「あなたは対話を楽しんでいますか？」
ー対話における内面状態推定の課題と展望ー

北陸先端科学技術大学院大学

岡田 将吾

Agenda

- 内面状態の推定技術の背景と周辺
- 内面状態の推定における課題と研究紹介
- 内面状態を推定することの効用とは？

Agenda

- 内面状態の推定技術の背景と周辺
- 内面状態の推定における課題と研究紹介
- 内面状態を推定することの効用とは？

感情・内面状態の理解が拓く未来．．

- ヘルスケア（うつ・認知症の推定や症状のモニタリング）
- コーチング・動機付けインタビュー
- 教育・チュータリング（集中度，覚醒状態のモニタリングと支援）
- 感情労働（コールセンター業務の代替）
- エンターテイメント（楽しんでいる状態を持続させる対話・ゲーム）
- 共感するAI (エージェント・ロボット)

会話から漏れ出る内面状態．．．

感情・感性
（ムード，センチメント）

ラポール

個人特性
（性格・スキル）

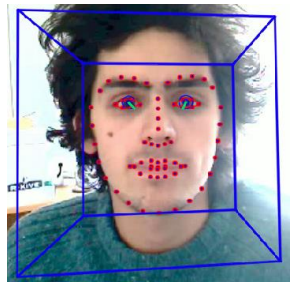
関係性・相性

他にも，認知機能など．．．

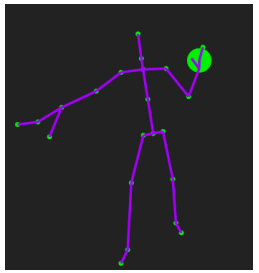
会話中に観測されるマルチモーダル情報

視覚情報

例：表情，ジェスチャ



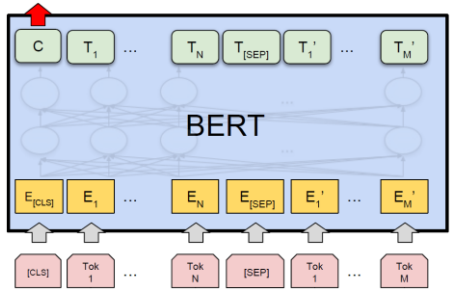
OpenFace



Kinect

言語情報

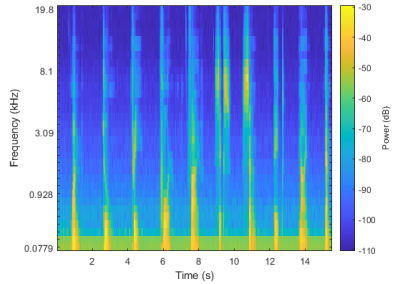
例：発話内容



BERT

韻律情報

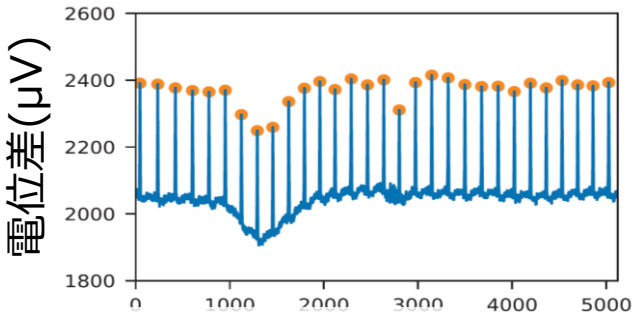
例：声の大きさ，高さ



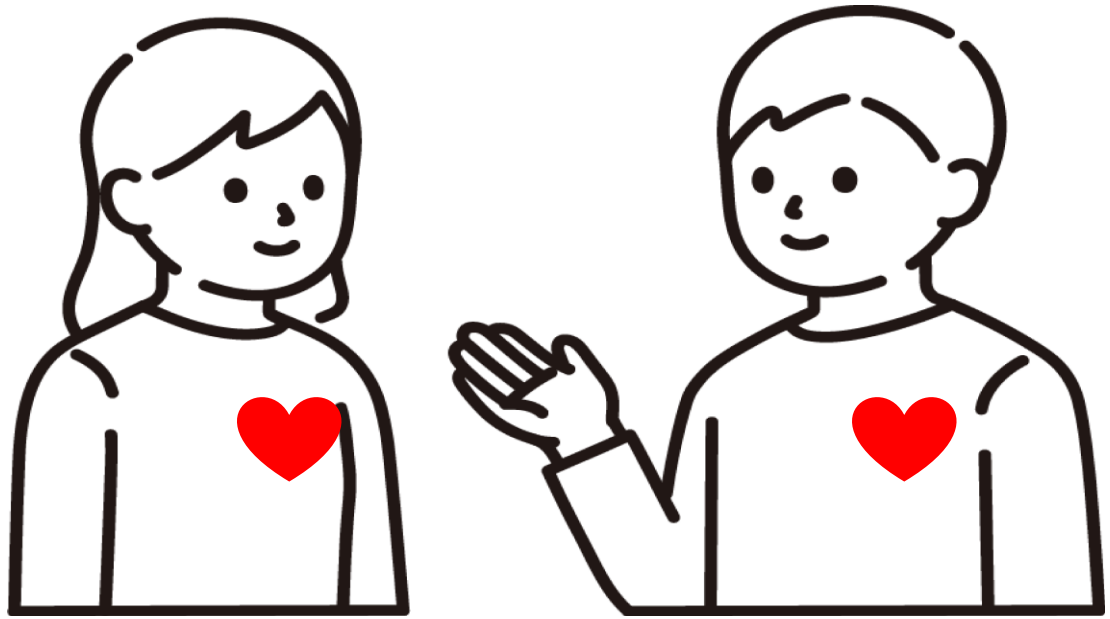
MelSpectrogram

生体信号

例：心拍，皮膚電位



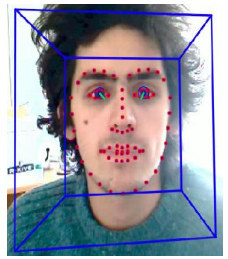
心拍データ(心拍波形)



