対話システムシンポジウム
2019/12/03

国際会議報告

ACL: 高山 隼矢（大阪大学）

NLP4CONVAI: 田中 翔平（NAIST）

SIGDIAL: 松山 洋一（早稲田大学）

INTERSPEECH: 小林 優佳 (東芝)、稲熊 寛文（京都大学）

ICMI: 田中 宏季（NAIST）

本報告の趣旨

■対話研究者の国際会議への投稿・参加のお手伝い

- 対話は関連する分野が多岐にわたるので、どこに投稿するかの判断が難しい

  自然言語処理？音声？HRI？HMI？マルチモーダル？機械学習？

- どんな学会に投稿するのがいいか
- どんなネタなら通りやすいか
- 採録されるためのテクニック

■国際会議の研究動向について共有

- 最新の研究動向について紹介

報告する会議

ACL	高山 隼矢（大阪大学）
NLP4ConvAI	田中 翔平（NAIST）
SIGDIAL	松山 洋一（早稲田大学）
INTERSPEECH(対話)	小林 優佳 (東芝)
INTERSPEECH(音声)	稲熊 寛文（京都大学）
ICMI	田中 宏季（NAIST）

第10回対話システムシンポジウム
3rd Dec 2019

国際会議参加報告
ACL2019@Florence

大阪大学大学院情報科学研究科　マルチメディア工学専攻
博士後期課程1年
高山 隼矢

ACL2019@Florence

➢開催概要

- ●開催地　：バッソ要塞
  （フィレンツェ, イタリア）
- ●開催期間：2019/07/28 – 2019/08/02

➢統計情報

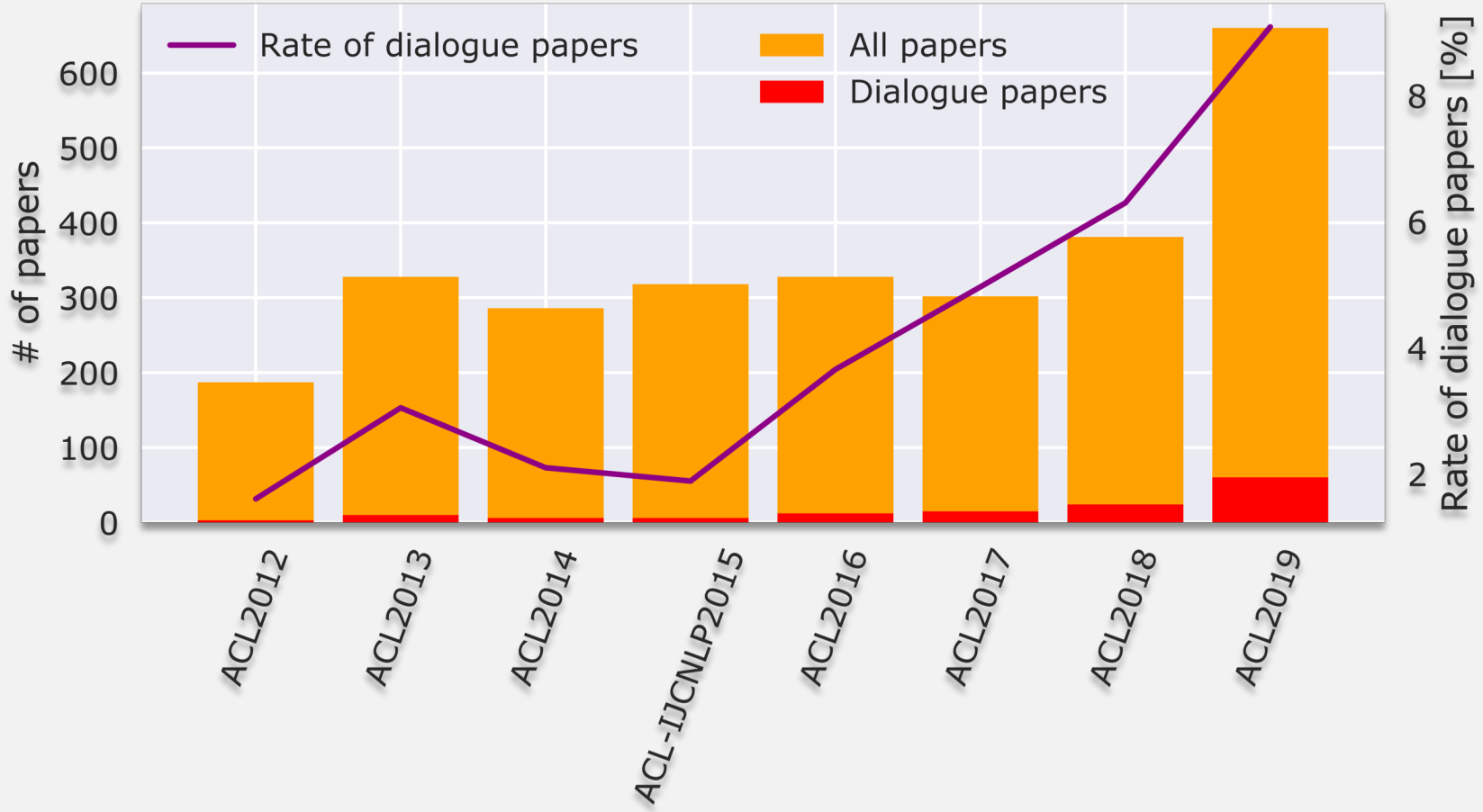
- ●投稿数が激増
  （1571本 → 2905本）
- ●Long Paper の採択率は
  例年とそれほど変わらず,
  Short Paper は低下

	Submitted	Accepted	Rate [%]
Long	1740	447	25.7
Short	1165	213	18.2
Total	2905	660	22.7
(対話*)	183	52	28.4

* : Dialogue and Interactive Systems エリアへの投稿論文

対話論文は増加傾向

➢採択論文のうち対話論文* が占める割合の変化



* : "dialog", "conversation", "chat", "response" のいずれかをタイトルに含む long/short 論文

応答生成・選択が主流

- ACL2019 対話論文と SIGDIAL2019 の Abstract における bi-gram の平均出現頻度の差
  - **ACL** > SIGDIAL, 上位20

    mult_turn, **respons_gen**, dialog_model, larg_scal, **respons_select**, expery_result, op_domain, sign_improv, sent_funct, sign_outperform, **encod_decod**, outperform_stat, **gen_model**, hum_evalu, stat_art, model_learn, dialog_hist, convers_model, **retriev_bas**, turn_convers

    →応答生成・選択に関する論文が多い

  - **SIGDIAL** > ACL, 上位20

    stat_track, **dialog_stat**, dialog_system, answ_quest, **slot_valu**, respons_typ, us_', spok_dialog, reinforc_learn, **slot_fil**, real_world, e_g, g_.,, log_dialog, convers_ag, **dialog_act**, valu_within, quest_/, +_slot

    →状態追跡, Slot filing, 対話行為 など多岐に渡る

今年のトレンドは対話履歴の活用

- ACL2019 ⇄ ACL2018 で同様の比較
  - **ACL2019** > ACL2018, 上位20

    op_domain, dialog_model, **dialog_hist**, domain_dialog, quest_answ, larg_scal, propos_new, machin_read, gen_quest, ground_convers, improv_qual, **convers_hist**, hum_evalu, publ_avail, convers_dataset, **discours_rel**, **relev_context**, dialog_task, **convers_context**, met_word

    →対話履歴の活用（一貫性の向上）に焦点を当てた論文が多い

  - **ACL2018** > ACL2019, 上位20

    end_end, typ_decod, 3_%, sequ_sequ, langu_model, train_dat, seq2seq_model, **gen_respons**, real_us, **stat_track**, divers_requir, **control_sent**, unlabel_dat, **profil_inform**, **senty_class**, langu_learn, build_dialog, **ordin_word**, rec_propos

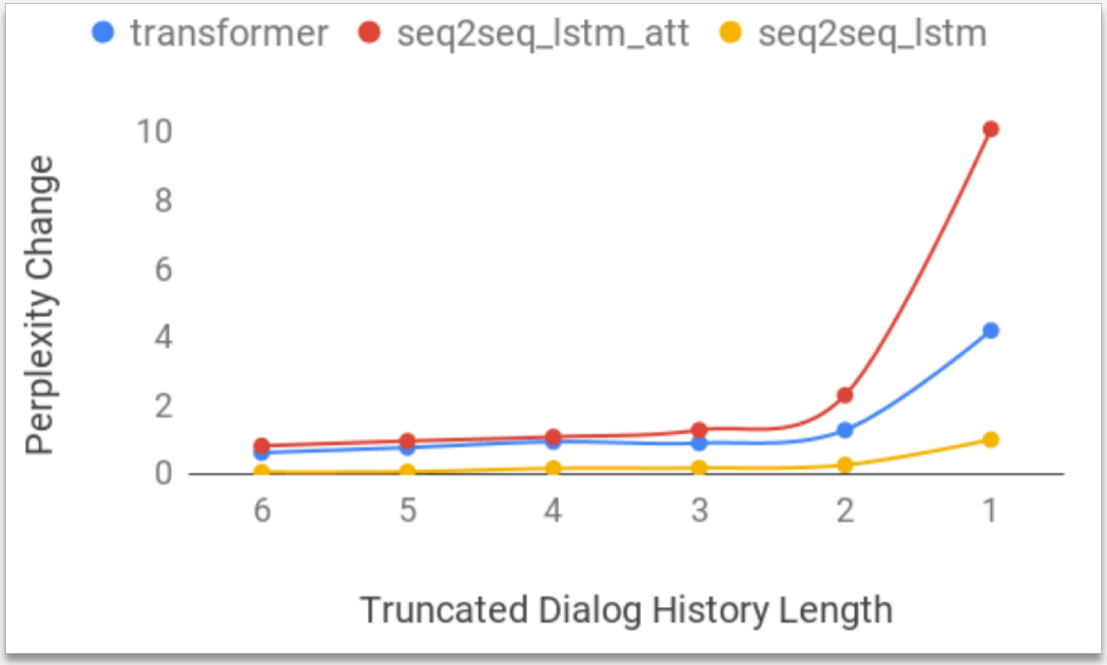
    →制御性, パーソナライズ, 感情, 多様性などのトピックはやや減少？

既存モデルは対話履歴を活用できていない

➢ **Do Neural Dialog Systems Use the Conversation History Effectively? An Empirical Study.**
Chinnadhurai Sankar et al. *Best Short Paper Nomination

