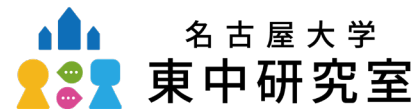


国際会議参加報告

SIGDIAL/YRRSDS 2022

名古屋大学大学院 情報学研究科 修士2年

大橋厚元





The 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2022) September 7-9, Edinburgh, UK



Photo by Connor Mollison

SIGDIALの基本情報

- 談話や対話に関する国際会議
- **SIGDIAL 2022 @Edinburgh, UK**
 - 開催場所 | Heriot-Watt大学
 - 開催形式 | 現地+リモートのハイブリッド
 - 登録者数 | 現地138, リモート168
 - 発表件数 | 口頭21, ポスター+デモ17, ビデオ28
 - **Special Session** | Natural Language in Human Robot Interaction (NLiHRI)
- 来年以降の情報
 - SIGDIAL 2023@(アメリカのどこか), SIGDIAL 2024@京都



統計

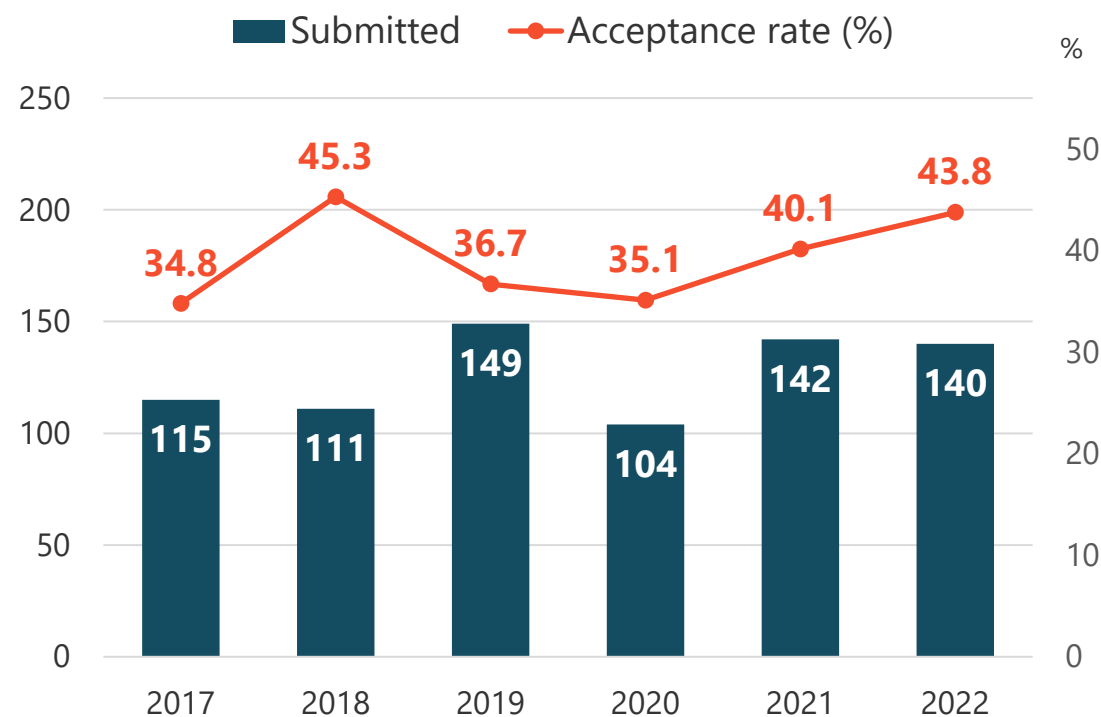
● 今年

	投稿数	採択数 (採択率)
Long	79	37 (46.8%)
Short	49	19 (38.8%)
Demo	12	8 (66.7%)
Total	140	64 (43.8%)

● ここ数年

- 投稿数, 採択率ともに大きな変動は無し

投稿数と採択率の推移 (直近6年)