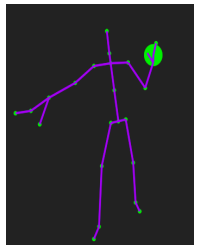
マルチモーダル行動に基づく内面状態推定

視覚情報

例：表情，ジェスチャ



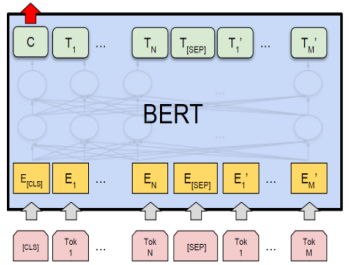
OpenFace



Kinect

言語情報

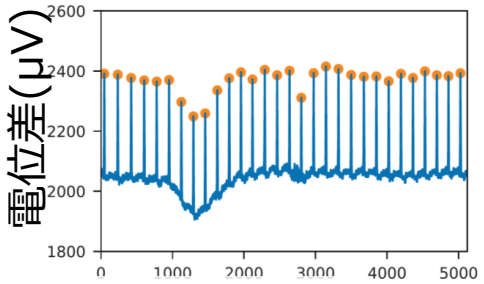
例：発話内容



BERT

生体信号

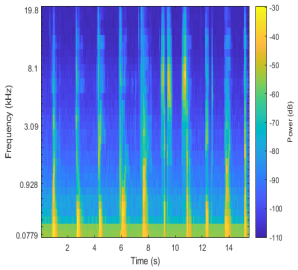
例：心拍，皮膚電位



心拍データ(心拍波形)

韻律情報

例：声の大きさ，高さ



MelSpectrogram

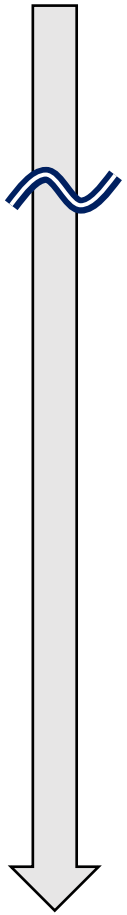
入力X

- 信号処理
- 特徴抽出
- 機械学習

出力Y

- 感情・感性
  （ムード，センチメント）
- ラポール
- 個人特性
  （性格・スキル）
- 関係性・相性

感情推定チャレンジの変遷



1970s~

ユニモーダル

- Facial Action Coding System (顔面動作の符号化) — Ekman & Friesen, 1978
- ブルンスウィキアン レンズモデル (音声感情認識モデル) — Scherer, 1978
- 単語の感情極性分類 (知識ベース教師無しモデル) — Hatzivassiloglou and McKeown, 1997

2000 Human-Agent (*SEMAINE*)

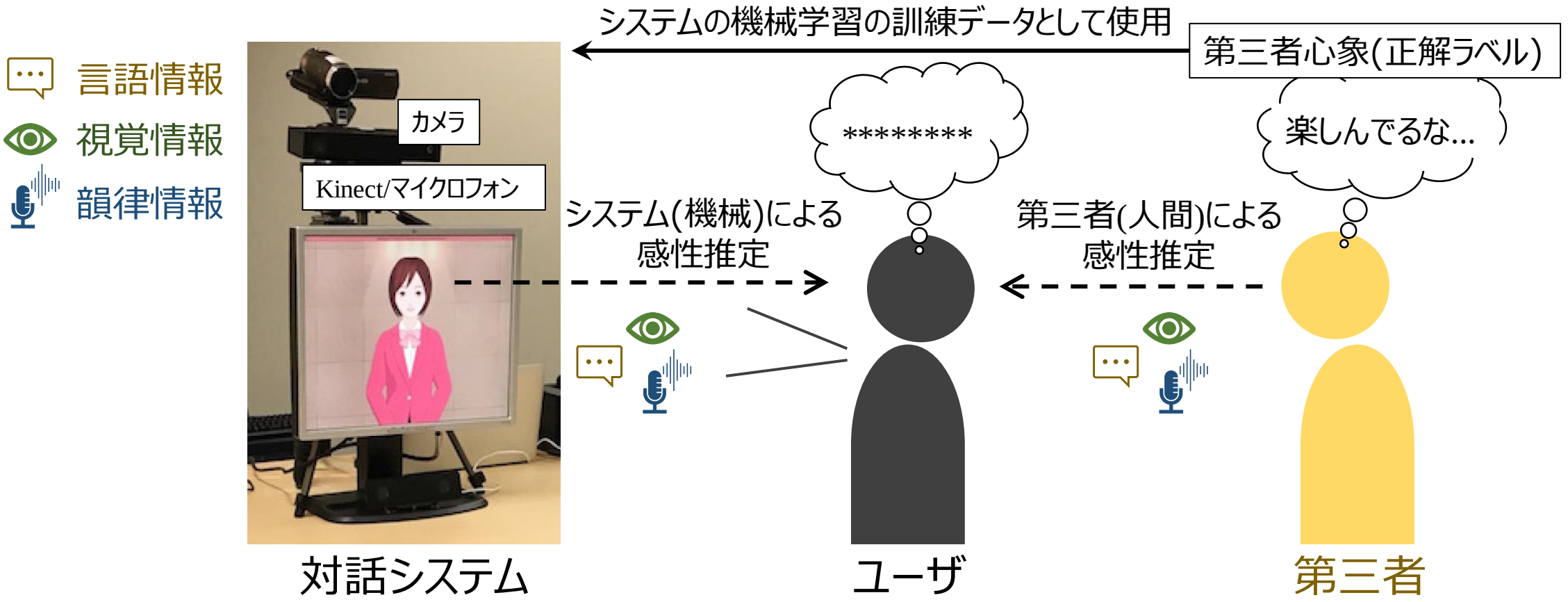
マルチモーダル

2008	*IEMOCAP*	演技による対話
2012	*SEMAINE*	自然な対話
2013	*RECOLA*	自然な対話
	AMIGOS	感情刺激材料による実験
2017	*MOSI/MOSEI*	独話(YouTube)
2019	*MELD*	演技による会話(TVドラマ)
	Hazumi	自然な対話