- ニューラル生成モデルが対話履歴を考慮できているかどうかを調査
- 対話履歴に人為的な摂動を加えたときの PPL の増加を観察する
  ΔPPL 大→履歴に対して sensitive
- 対話履歴のシャッフルや切り捨て，発話中の語順入れ替え等の摂動に対して RNN, Transformer 共に sensitivity が低かった

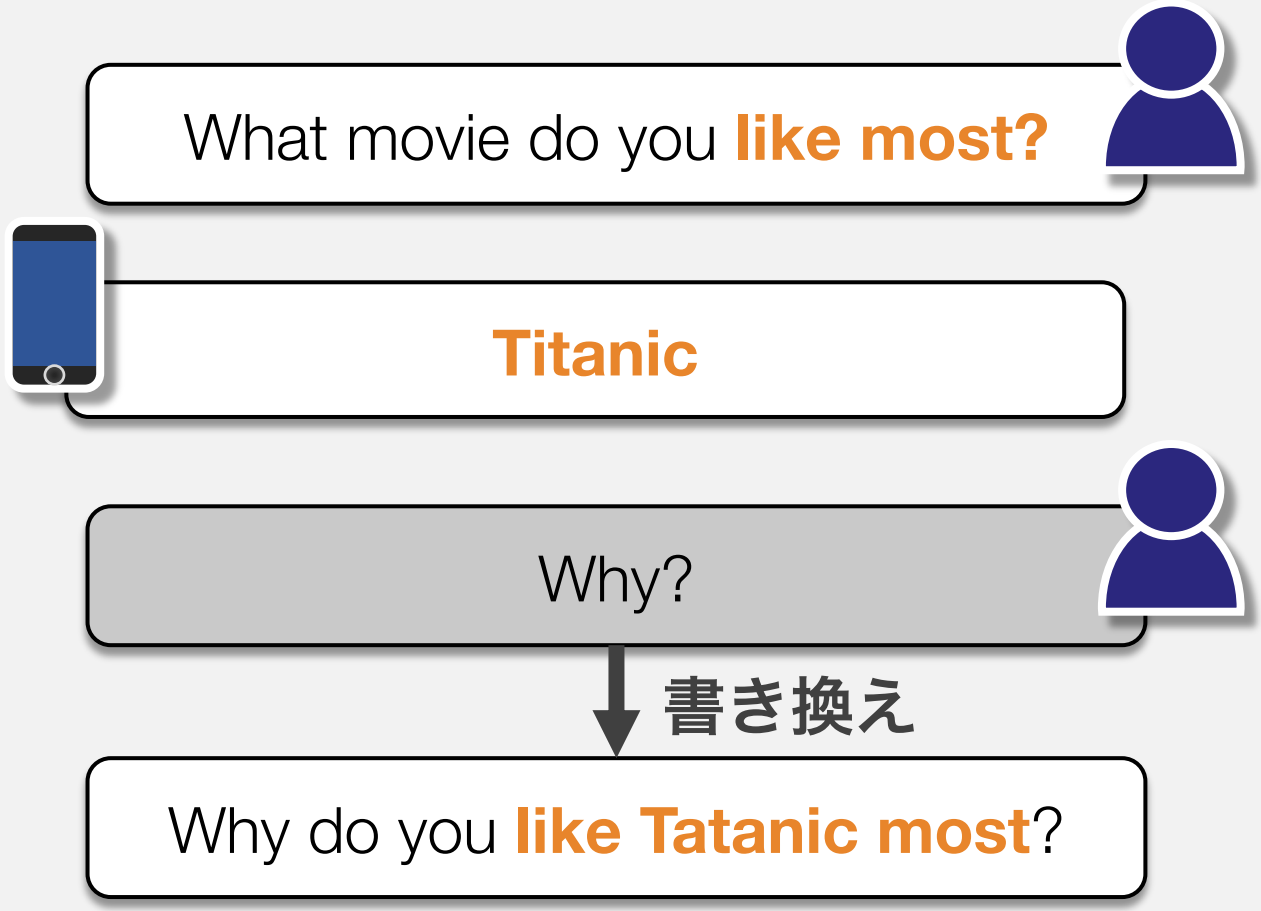
→十分に履歴を活用できていない



対話履歴の情報を含むように発話を書き換え

➢ **Improving Multi-turn Dialogue Modelling with Utterance ReWriter.**
Hui Su et al.

- 過去の対話履歴を用いて直前のユーザ発話を書き換え
- PointerNet ベースのモデルを用いて対話履歴とユーザ発話からコピーする単語を選択
- 評価は Online A/B テスト
- 書き換えデータセットは後に公開予定とのこと



対話履歴の情報を含むように発話を書き換え

- **Improving Multi-turn Dialogue Modelling with Utterance ReWriter.**
  Hui Su et al.
  - 過去の対話履歴を用いて直前のユーザ発話を書き換え
  - PointerNet ベースのモデルを用いて対話履歴とユーザ発話からコピーする単語を選択
  - 評価は Online A/B テスト
  - 書き換えデータセットは後に公開予定とのこと

Model	Intention Precision	CPS
Original	80.77	6.3
With Rewrite	89.91	7.7

Table 9: Results of integrated testing. Intention precision for task-oriented and conversation-turns-per-session (CPS) for chitchat.

対話履歴の活用に関する論文@ACL2019

- **Self-supervised Dialogue Learning.** Jiawei Wu et al.
  - 発話の順序が正しいかどうかを判定する Discriminator を用いて Adversarial に応答生成モデルを訓練

- **ReCoSa: Detecting the Relevant Contexts with Self-Attention for Multi-turn Dialogue Generation.** Hainan Zhang et al
  - Self-attention メカニズムによって遠く離れた依存関係を捉える

- **Dialogue Natural Language Inference.** Sean Welleck et al.
  - ペルソナ文と応答文の含意関係認識結果を用いて応答をリランキング
  - Persona-Chat Dataset を基にした Dialogue-NLI Dataset を公開

その他所感

- Best paper awards にノミネートされた論文の傾向（主観）
  - 経験的に知られていたことを定量的分析によって示す
    - Do Neural Dialog Systems Use the Conversation History Effectively? An Empirical Study.
  - 既存の "タスク" の問題点を指摘し，新たにタスク設定 + データ構築
    - OpenDialKG: Explainable Conversational Reasoning with Attention-based Walks over Knowledge Graphs.
    - Persuasion for Good: Towards a Personalized Persuasive Dialogue System for Social Good.
  - "何が" できるようになったのかが明確に示されている
    - Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems.

- SOTA であるかどうか以上に，目的に即したタスク設定と明確な分析がなされているかどうかが評価されている感じ

来年度の情報

➢開催概要

- ●開催地　：Hyatt Regency Seattle
  （シアトル, アメリカ）
- ●開催期間：2020/07/05 – 2020/07/10
- ●投稿締切：2019/12/09 23:59 AoE（来週！）

国際会議報告
NLP4ConvAI
ACL 2019 Workshop

田中翔平[1]

[1]奈良先端科学技術大学院大学

NLP for Conversational AI (ConvAI)

ACL と同じ会場（バッソ要塞）で 8/1 に開催

会議の傾向

68	25	36%	400
submissions	acceptances	acceptance rate	attendees

9本は ACL に採択された論文の二重投稿

Generation, DST など対話に関連する幅広いトピック

招待講演が豪華（みんなこっち目当て？）

個人的には Matthew Henderson (PolyAI) の話が面白かった

Should Conversational AI use neural response generation?