Keynotes



<https://2022.sigdial.org/speaker/vivian-chen/>

Yun-Nung (Vivian) Chen, National Taiwan University

Robustness, Scalability, and Practicality of Conversational AI

- ✓ タスク指向対話システムの実用性向上に向けた要素技術を紹介
(例: ASRのエラーにロバストなNLU, DSTのためのデータ拡張手法)



<https://2022.sigdial.org/speaker/angeliki-lazaridou/>

Angeliki Lazaridou, DeepMind

On opportunities and challenges on communicating using Large LMs

- ✓ Large LMは強力だが、学習データが古いと悪影響を及ぼす場合があることを紹介
- ✓ スケーラブルに最新データをモデルに導入することが課題 (例: Knowledge editの活用)



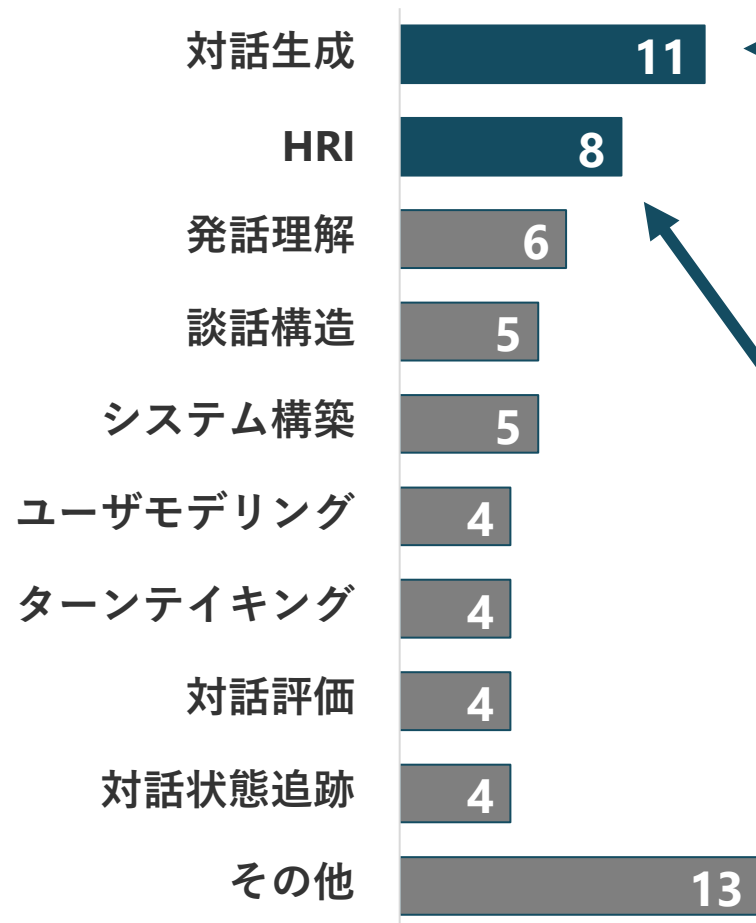
<https://2022.sigdial.org/speaker/giuseppe-carenini/>

Giuseppe Carenini, University of British Columbia

Unlimited discourse structures in the era of distant supervision, pre-trained language models and autoencoders

- ✓ 最新の談話解析コーパス自動生成手法 (例: 距離学習, BERT/BARTベース) を紹介
- ✓ 今後の談話構造の方向性を考察 (例: 離散表現→連続表現, 木構造→グラフ構造)

発表内容の傾向



● 大規模言語モデルを活用した対話システム研究が多数

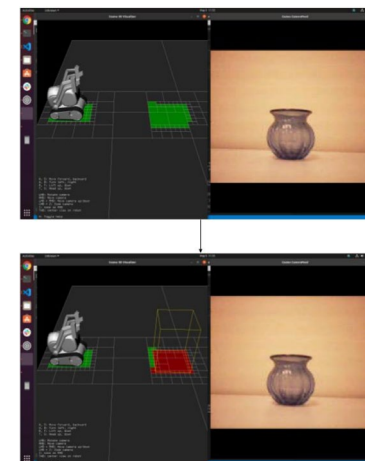
- ✓ ドキュメント情報・知識の導入 (Xue et al., Wu et al.)
- ✓ Few-shot / zero-shotのための、対話データ生成・拡張 (Lin et al., Mehri et al.)