- **入力データ：**ユニモーダル→　**マルチモーダルへ**
- **データ取得環境：**感情刺激材料による統制環境→　演技による対話→　**自然な対話へ**

本人心象と第三者心象

・これまでの感性推定では機械学習に使用される正解ラベルに関して第三者が付与



本人が実際にどう感じているか(本人心象)は考慮されていない

データセットから見る本研究の課題

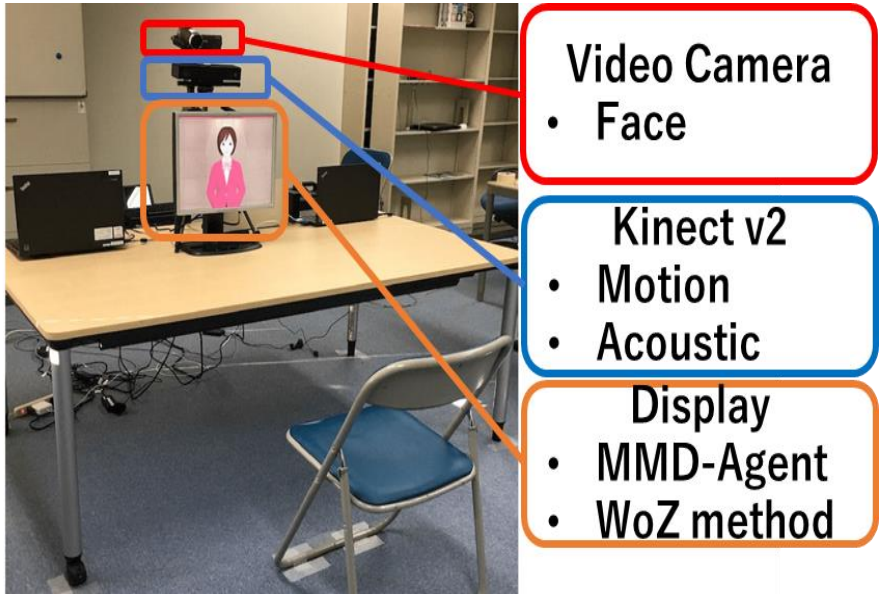
対話感性分析に利用可能なマルチモーダルデータセット

名称	内容	形式	モダリティ	アノテーション
IEMOCAP	台本のある対話	対話	言語，韻律，視覚	第三者 発話レベルのラベル
SEMAINE	エージェントとの対話	対話(エージェント)	韻律，視覚	
RECOLA	ビデオ会議での対話	対話	韻律，視覚，生体	
MOSI	映画レビュー	独話	言語，韻律，視覚	
MELD	TVドラマ	複数名の会話	言語，韻律，視覚	
Hazumi	エージェントとの会話	対話(エージェント)	言語，韻律，視覚，生体	本人，第三者 発話，対話 レベルのラベル

いずれのデータセットにおいても本人心象の動的な変化はアノテーションされていない

感性分析の多くは第三者によるラベルが使用される

Hazumi: マルチモーダル対話データセット



データ

- 対面対話条件
  - 89セッション，7744交換
- リモート対話条件
  - 126セッション， 10367交換

ユーザの主観評価ラベル

- 発話交換レベル

  心象（自己/第三者付与），話題継続
- 対話レベル

  対話の楽しさ，対話のぎこちなさ

Komatani and Okada. Multimodal Human-Agent Dialogue Corpus with Annotations at Utterance and Dialogue Levels. ACII 2021