By Prof. Verena Rieser

Heriot-Watt University の先生

最近の対話システムにおける応答生成全般の話

E2E NLG Challenge という応答生成の shared task の紹介など

The Curious Case of Degenerate Neural Conversation

By Yejin Choi

University of Washington の先生

結局今まともに動く対話システムって E2E だけでは難しいみたいな話

ATOMIC, COMET (cause-effect の知識グラフ) についての話もしていた

Best Paper Awards

Regular Workshop Papers

- Conversational Response Re-ranking Based on Event Causality and Role Factored Tensor Event Embedding
  by ***Shohei Tanaka***, *Koichiro Yoshino, Katsuhito Sudoh and Satoshi Nakamura*

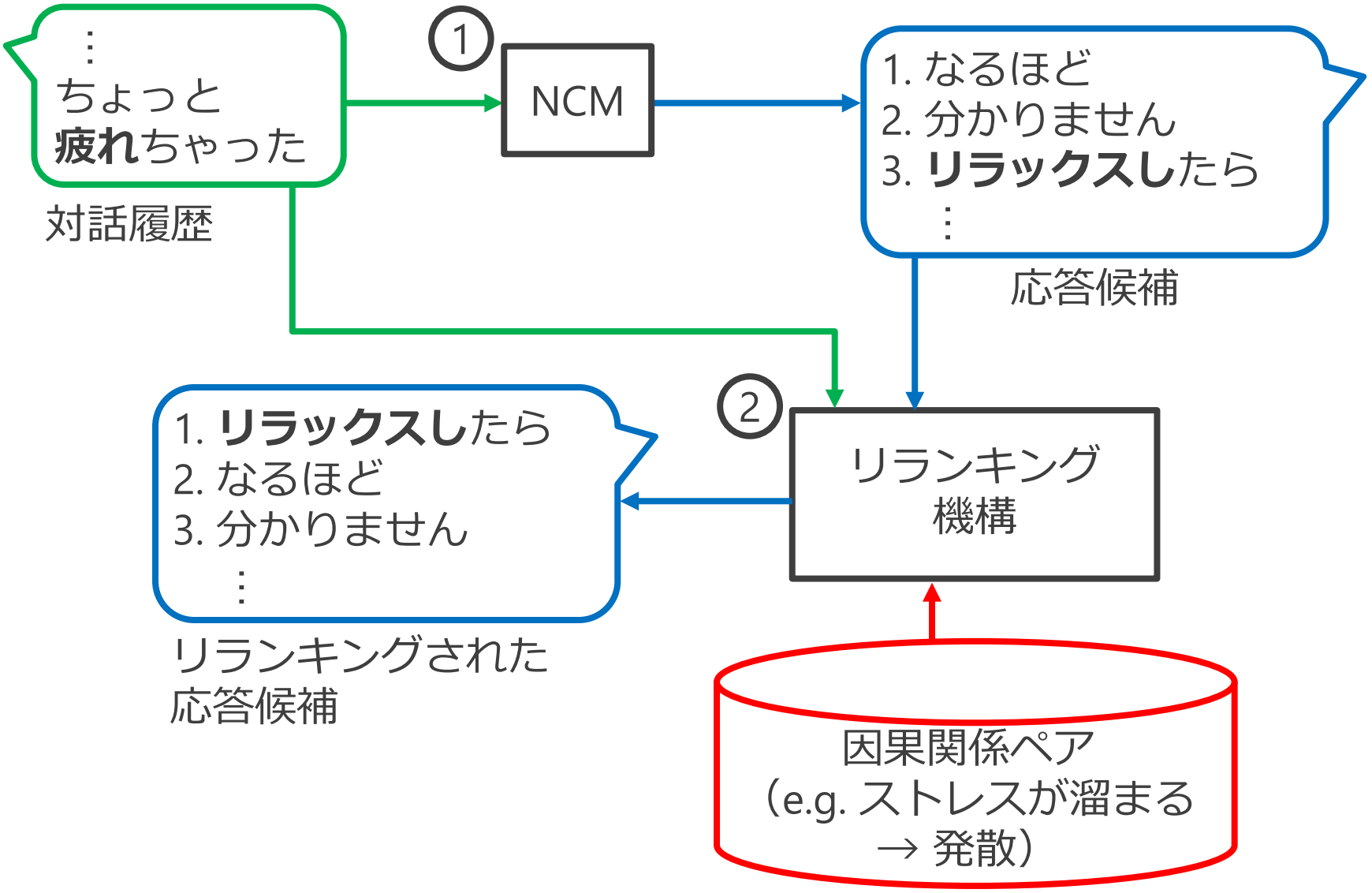
- Improving Long Distance Slot Carryover in Spoken Dialogue Systems
  by *Tongfei Chen, Chetan Naik, Hua He, Pushpendre Rastogi and Lambert Mathias*

Cross-submissions

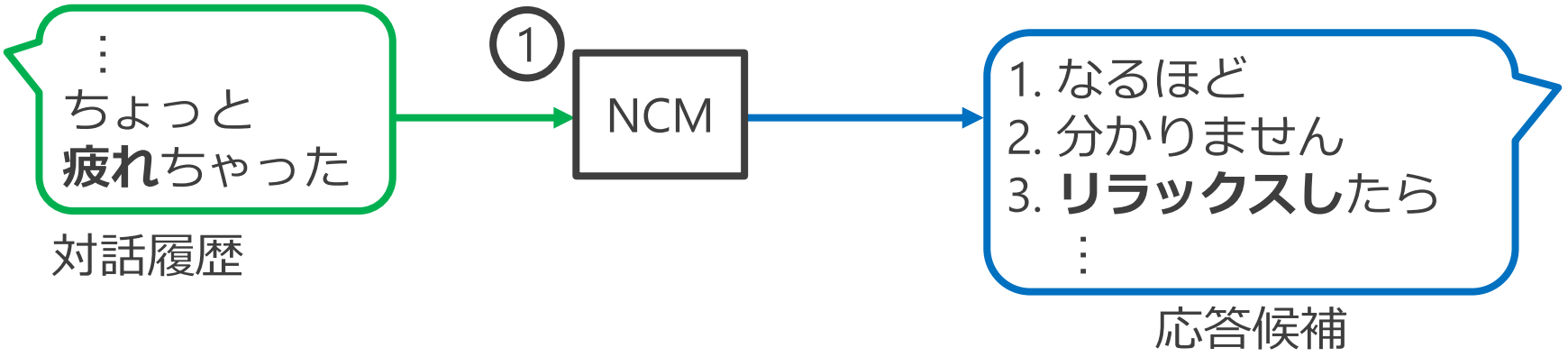
- Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems
  by *Chien Sheng Wu, Andrea Madotto, Ehsan Hosseiniasl, Caiming Xiong, Richard Socher and Pascale Fung*

Conversational Response Re-ranking Based on Event Causality and Role Factored Tensor Event Embedding

