July, Paris' is on the right. A 3D rendered woman in a blue dress stands on the right side of the banner.

IVA 2019

2-5 July, Paris

国際会議報告 ICMI 2018



福田 悠人
(埼玉大学)

ICMI 2018 とは

■ 開催地

- ボールダー, コロラド.
- (ツアーで本会議終了後にハイキング)

ICMI 2018

20th ACM International Conference on Multimodal Interaction
Boulder, Colorado, 16th-20th of October 2018

<https://icmi.acm.org/2018/>

■ 期間

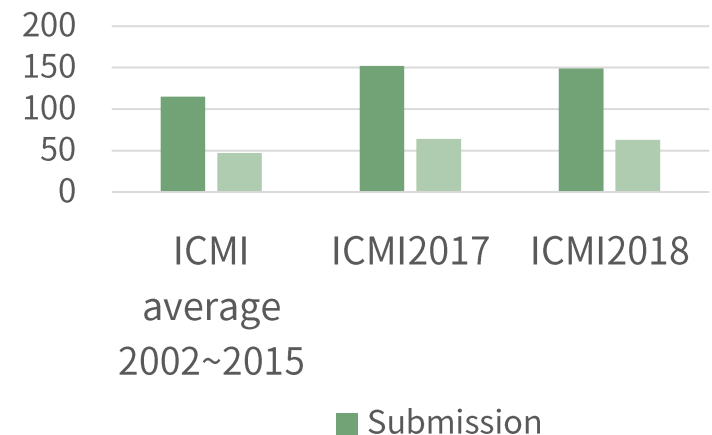
- 2018年10月16日～20日 (5日間)
- 初日はワークショップ, ドクターコンソーシアム
- 最終日はチュートリアル, グランドチャレンジ

■ 参加人数

- 約 200 人 強
- うち日本人 約20人

■ 採択率

- 42.3 % (63 採択 / 149 投稿)
- オーラル : 28 件 (18.7%) , ポスター : 35 件



ICMI 2018 とは

- **テーマ** : マルチパーティインタラクション
- **発表の形態**
 - オーラルとポスター
 - シングルトラック, 6セッション
(Multiparty Interaction, Artificial Agent, Sound and Interaction, など)
- **ワークショップ** : 5セッション
- **ドクターコンソーシアム**
- **チュートリアル** : 行動解析, 深層学習
- **グランドチャレンジ** : 感情認識, 食事解析

※ドクターコンソーシアム, グランドチャレンジの発表は
本会議のポスター発表においても発表

対話系の発表

■ 会議全体に対する割合

- 全体の3割強ほど
- 対話システムや感情認識ターンテイキング検出など

■ 採択傾向 (対話系に関して)

- モダリティ
 - ◆ マルチモーダル (音声, 言語, 画像, 生体情報, 等)
- タスク
 - ◆ (対話に限らず) 幼児, 高齢者, 精神疾患者を対象にしたもの
 - ◆ 多人数対話 (会議支援など)
- 検出, 認識はほとんどが深層学習を用いた手法

バンケットの様子（その1）



バンケットの様子（その2）



研究紹介（その1）

■ 発話交代の予測

- ✓ マルチスケールなRNNを用いて，マルチモーダル（音声，言語）のデータを扱う

Multimodal Continuous Turn-Taking Prediction Using Multiscale RNNs

- ✓ リモートコミュニケーション（テレビ電話）において，対象者以外のデータを用いて発話を予測

Multimodal Modeling of Coordination and Coregulation Patterns in Speech Rate during Triadic Collaborative Problem Solving

■ 対話システム

- ✓ 幼児の学習を支援する対話システム (**Best paper**)

Multimodal Dialogue Management for Multiparty Interaction with Infants

研究紹介（その2）

■ 感情認識

- ✓ fMRIの情報を用いて学習したネットワークで音響特徴を生成する
Generating fMRI-Enriched Acoustic Vectors using a Cross-Modality Adversarial Network for Emotion Recognition
- ✓ 相対的な感情の変化を認識に利用
Multimodal Local-Global Ranking Fusion for Emotion Recognition
- ✓ 対話相手のユーモアの好みの理解
How to Shape the Humor of a Robot – Social Behavior Adaptation Based on Reinforcement Learning

■ Personality の認識

- ✓ 発話者と発話交代時の他者の発話を用いて推定
Using Interlocutor-Modulated Attention BLSTM to Predict Personality Traits in Small Group Interaction

来年度の情報

■ ICMI 2019

- 蘇州（中国）
- 2019年10月14日～18日
- 投稿締切：5月7日（abstractは5月1日）
 - ◆ Long paper (8 pages) , Short paper (4 pages)
 - ◆ 小規模でアットホームな雰囲気会議
 - ◆ マルチモーダルであればどんなものでも受け入れられる



国際会議報告 Interspeech 2018



井上 昂治
(京都大学)