Agenda

- 内面状態の推定技術の背景と周辺
- **内面状態の推定における課題と研究紹介**
- 内面状態を推定することの効用とは？

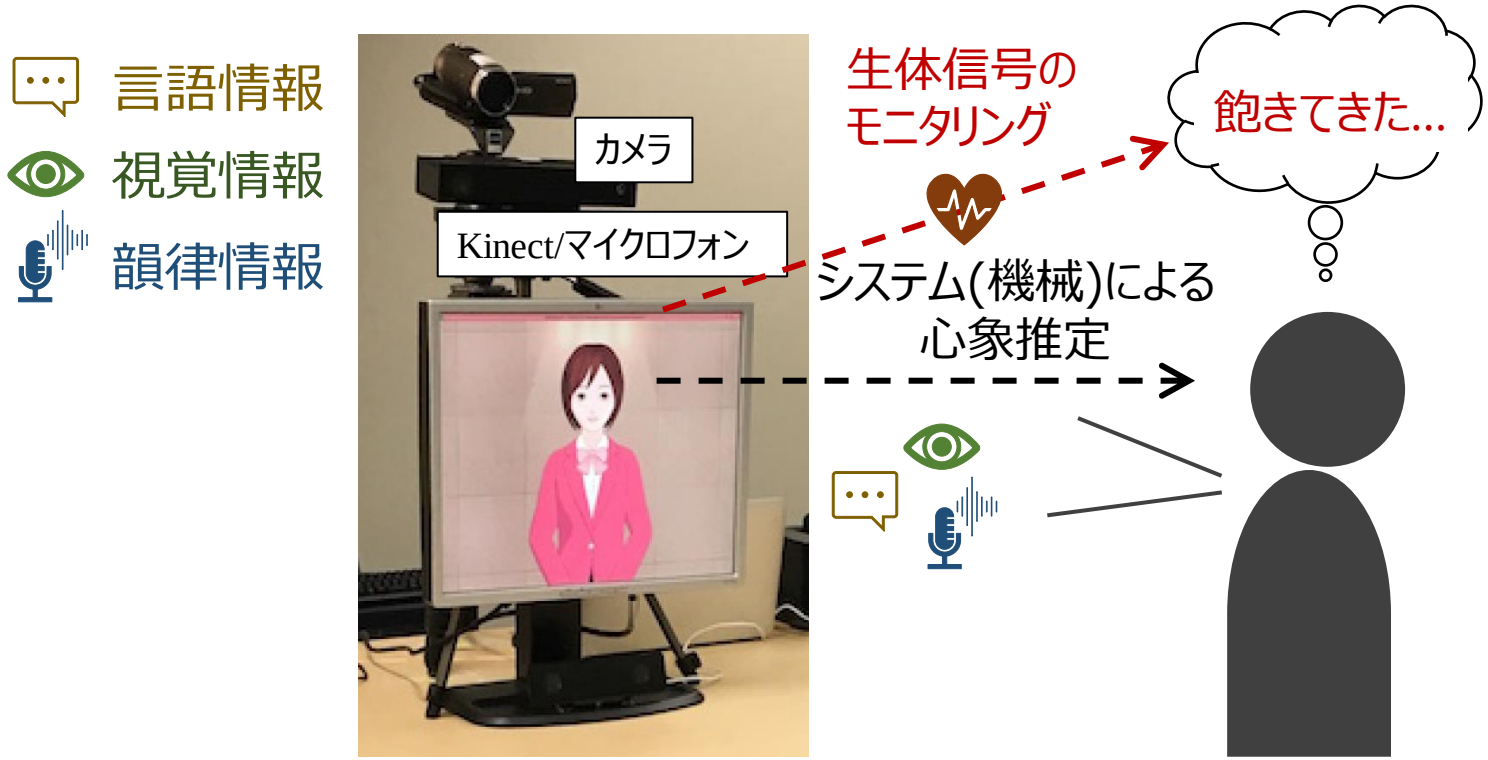
内面状態推定のためのMM機械学習の現状・課題

1. 感情が外面的な行動として表出しない問題
   - （例）笑っていても，本心では悲しいと感じている.
2. 学習に悪影響 なサンプルの影響
   - （例） アノテーション一致率が低い
3. 訓練サンプルの少量問題
   - （例）アノテーション収集コスト，マルチモーダルデータの収集コスト
4. 話し言葉における情報の欠損
   - （例）省略された単語に重要な感情の手がかりがある．．．

内面状態推定のためのMM機械学習の現状・課題

1．感情が外面的な行動として表出しない問題

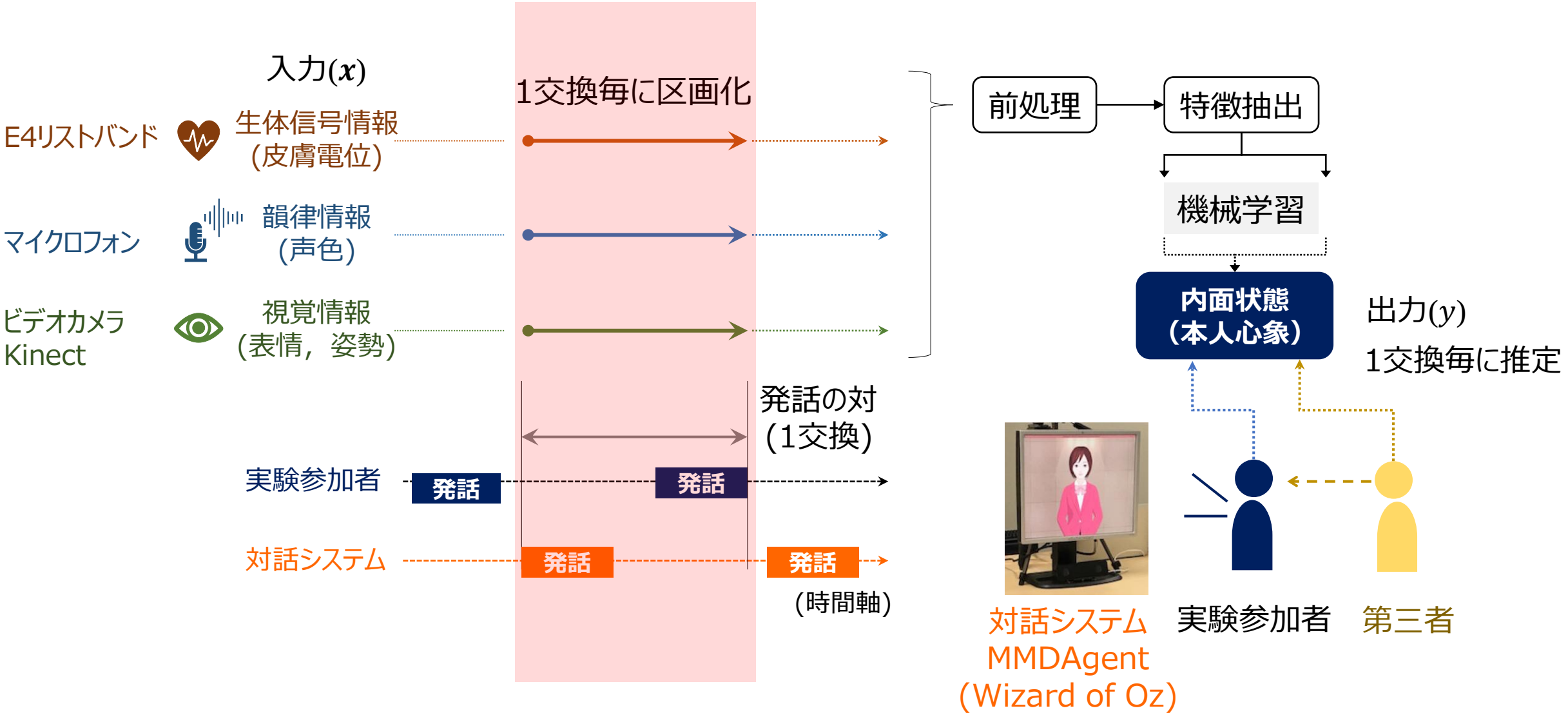
⇒　**意識的な制御が困難な生体信号情報を利用**



Shun Katada,et al. Is She Truly Enjoying the Conversation? Analysis of Physiological Signals toward Adaptive Dialogue Systems.ICMI2020
Shun Katada, et al, Effects of Physiological Signals in Different Types of Multimodal Sentiment Estimation, IEEE Trans. on Affective Computing 2022