因果関係に基づく雑談対話応答のリランキング



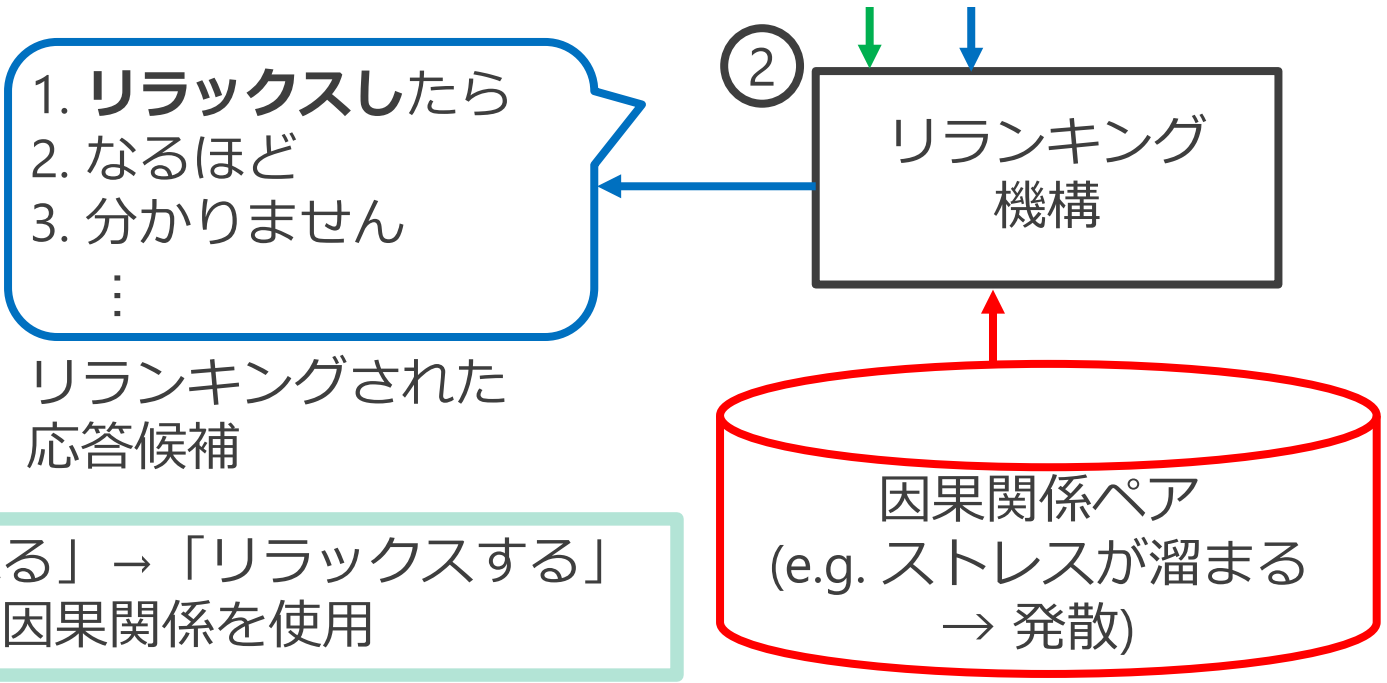
応答候補の生成



対話履歴から応答候補を生成

因果関係に基づくリランキング

対話履歴との間に因果関係を持つ
応答候補を高い順位にリランキング



「疲れる」→「リラックスする」
という因果関係を使用

因果関係ペア

各事態は述語項構造を用いて表現

述語：必須，格要素：持たない場合も

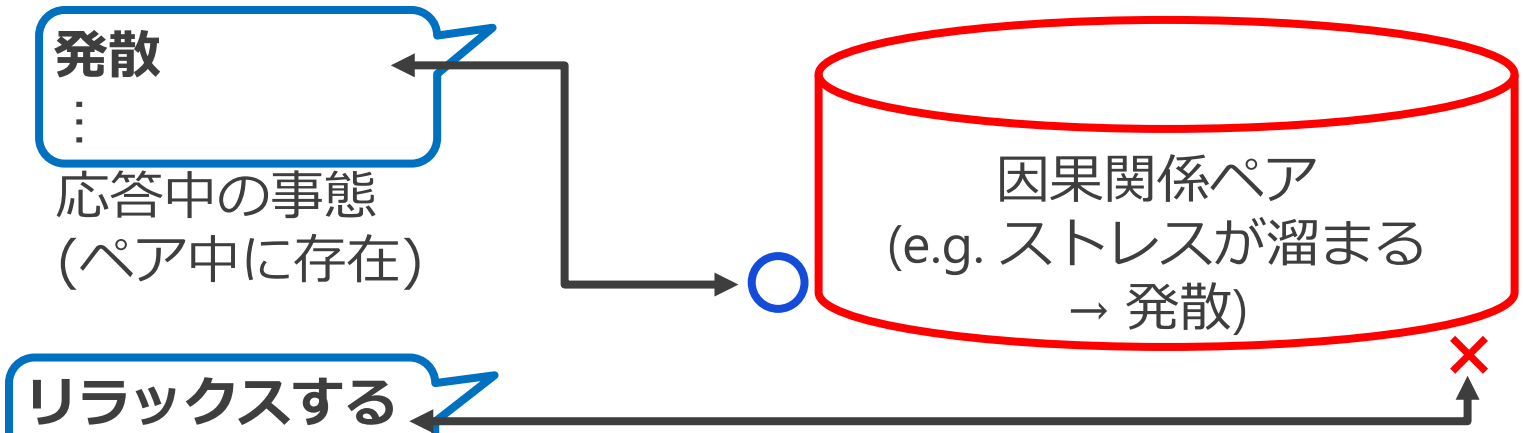
因果関係			
原因		結果	
述語	格要素	述語	格要素
溜まる	ストレス	発散	-

因果関係ペアを用いて対話履歴と応答候補の間に存在する因果関係を探索

因果関係ペアのカバレージ問題

限られた Web コーパスから構築されているため

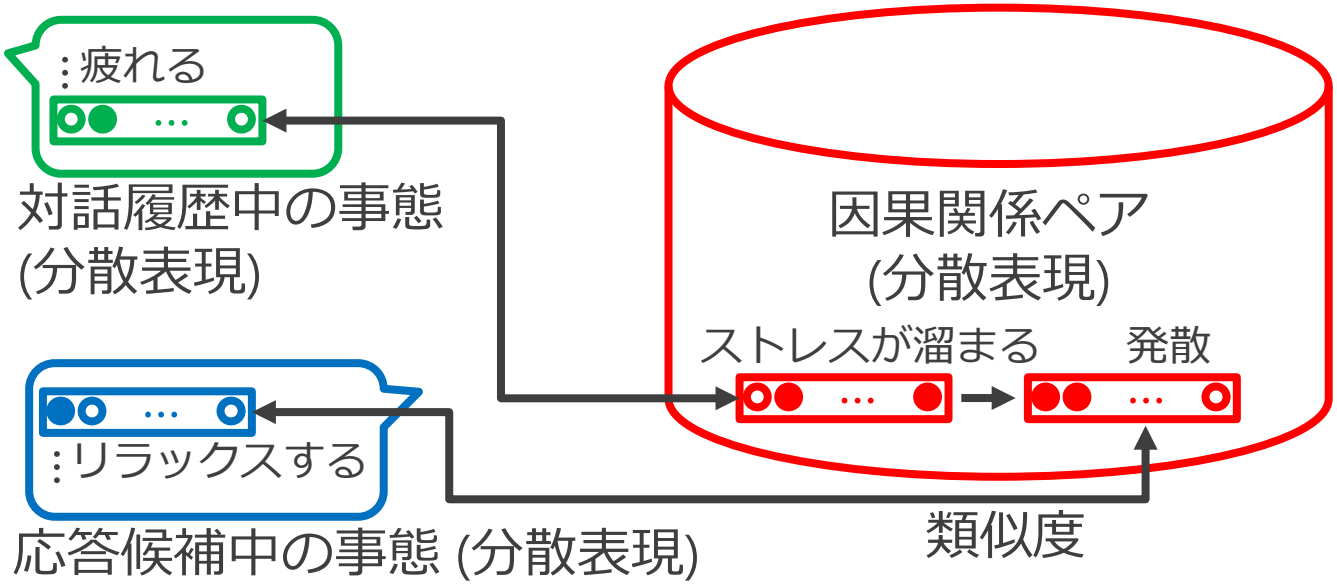
因果関係ペアは対話中の全ての因果関係を網羅していない



「リラックスする」という事態は
ペア中に存在しないため考慮できない

事態分散表現に基づくマッチング

類似した因果関係をベクトル空間上で探索



対話中の「疲れる」→「リラックスする」という
因果関係は「ストレスが溜まる」→「発散」という
類似した因果関係がペア中に含まれるため考慮可能

リランキングの例

対話1：

ユーザ：もう不安なことが多すぎて**ストレスが溜ま**ってく

システム (1-best)：大丈夫ですか

システム (Re-ranked)：大丈夫ですか**無理し**ないでくださいね

「無理をする」→「ストレスが溜まる」という妥当な因果関係により一貫した応答を選択

Improving Long Distance Slot Carryover in Spoken Dialogue Systems

概要

Dialogue State Tracking（DST）において
Slot carryover（過去の対話履歴中で確定したスロットの値を保持しておくこと）の精度を上昇

モデルは Pointer Networks ベース，Transformer ベースを比較

データセットは内部で集めたものと DSTC2

Figure 1: An example of a conversation session. Slots are listed on the right. Related slots often co-occur, such as (1) [WEATHERCITY: *San Francisco*] and [WEATHERSTATE: *CA*], and should be carried over together due to their interdependencies (2) PLACE slot is often seen to occur along with TOWN.

Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems

概要

同じく DST を対象にした研究

マルチドメインに対応可能な点をアピール

モデルは Pointer Networks ベース，著者らは TRADE と呼んでいる

データセットは MultiWOZ と DSTC2

ACL best paper 候補にもなった論文

確かにとても読みやすくお手本にしたいと感じた

SIGdial

SIGdial conference@sigdial.org

SIGDIAL 2019

松山 洋一 matsuyama@pcl.cs.waseda.ac.jp

早稲田大学GCS研究機構 知覚情報システム研究所主任研究員（研究院准教授）

2019.12.3

Stockholm


7
Djurgården
Waldemarsudde
SL
3


SÖDERMALM
SERGELS TORG
E4 E18 E20
STUREPLAN
Premiär!
Premiär!

KTH Royal Institute of Technology in Stockholm

Special Interest Group (SIG) of the Association for Computational Linguistics (ACL)

Special Interest Group (SIG) of the International Speech Communication (ISCA)

Keynote Speech **Dan Bohus** (Microsoft Research)