今後は**Few-shot / zero-shot** 対話生成に向け、
活発に**超大規模言語モデル (GPT-3等)**が使用されそう

● 実世界指向の高まり

- ✓ Embodied agentを用いたタスク
例：料理 (Gella et al.) / 積み木 (Younes et al.)
- ✓ 記号接地を用いた言語獲得 (Torres et al.)

テキストだけでなく、**画像や物理情報など**
を扱った**HRI分野の研究が増加している**



Torres et al., 2022, Symbol and Communicative Grounding through Object Permanence with a Mobile Robot, Figure 1

Best Paper Nominations

ベストペーパー

- **Reducing Model Churn: Stable Re-training of Conversational Agents** 🏆

- ✓ 深層学習モデルにおける *churn* (パフォーマンスのばらつき) の削減手法を提案

- **How Much Does Prosody Help Turn-taking? Investigations using Voice Activity Projection Models** 🏆

- ✓ ターンテイキングのモデル化において韻律特徴がどの程度重要であるかを調査

- **EDU-AP: Elementary Discourse Unit based Argument Parser**

- ✓ 従来のトークンレベル構造解析器における局所ミスを解消するため、EDU*レベルのモデルを提案

*Elementary Discourse Unit: ある単一の意味を表す、文節やフレーズのこと (単語以上文未満)

- **Redwood: Using Collision Detection to Grow a Large-Scale Intent Classification Dataset**

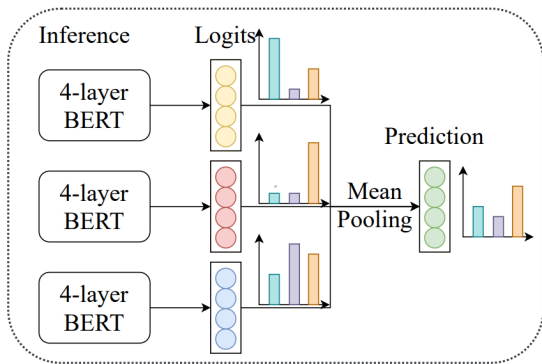
- ✓ 複数の意図分類データセットを結合する際、意味的に重複した意図が混在してしまうと、モデルパフォーマンス悪化の原因となる。自動的に意図の重複を検出するタスクを提案。

研究紹介

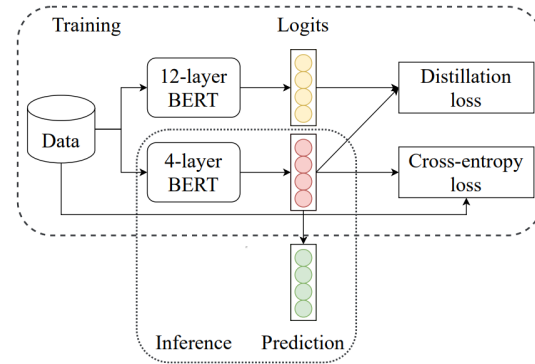
Reducing Model Churn: Stable Re-training of Conversational Agents



Christopher Hidey, Fei Liu, Rahul Goel (Google)



(a) Ensemble.



(b) Soft distillation.