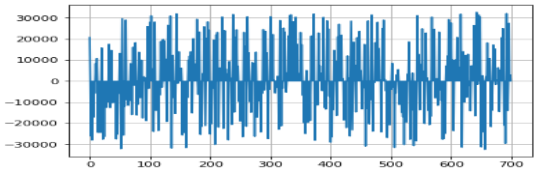
対話交換毎の本人心象の推定タスク

マルチモーダル対話コーパス: Hazumi1911データセット

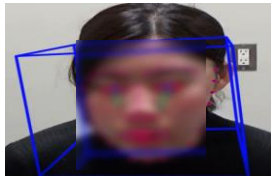


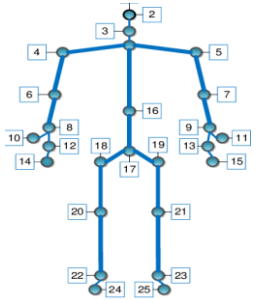
マルチモーダル非言語特徴抽出

- 生体信号特徴量
  - 皮膚電位特徴量 (GSR数，統計量)
  - 心拍特徴量
- 韻律特徴量
  - OpenSMILE

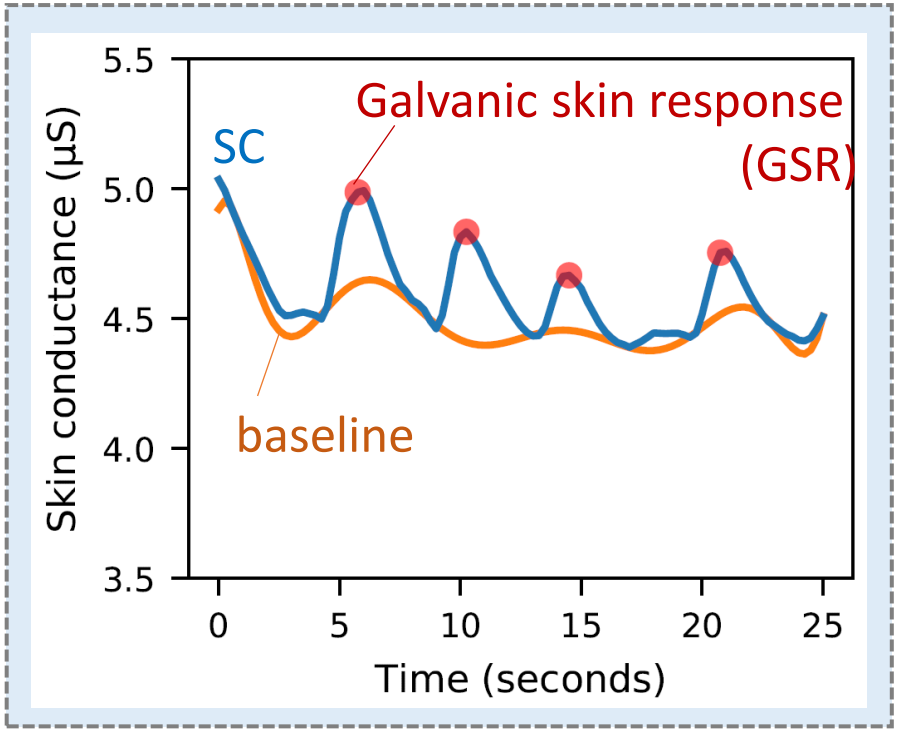


- 視覚特徴量
  - OpenFace (表情)
  - Kinect (姿勢，ジェスチャ)