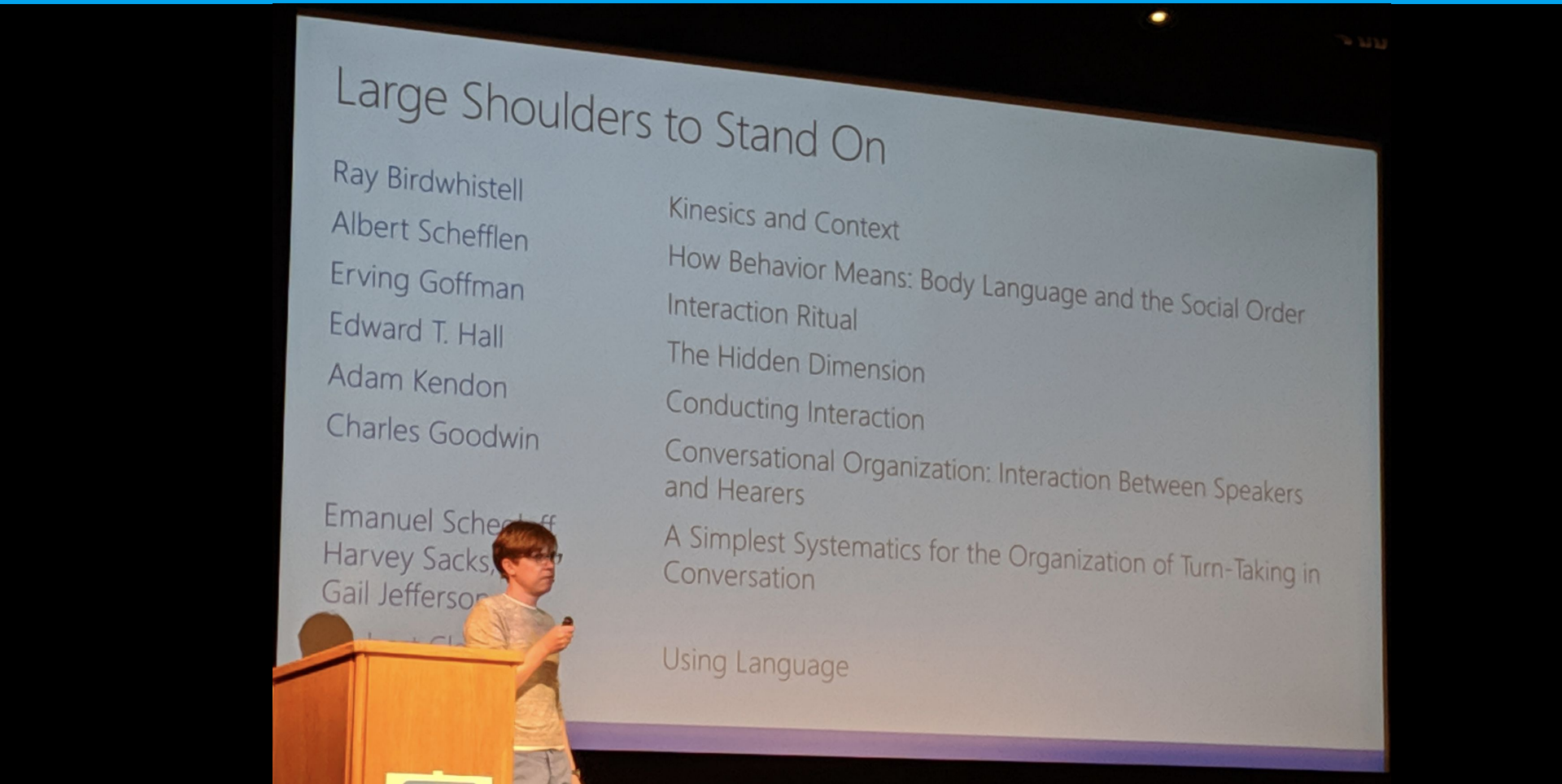
SIGDIAL

Special Interest Group (SIG) of the Association for Computational Linguistics (ACL)
Special Interest Group (SIG) of the International Speech Communication (ISCA)

メインテーマ

対話システム
談話処理
対話コーパス収集・分析

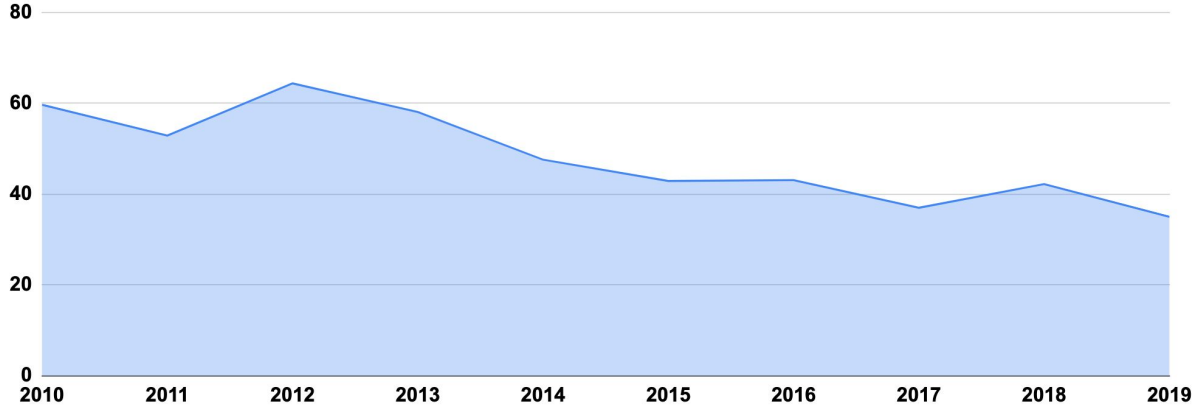
Look Back Human Communication Dynamics Research



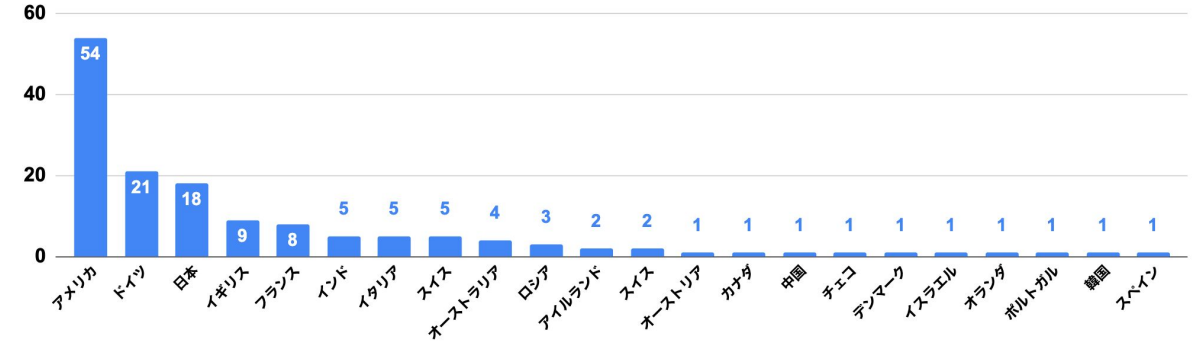
Young Researchers' Round Table on Spoken Dialogue Systems (YRRSDS)

投稿数・採択率

- 投稿数：146（96 Long, 43 Short, 10 Demo）
- 採択率：35%（採択数 51）
  - 2018：42.2%
  - 2017：37.0%
  - 2016：43.1%
  - 2015：42.9%
  - 2014：47.6%
  - 2013：58.1%
  - 2012：64.4%
  - 2011：52.9%
  - 2010：59.7%

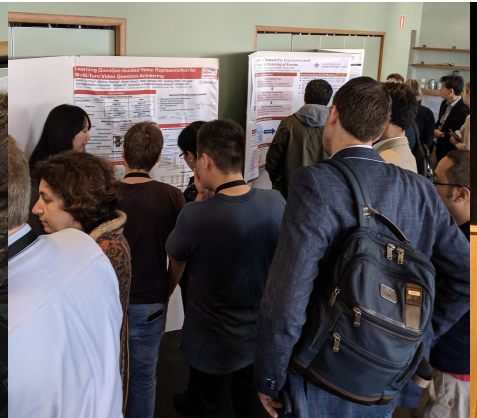


- 国別投稿数
  - 日本は投稿数3位
  - <u>しかし採択数はゼロ！</u>



採択テーマ（オーラルのみ）

1. **Policy and Knowledge**：2件
2. **Implications of Deep Learning for Dialogue Modeling**：3件
3. **Generation and End-to-end Dialogue Systems**：3件
4. **Understanding and Dialogue State Tracking**：2件
5. **Acoustics**：2件
6. **Evaluation and Data**：3件
7. **Discourse**：3件



Best Papers

Structured Fusion Networks for Dialog

Shikib Mehri, Tejas Srinivasan and Maxine Eskenazi
Carnegie Mellon University

伝統的な**パイプライン方式**（言語理解・対話制御・言語生成）の構造を明示的に**End-to-End方式**のアーキテクチャに組み込むハイブリッドな対話システムを提案.

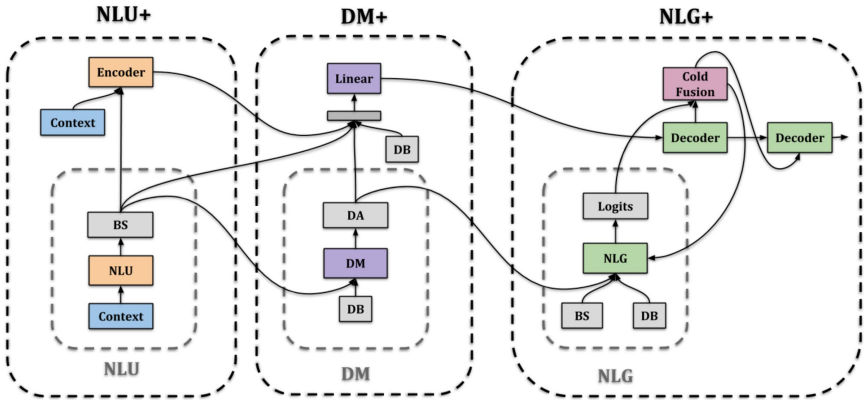


Figure 5: The Structured Fusion Network. The grey dashed boxes correspond to the pre-trained neural dialog modules. A higher-level is learned on top of the pre-trained modules, as a mechanism of enforcing structure in the end-to-end model.

Deep Reinforcement Learning For Modeling Chit-Chat Dialog with Discrete Attributes

Chinnadhurai Sankar and Sujith Ravi
Google Research

文脈情報と強化学習よって選択される対話行為・感情クラスを条件として，返答をseq2seqによって生成

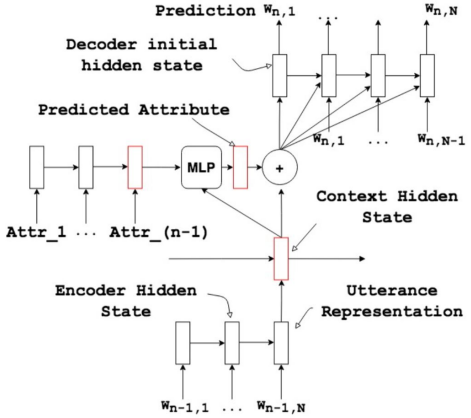


Figure 2: Attribute Conditional HRED : Token generation is additionally conditioned on the predicted dialog attributes. The dialog attribute's embedding is concatenated with the context vector.

SIGDIAL 2020

- 日時：2020年7月1日～3日
- 場所：ボイシー（Boise），米国アイダホ州
  - Jack’s Urban Meeting Place (JUMP)
  - ACL 2020（シアトル，7月5日～10日）と共催
- 論文締め切り：2020年3月6日

メンタリング・サービス
Acceptable submissions that require language (English) or organizational assistance will be flagged for mentoring, and accepted with a recommendation to revise with the help of a mentor.