Interspeech 2018 とは

■ 開催地

- ハイデラバード (インド)
- Google や Microsoft のオフィスがある

■ 期間

- 2018年9月2日～6日 (5日間)
- 初日はチュートリアルのみ

■ 参加人数

- 約 1,000 人 (それ以上)
- うち日本人 約50～100人

■ 採択率

- 44.9 %
- 749 採択 / 1668 投稿



Interspeech 2018 とは

■ 発表の形態

- オーラル (口頭発表) と ポスター
- マルチトラック
 - ◆ 流行りのセッションは大会場
 - ◆ 対話系は小会場

■ チュートリアル

- 8 件
- 例年 1 件ほどが対話関係
- Multimodal Speech and Audio Processing in Audio-Visual Human-Robot Interaction

対話系の発表

■ 会議全体に対する割合

- 全体の1割かそれ以下
- メインは音声認識や話者照合など

■ 採択傾向 (対話系に関して)

- モダリティ
 - ◆ 音声の基本
- タスク
 - ◆ スマートスピーカーを想定したものが多い
- ほとんどの手法が機械学習 (ニューラルネットワーク)
- 公開データセットを用いた客観評価が多い

会場の様子（その1）



会場の様子（その2）



会場の様子（その3）



会場の様子（その4）



会場周辺の様子（その1）



会場周辺の様子（その2）



研究紹介（その1）

■ Dialogue act (対話行為ラベル) の認識

- ✓ Attention-mechanism (注意機構) を用いた文脈情報の考慮
Conversational Analysis using Utterance-level Attention-based Bidirectional Recurrent Neural Networks
- ✓ 言語と韻律の両方の情報を用いる (事前学習)
Conversational Analysis using Utterance-level Attention-based Bidirectional Recurrent Neural Networks

■ 感情認識

- ✓ 音響と言語の両方の情報を用いる (結果統合)
Deep Neural Networks for Emotion Recognition Combining Audio and Transcripts
- ✓ 時間方向と特徴量次元の2次元空間に対して、CNNで全体を、BLSTMで時間方向のAttentionを学習
Exploring Spatio-Temporal Representations by Integrating Attention-based Bidirectional LSTM-RNNs and FCNs for Speech Emotion Recognition

研究紹介（その2）

■ Personality の認識

- ✓ 韻律，言語，心理学的特徴に基づく BigFive の二値分類
Lexical And Acoustic Deep Learning Model For Personality Recognition

■ 年齢層の推定

- ✓ 時系列の音声情報からEnd-to-end で年齢層を推定
End-to-end Deep Neural Network Age Estimation

■ 性別の推定

- ✓ 大規模な別タスク(音声イベント分類) のデータで学習したモデルを転移学習
Improving Gender Identification in Movie Audio Using Cross-Domain Data
- ✓ 雑音除去と性別推定のネットワークとワークを結合することで雑音音声でも認識できるように
Denoising and Raw-waveform Networks for Weakly-Supervised Gender Identification on Noisy Speech

来年度の情報

■ Interspeech 2019

- グラーツ (オーストリア)
- 2019年9月15日～19日
- 投稿締切：3月29日
 - ◆ 卒論・修論の内容を投稿しやすい時期 (!?)
 - ◆ ページ数は4ページで少なめ (!?)



« Crossroads of **S**peech and **L**anguage »

音声系の研究であれば Interspeech へ !

まとめ

	モダリティ	対話系の割合	採択率	流行のタスク
<u>SIGDIAL</u>	言語 / 音声 / マルチモーダル		54 % (L) 31 % (S)	対話管理 応答生成
<u>SemDial</u>	何でも 言語 / 音声 / 画像 / ジェスチャ	ほぼ 100 %	60 %	システム構築なら タスク指向対話
<u>ACL</u>		10-15 %	25 %	
<u>COLING</u>		5 %	33 %	
<u>EMNLP</u>	言語	10-15 %	24.6 %	応答生成
<u>IJCAI</u>		10 % 以下	20.5 %	
<u>IVA</u>	エージェント系	30 % 超	34.7 % (L) 58.8 % (S)	システム提案 ユーザ評価
<u>ICMI</u>	マルチモーダル	30 % 超	42.3 %	幼児・高齢者 多人数対話
<u>Interspeech</u>	音声	10 % 以下	44.9 %	スマートスピーカを 想定したもの

※各報告者の主観を含みます

投稿時期まとめ

- IJCAI 2019 - 2019年 2月 25日
- IVA 2019 - 2019年 3月 1日
- ACL 2019 - 2019年 3月 4日
- Interspeech 2019 - 2019年 3月 29日
- ICMI 2019 - 2019年 5月 7日
- SIGDIAL 2019 - 2019年 5, 6 月 くらい
- EMNLP-IJCNLP 2019 - 2019年 5, 6 月 くらい
- SemDial 2019 - 2019年 6月 くらい
- COLING 2019 - 未定

会議会場でお会いしましょう！