システム対話時皮膚電位解析の例



Empatica 社
E4センサ

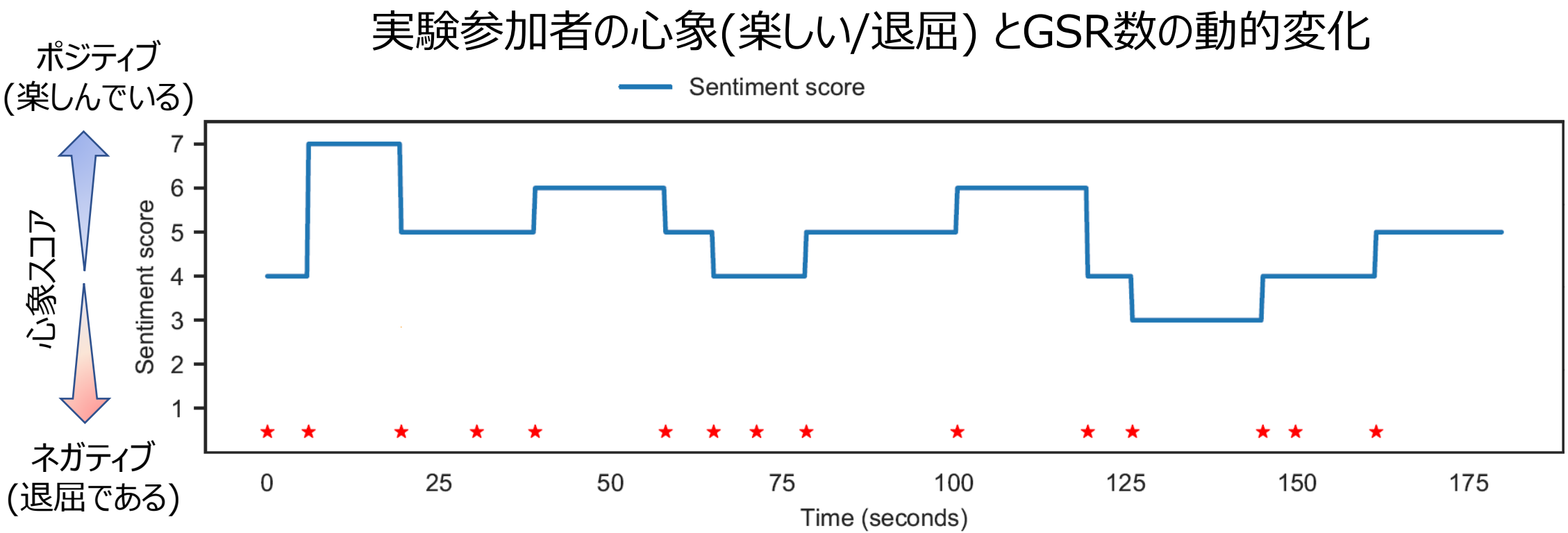
本人心象推定性能 (二値分類正解率)の比較

P, 生体信号特徴量; A, 韻律特徴量; V, 視覚特徴量; Uni, ユニモーダル; Multi, マルチモーダル

Physiological features	Model	Uni	Multi						Uni		Multi		Human model
		P	A+P		P+V		A+P+V		A	V	A+V		
			EF	LF	EF	LF	EF	LF			EF	LF	
EDA+HR (27)	SVM	57.7	57.0	60.3	57.5	58.7	56.8	60.2	57.7	58.2	57.1	58.9	63.0
EDA (14)		**61.6**	60.4	61.4	60.7	61.2	58.4	61.2					
HR (13)		52.5	57.0	55.0	56.7	54.9	56.9	57.1					
EDA+HR (27)	DNN	60.1	58.9	58.7	60.5	60.0	59.7	60.1	57.3	57.7	58.4	58.1	
EDA (14)		**62.2**	60.2	59.4	**63.2**	62.9	60.8	61.0					
HR (13)		48.6	56.1	55.4	53.7	54.3	55.7	56.9					

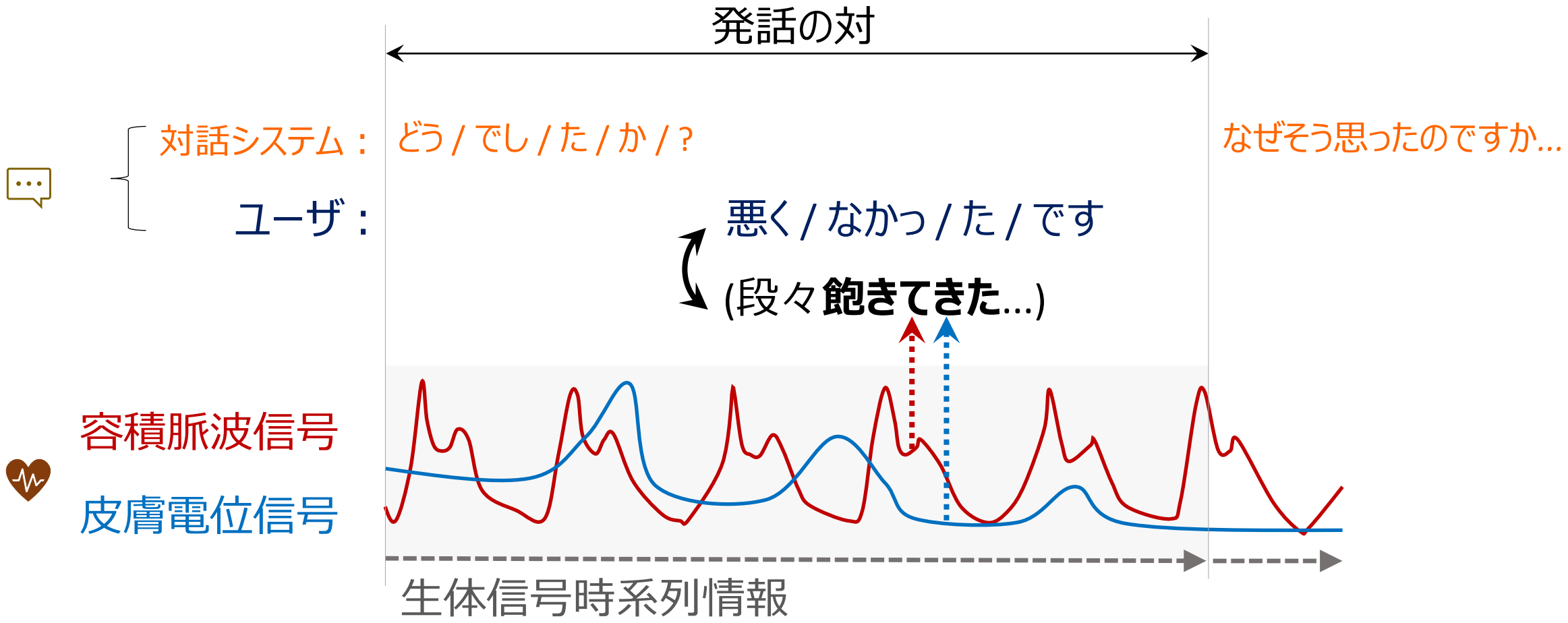
- 皮膚電位特徴量モデルが韻律/視覚特徴量モデルを上回る
- Deep Neural Network (DNN)モデルにおいて推定性能がさらに向上
- DNNモデルの正解率は第三者(人間)による推定と同等