TOSHIBA

INTERSPEECH2019 参加報告

東芝　小林　優佳

2019/12/03

INTERSPEECH2019とは

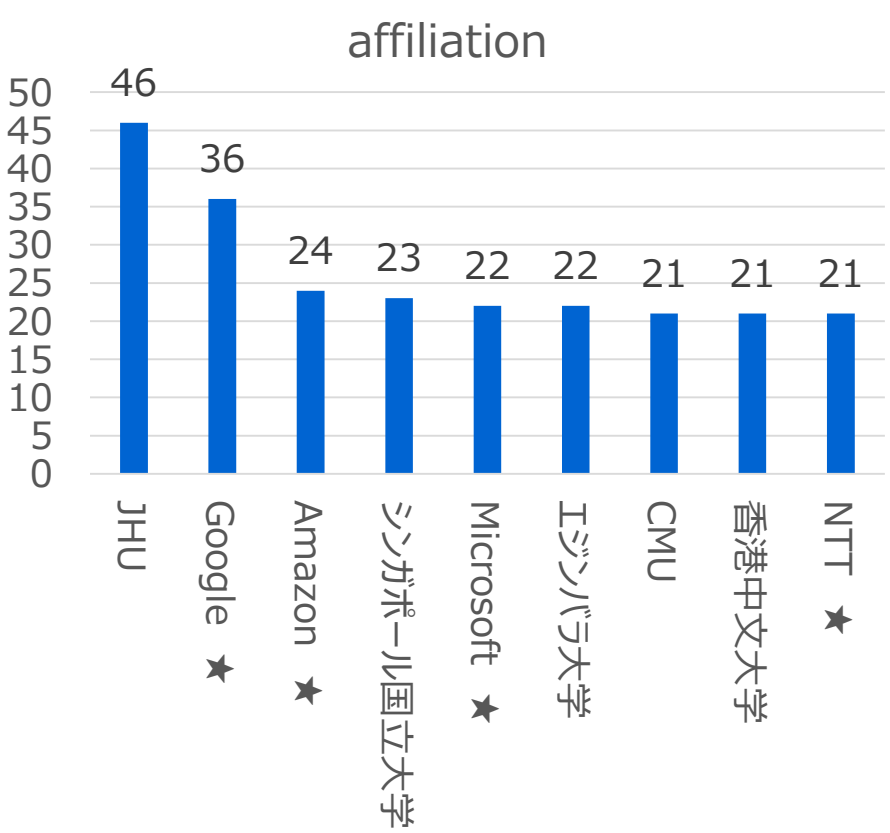
ISCA (International speech communication association)主催の学会
音声認識、音声合成などがメイン
最近は音声対話に関係する発表も増えている
開催地：グラーツ（オーストリア）
日程：2019年9月15日～19日（5日間）
　　初日はチュートリアルのみ
参加人数：約2000人
採択率：914/1855＝49.2%

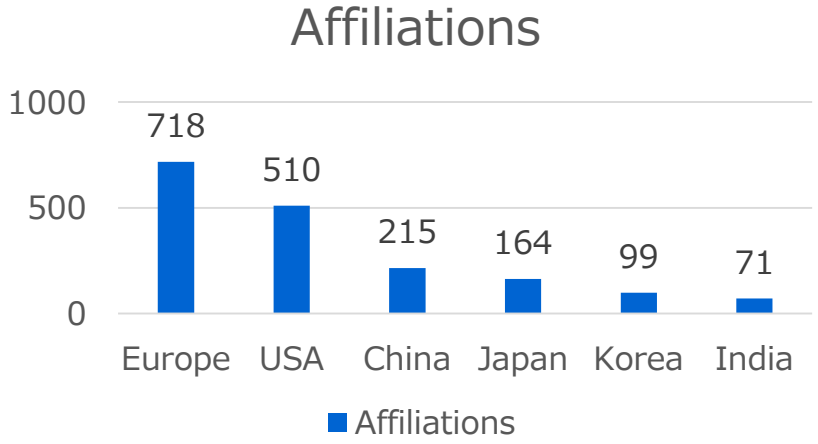
会場の様子

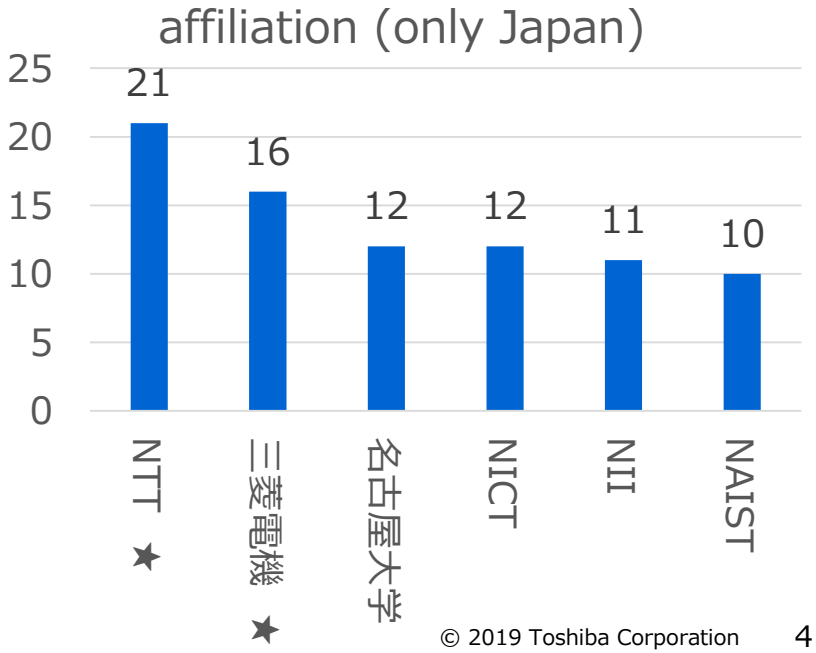
スポンサー

Founding sponser : amazon alexa
Platinum sponser : Apple
Diamond sponser : DiDi, ASAPP, facebook
Gold sponser : Google, DataBaker, Alibaba Group, Microsoft, NUANCE, NAVER LINE
Silver sponser : IBM Research AI, Yahoo! JAPAN

発表件数ランキング







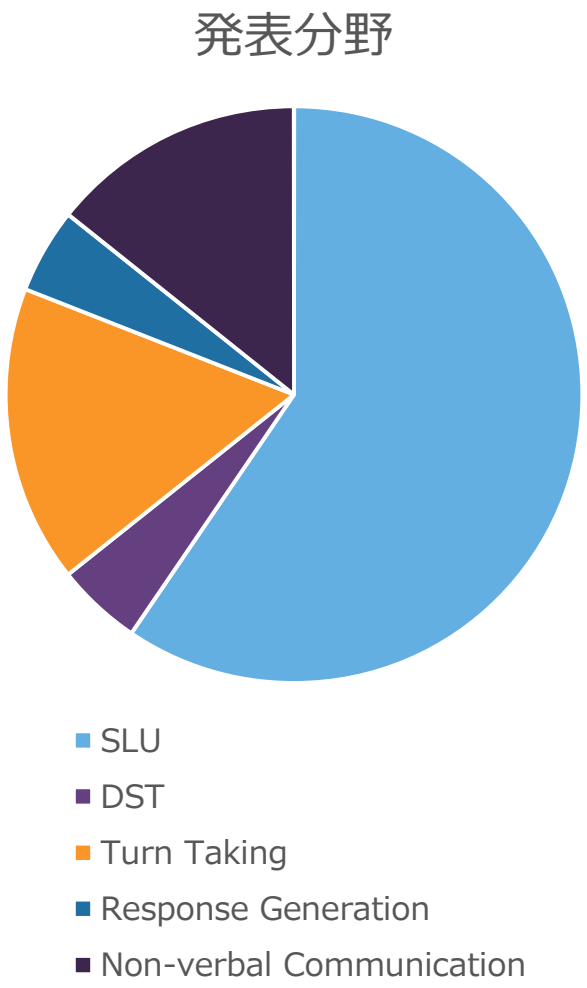
企業ではGoogle, Amazon, Microsoft, NTTが多い
国内ではNTT、三菱電機が多い

セッション振り分け

音声認識のセッションも多いが、それ以外の音声の解析に関するセッションが多い

Speech Perception and Production	7
Phonetics, Phinology, and Prosody	5
Paralinguistic Analysis	8
Speaker and Language Identiciation	8
Analysis of Speech and Audio Signals	12
Speech Coding Enhancement	7
Speech Synthesis	9
Speech Recognition 1:Signal Processing	14
Speech Recognition 2:Architecture	4
Speech Recognition 3:Speech Recognition 3:New Applications	5
Spoken Dialog Systems	6
Spoken Language Processing	6

音声対話領域の内訳



SLUが断然多い
純粋なNLPではなく、
音声認識とend2endで実施するもの
音声認識誤りを考慮したもの
音響特徴量を使用したもの
などが多い
応答生成、対話制御などはほとんどない

キーワード
end2end, GAN, triplet loss, CCA, ASR error, low resource, phoneme, grapheme, Out-of-vocabulary, multi-lingual, BERT, subword, transfer learning

ピックアップ論文(1) SLU,DST

- **Speech Model Pre-training for End-to-End Spoken Language Understanding**

ASRとSLUをend2endで行う。ASR部分はpre-trainingしておく。学習時にword layerだけ再学習すると全体を再学習するよりも性能がいい。