Model	TOP		TOPv2		MTOp	
	EM (@10)	AGR	EM (@10)	AGR	EM (@10)	AGR
BERT-4	80.65 (70.29)	75.48	83.88 (73.12)	78.15	79.31 (69.04)	73.64
Ensemble	84.60 (78.55)	86.18	86.42 (80.38)	88.17	84.59 (78.52)	84.39
SD (ensemble)	81.20 (70.80)	76.16	84.00 (73.47)	78.75	79.29 (67.40)	71.38
SD (BERT-12)	80.93 (71.14)	76.80	84.12 (73.87)	79.02	79.23 (68.71)	73.23
HD (BERT-12)	80.72 (70.01)	75.03	83.84 (72.57)	77.37	78.96 (68.61)	73.07
Co-distillation	81.43 (73.56)	80.41	84.21 (76.10)	82.99	79.45 (69.73)	74.87

(a) Original dataset (label smoothing with $\alpha = 0.1$).

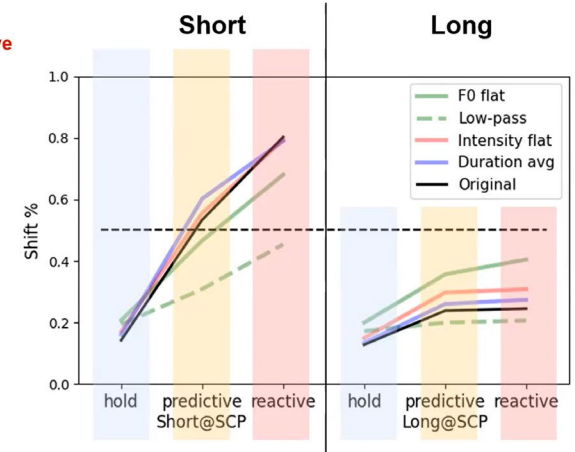
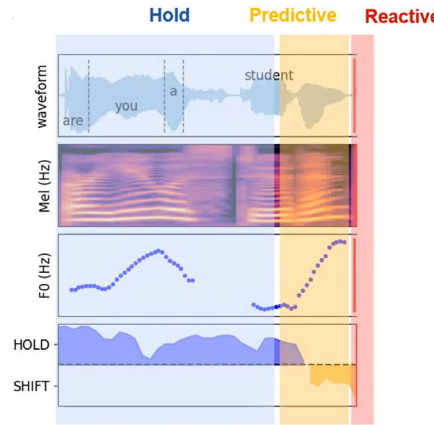
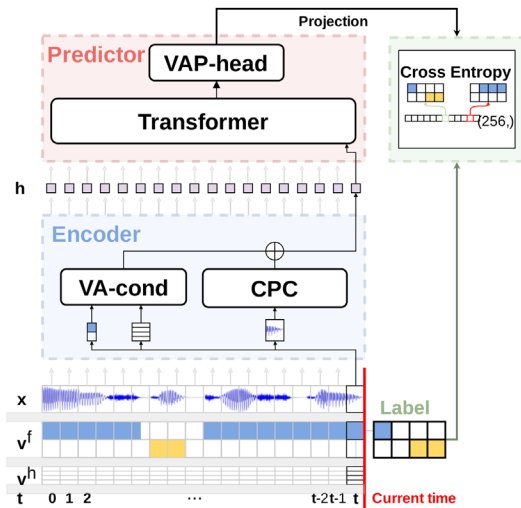
- **背景** | 深層学習モデルの実用において、*churn*は問題
 - Churn: 同じデータ&同じハイパラで学習し直しても、推論結果が変わってしまう現象
 - 実世界では、学習し直したりデータにノイズが混入したりする
- **提案** | 複数モデルのEnsemblingやdistillation学習を用いたchurn削減手法
- **実験** | 発話理解タスクにおいて、提案手法がchurn削減に有効であることを確認

研究紹介

How Much Does Prosody Help Turn-taking? Investigations using Voice Activity Projection Models