皮膚電位特徴量の解析結果



- 適応的対話システム構築のためにはこのような動的変化を捉える必要
- 心象スコアとGSR数の共起が推定性能に役立つ可能性

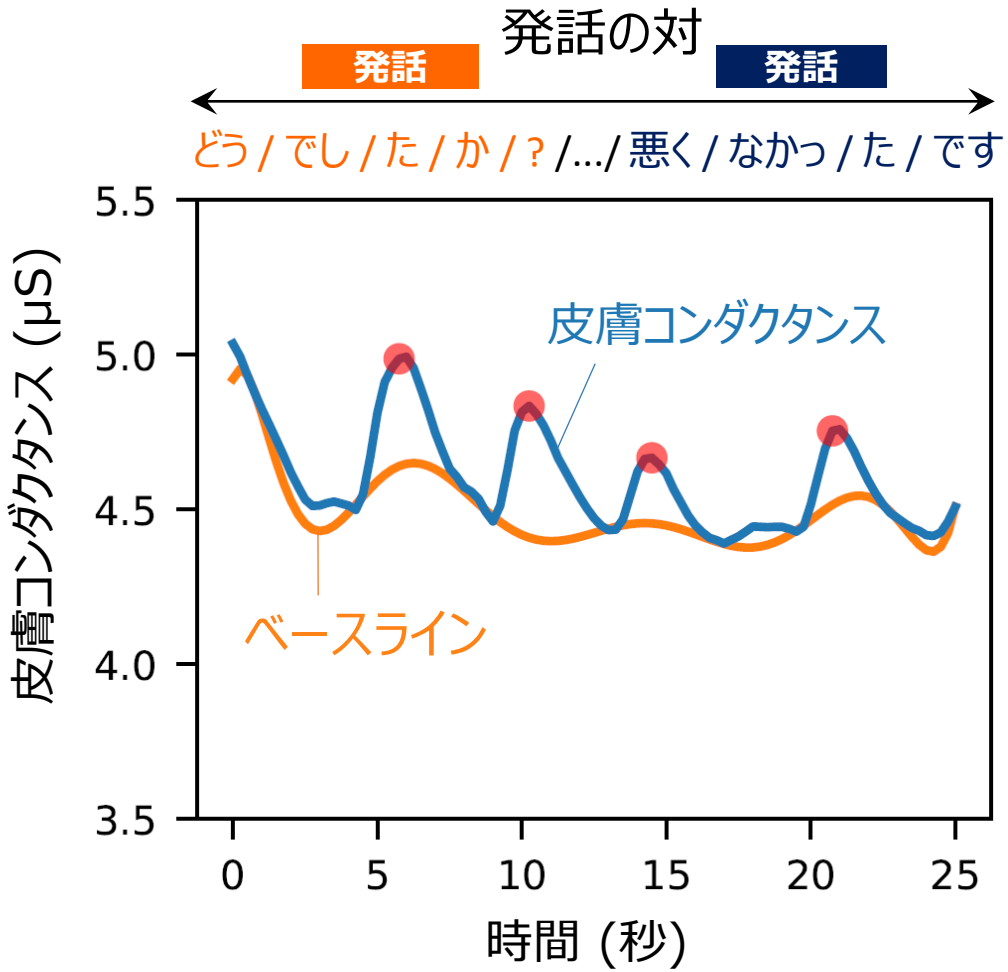
言語情報と生体信号情報の時系列統合



生体信号時系列情報が発話内容に表出しない変化を捉える可能性

Katada et.al, Transformer-Based Physiological Feature Learning for Multimodal Analysis of Self-Reported Sentiment, ICMI2022

皮膚電位信号処理と感情変化の関連

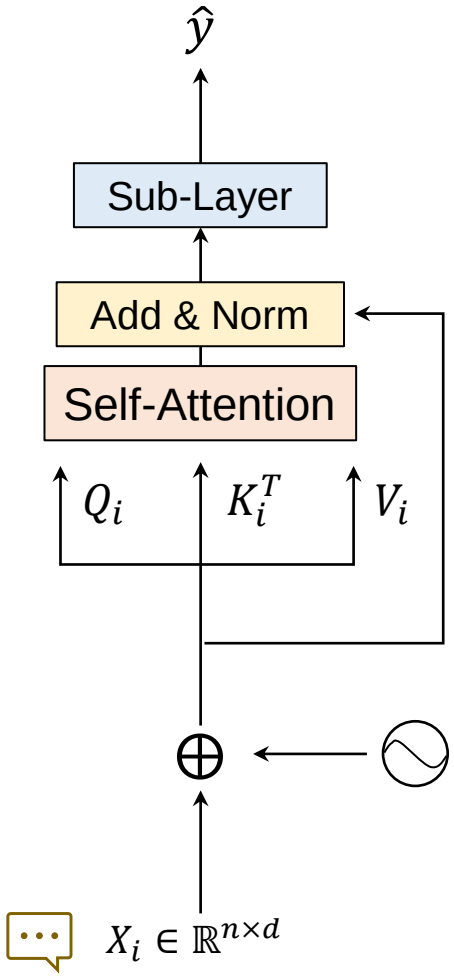


皮膚コンダクタンスとベースラインの差分
= fast phasic component
⇒ この変化が感情と関連があるとされる
⇒ この成分を中心にモデル構築を検討
EDA_{fast}と表記する