- **Iterative Delexicalization for Improved Spoken Language Understanding**

学習データの中にvalueの数が多いslot-valueがある場合、スロット名に置き換えてマスクして学習する。テスト時は発話文からslot-valueを検出し、検出した箇所をスロット名に置き換えてSLUを行う。

- **HyST: A Hybrid Approach for Flexible and Accurate Dialogue State Tracking**

valueリストを使わないDSTと使うDSTのハイブリッドモデル。valueリストを使わないDSTのおかげで未知語に強い。

- **Mining Polysemous Triplets with Recurrent Neural Networks for Spoken Language Understanding**

同じvalueに複数のslotが付与される場合がある。triplet lossを使用し、正解が複数あるという状況をうまくモデルにする。

ピックアップ論文(2)　NLG, Non-verbal

- **Personalized Dialogue Response Generation Learned from Monologues**

GANでend2endの音声対話システムを生成。generatorはユーザ発話を入力するとシステム応答を出力する。reconstructorはシステム応答を入力するとユーザ発話を推定する。推定ユーザ発話とユーザ発話の差分をlossとして学習することで入力に即した応答を生成できるようになる。対話コーパスと話者別対話コーパスを使用して学習する。discriminatorはgeneratorによって生成された応答か人間の応答かを判別する。モノローグのコーパスを使用して学習する。

- **The greennn tree - lengthening position influences uncertainty perception**

人間は自信がないときに単語を伸ばして発話することがある。システムも同じように単語を伸ばして発話させて印象の違いを評価する。単語の前半を伸ばす場合、後半を伸ばす場合の違いを評価。

- **Mirroring to Build Trust in Digital Assistants**

ユーザのおしゃべり度合いに応じてシステムのおしゃべり度合いも変更して応答する。おしゃべりかどうかは発話の情報密度で測る。

所感

- 純粋な言語処理・対話処理よりは音声をからめた方が通りやすそう
- システムの出力の制御よりはシステムへの入力の理解の方が多そう
- 何語で研究するか？

  言語に依存しない技術を提案して**英語**の**公開コーパス**で評価するのが王道
  それ以外なら、中国語、話者が数百人しかいない言語、言語間のモデルの転用など
  日本語固有の技術、日本語のみで評価した技術は若干不利かも

次回のINTERSPEECH

開催地：上海インターナショナルコンベンションセンター
投稿締め切り日：2020/3/30
開催日：2020/9/14-18

採録されるには・・・新規性が重要！
他の学会ほど再現性は
重視されていない

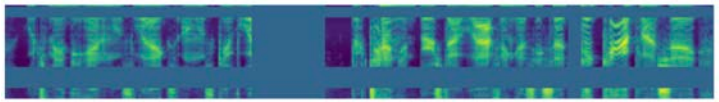
chance of being accepted if receiving the tops score in

novelty 77%
clarity/correctness 71%
reproducibility 60%
overall recommendation 90%

ピックアップ論文: 音声認識①

「SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition」（Google）

- 入力メルスペクトログラムに対して，周波数方向と時間方向にそれぞれマスクをかけるdata augmentation（regularization）
- 大きなモデルを長く学習（600epochs/24days）
- Encoder-decoderモデルのDecoder側の改善



「Forget a Bit to Learn Better: Soft Forgetting for CTC-Based Automatic Speech Recognition」（IBM）

- SpecAugmentと類似したアイデア，CTCに適用+Twin regularization

「Joint Speech Recognition and Speaker Diarization via Sequence Transduction」（Google）

- RNN transducerでASRとspeaker diarizationを同時最適化
- 医者と患者の対話
- 話者のラベルと書き起こしを交互に出力するように学習

hello dr jekyll <spk:pt> hello mr hyde what brings you here today <spk:dr> I am struggling again with my bipolar disorder <spk:pt>

ピックアップ論文: 音声認識②

「Cross-Attention End-to-End ASR for Two-Party Conversations」（CMU）

- 2話者間の履歴を考慮したEnd-to-End音声認識
- 話者ごとの発話履歴をBERTでエンコードしてattention（ACL2019の拡張）

「Language Modeling with Deep Transformers」（RWTH, best student paper）

- 深いTransformer言語モデル（64層！）を音声認識のリスコアリングに適用
- インクリメンタルにエンコードするのでpositional encodingは不要
- RWTHのLibrispeech state-of-the-artモデルに大きく貢献

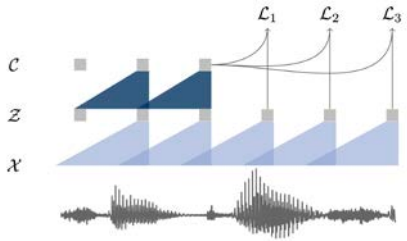
「Personalizing ASR for Dysarthric and Accented Speech with Limited Data」（Google）

- 構音障害者の音声認識のためのモデル適用
- RNN-T/LASをLibrispeechで学習し，少量のデータに適用
- テクニカルな新規性はないが，重要な問題

ピックアップ論文: 音声認識のための教師なし事前学習

「wav2vec: Unsupervised Pre-training for Speech Recognition」（Facebook）

- Contrastive predictive coding (CPC）による，音声（波形）を使った教師なし識別的事前学習
- 未来の音声情報を予測（相互情報量を最大化）するように特徴量エンコーダ（CNN）を学習
- 未来の音声フレームの正例と負例を識別するcontrastive loss
- 事前学習した特徴量をフィルタバンクの代わりに使う



「An Unsupervised Autoregressive Model for Speech Representation Leaning」（MIT）

・Autoregressive predictive coding (APC）による音声（スペクトログラム）を使った教師なし生成的事前学習

・現在までのコンテキストから未来のフレームを予測してL1ロスを最小化（回帰）

・phone classification，speaker verificationでCPCを上回る

・下層は話者情報，上層は音素情報を捉えている

・拡張版（ICASSP2020投稿中）ではASR，speech translation, speaker identificationでCPCを上回る

音声認識の動向

壮絶なLibrispeech（1000h）での精度競争

- 「RWTH ASR Systems for LibriSpeech: Hybrid vs Attention」
- test-clean/otherのSOTAはWER2.3/5.0%
- SpecAugmentによりE2Eの精度が格段に改善（2.5/5.8%）
- ASRU2019ではTransformerで2.6/5.7%（Karita et al.）
- 精度だけで勝負するのは危険だが，最新の論文を追うのは重要

NLPに続いて音声を用いたunsupervised pre-training

- 昨年のトレンドはTTS，オートエンコーダ，cycle consistencyによるデータ拡張
- 今年はCPC，APC
- 次はBERTのspeech版（ICASSP2020投稿中の論文だけですでに4件）

ピックアップ論文: speech2speech

「Direct Speech-to-Speech Translation with a Sequence-to-Sequence Model」（Google）

- ソース言語の音声（スペイン語）からターゲット言語への音声（英語）へのマッピングをend-to-endで学習
- ASR，MT，TTSを一つのモデルに（proof-of-concept）
- スペイン語（ASR），英語（ST）の音素ラベルを予測するタスクとマルチタスク学習（必須）
- E2E音声翻訳->TTSには負ける
- まだソース側の声質をターゲット側にうまく反映できてない

「Parrotron: An End-to-End Speech-to-Speech Conversion Model and its Applications to Hearing-Impaired Speech and Speech Separation」（Google）

- End-to-endのmany-to-one音声変換
- 聴覚障がい者の音声を聞きやすいように変換
- 音源分離にも適用

所感

- 新しいコンセプトを示すような論文がインパクトがあった
  - Speech-to-speech translation
  - ASR + speaker diarization
- SpecAugmentによりハイブリッドモデルとの差は今後もさらに縮まると思われる
- SpecAugmentはBERTやMASSなどに似ており，応用の幅がかなりありそう
- NLPの進展に続いて教師なし事前学習は今後流行っていくと思われる

学生イベント

Student reception

- 学生なら知り合いがいなくても参加した方がよい（しかも無料！）
- 個人的には2回目の参加だが，どちらも良かった
- 日本人で群れるのは良くない

企業イベント

- 海外の学生，リサーチャーと仲良くなるチャンス！
- 個人的にはFacebook，Microsoftのイベントに参加
- Facebookのイベントはstudent receptionでできた友達に紹介してもらい，Facebookのイベントでさらに友達ができた
- インターン先を見つけたい場合，会場の企業ブースにはリサーチャーはあまりいないので，このようなイベントに行けばゆっくり話してアピールできるチャンスあり

国際会議報告
ICMI 2019

田中 宏季
AHC-Lab., NAIST

基本情報 (1/2)

- ACM International Conference on Multimodal Interaction (ICMI) 2019

- 開催地
  - Suzhou, Jiangsu, China
- 期間
  - October 14-18, 2019
- 参加人数
  - 200人ほど
  - 日本からは20人ほど
- 採択率
  - 138 long and short paper submissions (89 long and 49 short)
  - Overall acceptance rate: 36%

基本情報 (2/2)

- 対象
  - Multidisciplinary research on multimodal human-human and human-computer interaction, interfaces, and system development
- 発表の形態
  - オーラル、ポスター、デモ
  - シングルトラック、６セッション
- ワークショップ：2種類
- チュートリアル
  - Getting Virtually Personal: Power Conversational AI to Fulfill Tasks and Personalize Chitchat for Real-World Applications by Michelle Zhou
  - Spoken Dialogue Processing for Multimodal Human-Robot Interaction by Tatsuya Kawahara
- グランドチャレンジ：感情、言語理解