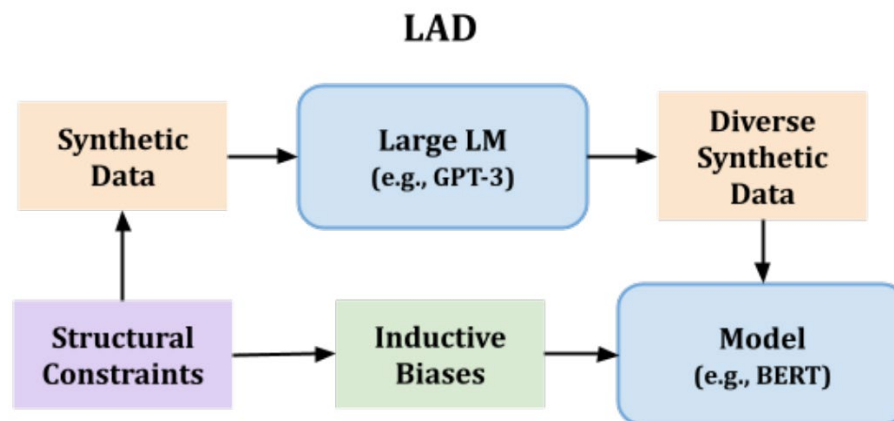
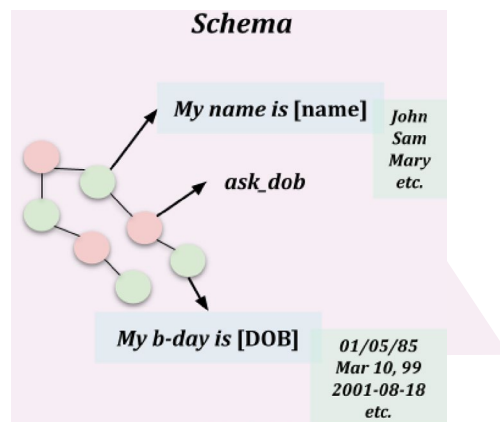
Erik Ekstedt, Gabriel Skantze (KTH)



- **疑問** | ターンテイキングのモデル化において韻律特徴はどの程度重要なのか
 - 人手実験による制約 (表面的な特徴, 被験者コスト) によって, まだ包括的な検証はない
- **実験** | VAPモデル (Ekstedt+, 2022) と大規模な音声会話データセットを用いて自動的に調査
 - **方法** | 入力音声を変化させ (例: F0の平滑化), VAPの予測精度を確認
 - **結果** | Low-pass変換にセンシティブであることが判明 → 音韻情報が特に重要

研究紹介 | LAD: Language Models as Data for Zero-Shot Dialog

Shikib Mehri¹, Yasemin Altun², Maxine Eskenazi¹ (1CMU, 2Google)

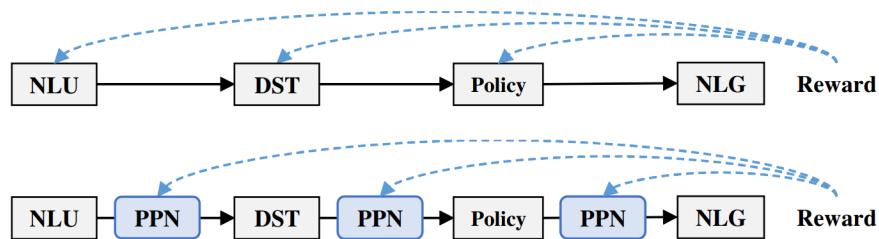


Model	F-1
Zero-Shot Results	
CONVEX (HENDERSON AND VULIĆ, 2020)	5.2
COACH+TR (LIU ET AL., 2020)	10.7
GENSF (MEHRI AND ESKENAZI, 2021A)	19.5
GENSF + LAD	50.9
Zero-Shot Results	
BERT+S (MOSIG ET AL., 2020)	28.12
SAM (MEHRI AND ESKENAZI, 2021B)	53.31
SAM + LAD	64.36