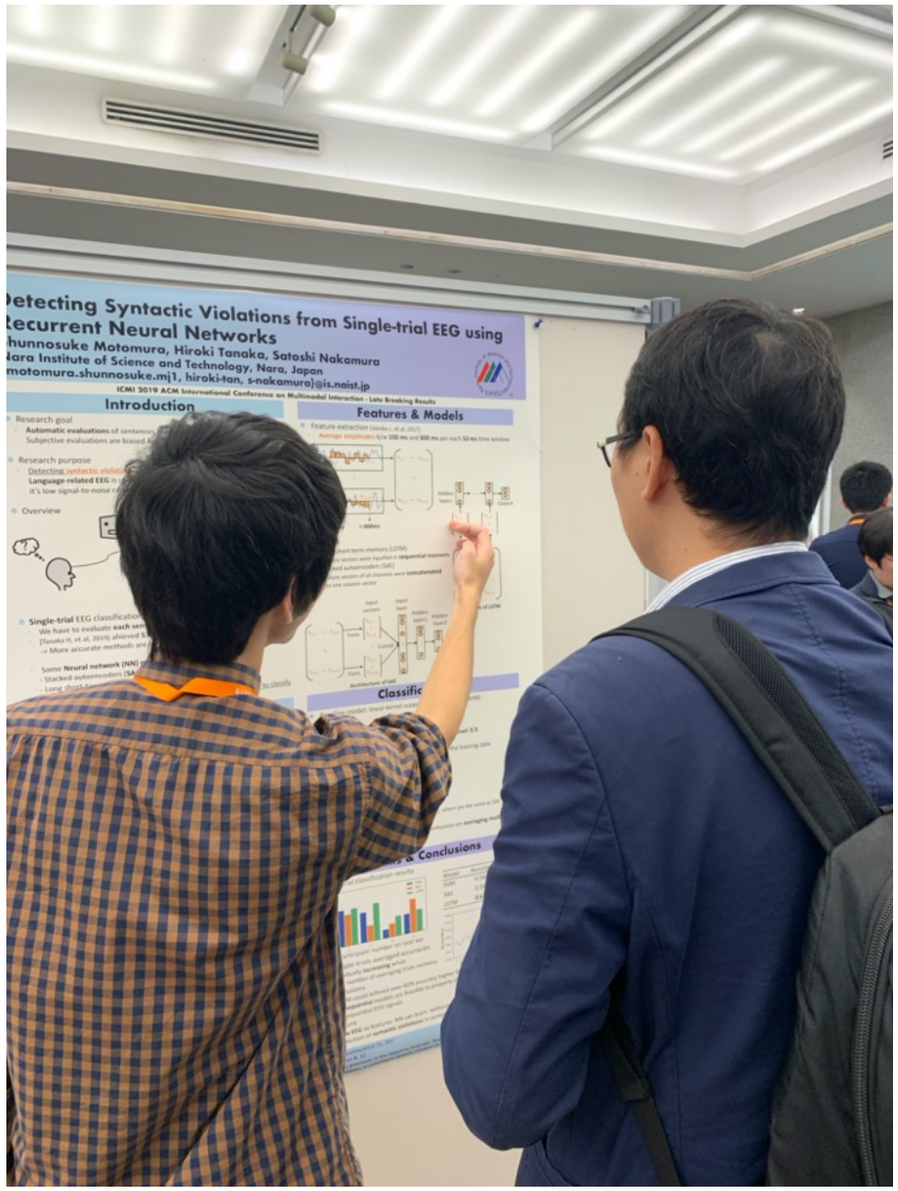
対話系の発表

- 会議全体に対する割合
  - 3-4割
  - 対話システムや検出、フィードバックなど

- 採択傾向
  - マルチモーダル（聴覚、視覚、触覚、など）
  - タスク（疾患、マルチパーティ、スキル検出）
  - 機械学習、深層学習によりディテクション
  - 認知モデルなど

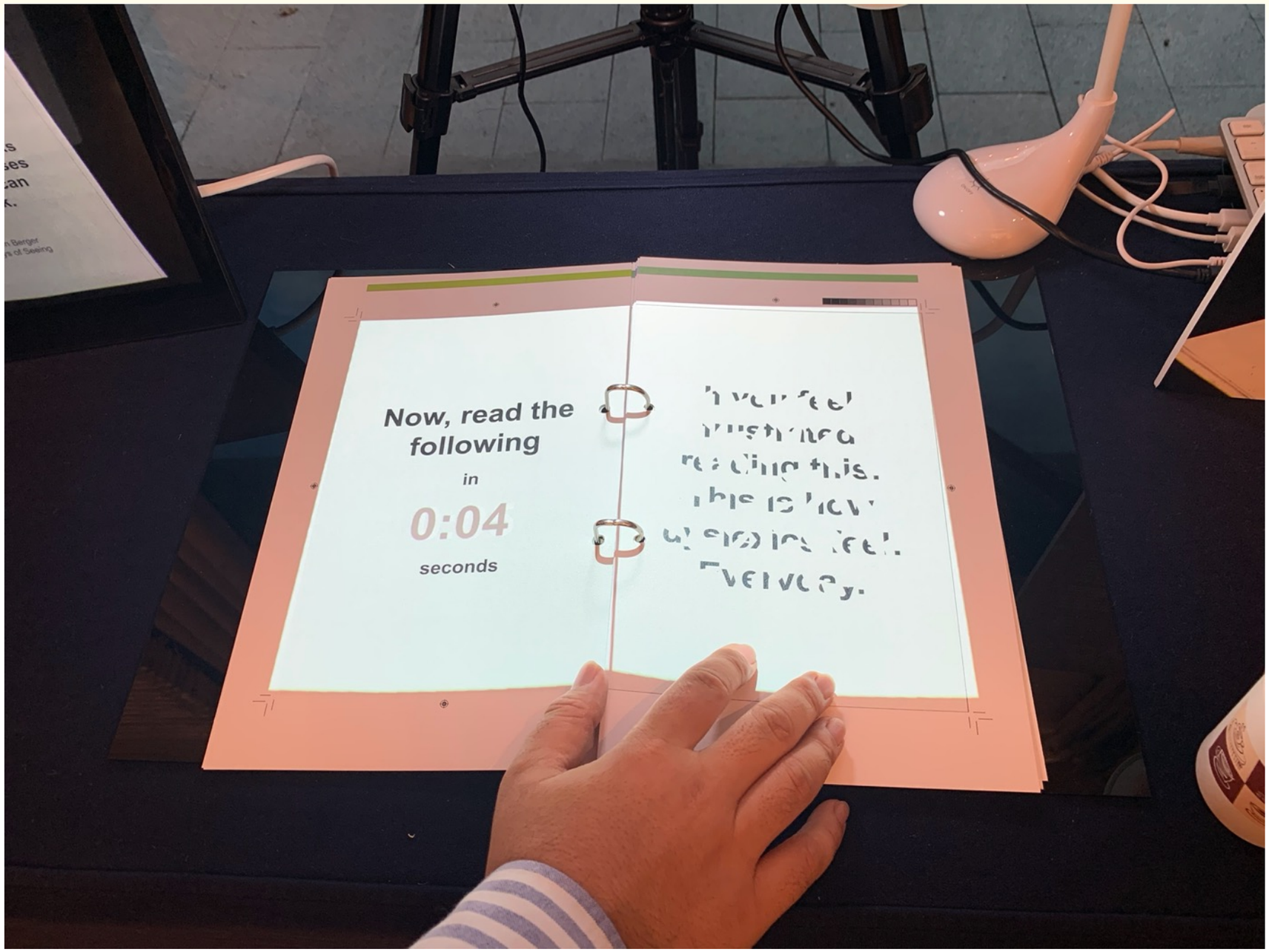
会議の様子 (1/3)

会議の様子 (2/3)


Detecting Syntactic Violations from Single-trial EEG using Recurrent Neural Networks
Introduction
Features & Models


ICMI 2019
21st ACM INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION
2019.10.14-10.18 SUZHOU,CHINA
SIGCHI

会議の様子 (3/3)



研究例 (1/2)

- Session 1: Human Behavior
  - Multi-modal Active Learning From Human Data: A Deep Reinforcement Learning Approach [Rudovic et al., 2019]

- Session 2: Artificial Agents
  - Multitask Prediction of Exchange-Level Annotations for Multimodal Dialogue Systems [Hirano et al., 2019]

- Session 3: Touch and Gesture
  - Dynamic Adaptive Gesturing Predicts Domain Expertise in Mathematics [Sriramulu et al., 2019]

研究例 (2/2)

- Session 4: Physiological Modeling
  - Multimodal Classification of EEG During Physical Activity [Ding et al., 2019]

- Session 5: Sound and Interaction
  - Towards Automatic Detection of Misinformation in Online Medical Videos [Hou et al., 2019]

- Session 6: Multiparty Interaction
  - A Multimodal Robot-Driven Meeting Facilitation System for Group Decision-Making Sessions [Shamekhi et al, 2019]

ICMIへの感想

- 「人間」マルチモーダルに焦点、どう多元情報を統合する
- 学際研究：対話にとどまらない、インタラクション (VR/プロジェクションマッピング)
- 違い: ACII; 感情、IVA; エージェント、SigDial; 対話
- 幅広く受け入れられる、日本からも多め
- まだわかっていない研究できることは沢山
- 認知科学、対話、マルチモーダル、にとっては、適切だと思う

来年度の情報

- ICMI 2020
  - Utrecht, the Netherlands
  - October 25-29, 2020
  - 投稿締切: 5月初旬頃
  - Long paper (8 pages) , Short paper (4 pages)

http://icmi.acm.org/2020/