- **背景** | タスク指向型対話におけるzero-shot対話生成の実現は課題
 - タスク専用の知識が必要なので，大規模LMの汎用的なNLPの能力をもってしても困難
- **提案** | LAD: Language Models as Data
 1. GPT-3を用いて，対話の構造情報（schema）から対話データを自動生成
 2. 生成されたデータを用いて，タスク用モデル（BERT, DialoGPTなど）をfine-tuning
- **実験** | 複数のタスク（意図推定, slot filling, 行動決定）で，zero-shotにおけるSoTAを達成

研究紹介 | Post-processing Networks: Method for Optimizing Pipeline Task-oriented Dialogue Systems using Reinforcement Learning

Atsumoto Ohashi, Ryuichiro Higashinaka



Model Combination				w/ PPN	Task Success	Inform F1	Match Rate	Turn
NLU	DST	Policy	NLG					
BERT	Rule	Rule	Template	✓	84.1 84.0	87.4 86.3	90.2 92.4	5.92 6.33
BERT	Rule	MLE	Template	✓	43.3 66.1	62.4 71.5	27.8 78.7	9.03 8.61
BERT	Rule	PPO	Template	✓	54.9 68.8	65.5 72.1	55.2 77.8	8.41 8.37
BERT	Rule	Rule	SC-LSTM	✓	38.3 44.2	57.5 71.7	56.7 71.8	13.53 11.04
	TRADE	Rule	Template	✓	19.0 18.8	45.6 49.2	36.4 31.6	12.08 12.14
BERT	Rule		LaRL	✓	21.6 23.9	44.9 50.9	27.6 34.1	13.24 12.77

- **背景** | 対話システムをE2Eに最適化するには、システム全体が微分可能であることが前提
 - パイプライン構造をとるシステムの能力を、微分可能性に依存せずに最適化したい
- **提案** | Post-processing Networks (PPNs)
 - システム全体のタスク遂行能力が向上するように、各モジュールの出力を修正
 - 各モジュールにアクセスすることなく、パイプライン全体を最適化可能
- **実験** | MultiWOZを用いた自動評価と人間評価の両方で、タスク遂行能力の向上を確認



The 18th Young Researchers' Roundtable on Spoken Dialogue Systems (YRRSDS 2022)

September 5-6, Edinburgh, UK



Photo by Connor Mollison

YRRSDSの基本情報

- 対話システムの若手研究者が議論するワークショップ (SIGDIALと共催)
- 参加要件 | ポジションペーパーの提出
 - 「自身の研究内容」「議論したいテーマ」に関する原稿 (max 2ページ)
- 主なイベント
 - ラウンドテーブル, ポスターセッション, Keynote Talks
- YRRSDS 2022
 - 開催場所 | Heriot-Watt大学
 - 参加者数 | 34名 (ほとんどがPhDの学生, 日本人は自分のみ)

ラウンドテーブル

- **最近の対話システム分野のテーマについて、10名弱に分かれて議論** (計4回)

- 大橋が選択したテーマ：

Multimodality, Data Collection & (Low) Resources, LMs & Explainability, Evaluation

- **議論抜粋**

テーマ：Evaluation

- 人間によるシステム評価はコストが高すぎる → 内発的動機づけのある評価者の導入はどうか？
- 対話タスクによっては、一般的な評価尺度では測れない項目も多い
 - ✓ 例：multi-party対話では「被発話者の推定」「介入」「協力/競争目標」等も大切

テーマ：Data Collection & (Low) Resources

- 低コストにデータを収集する方法として何があるか
 - ✓ データ拡張手法を使う場合が多い。またデータの共有・寄付, Gamificationもありでは？
- 認知症系や日常会話系では倫理的な問題もあるため、簡単にはデータを公開できない

その他現地の様子 & 所感

- 参加者同士の親睦を深めることができる機会が多い



https://twitter.com/YRRSDS_Official/status/1567434204157296641

- ✓ コミュニティの雰囲気を知ることができた
- ✓ 同世代の研究者と知り合えた & 仲良くなれた
- ✓ 研究に対するモチベーション・刺激をもらった