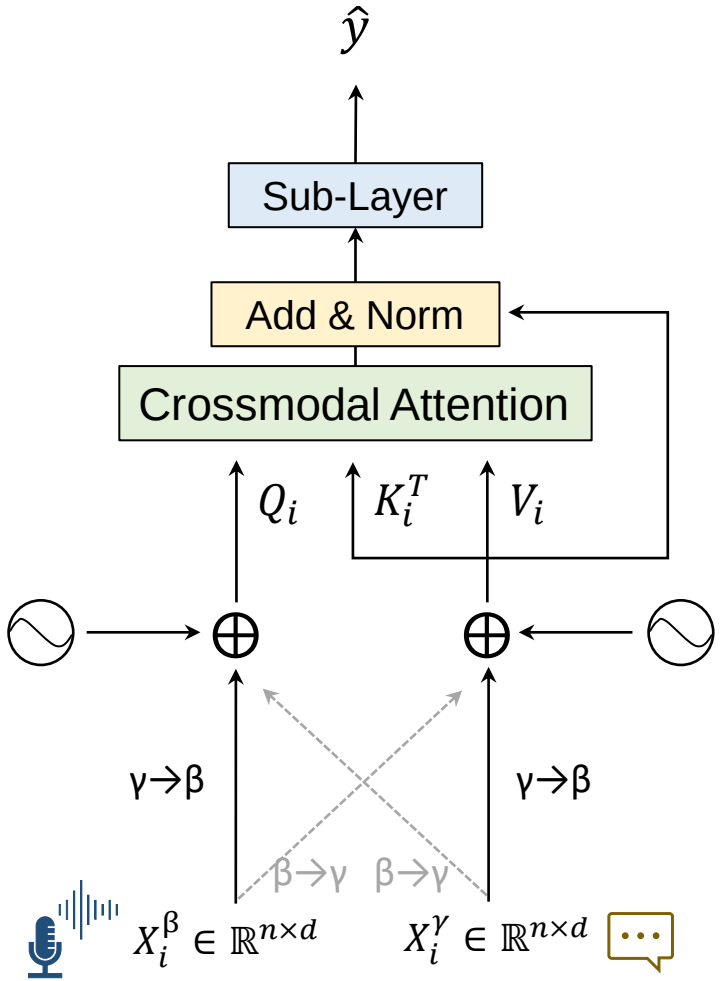
ベースライン
= tonic component
⇒ 気温(ノイズの一つ)の影響を受ける

多項式近似によりベースラインを定めfast phasic componentを算出

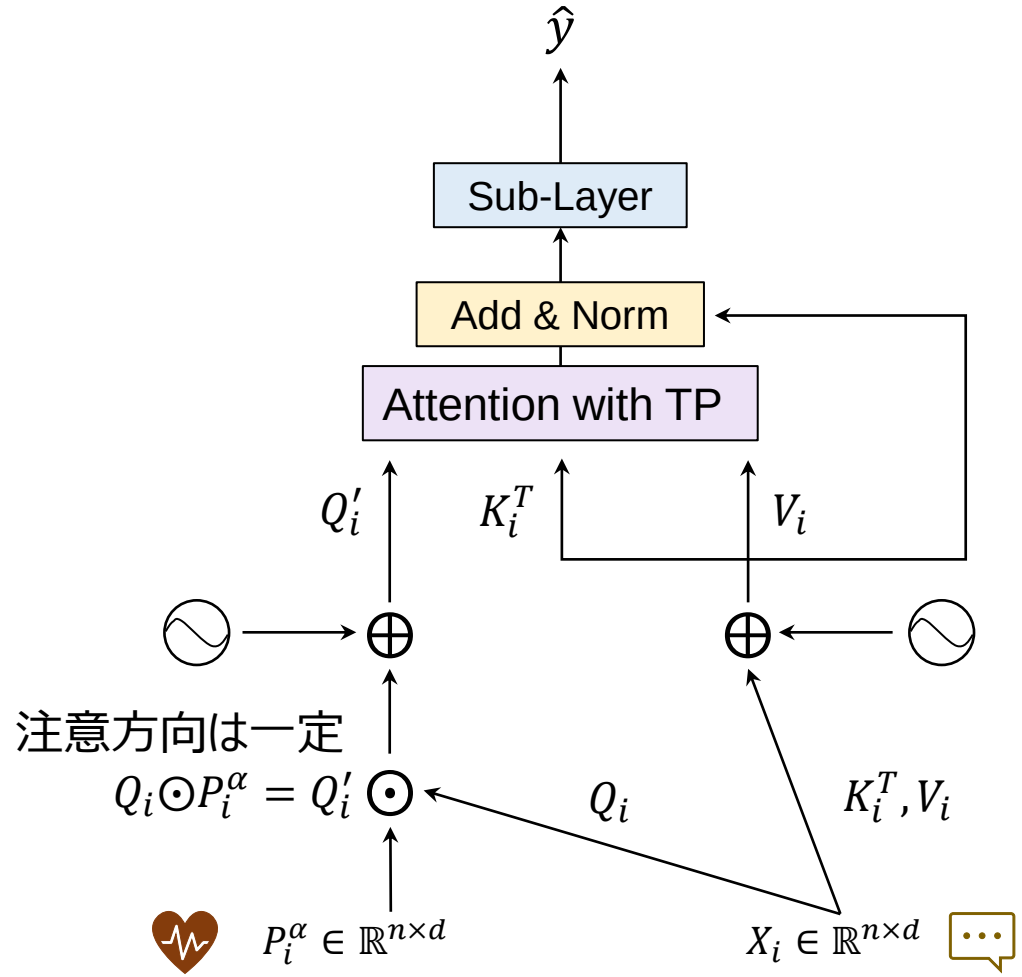
Time-series Physiological signal with Transformer (TPTr, 提案手法)



Transformer
(Vaswani *et al*., 2017, NIPS)



Crossmodal Transformer
(Tsai *et al*., 2019, ACL)



TPTr, 提案手法
(Katada *et al*., 2022, ICMI)

実験結果：提案手法TPTr （EDA_{fast})を用いた性能評価

言語のみ

言語のみ

言語・生体

各モデルによる本人心象推定結果 (回帰)

モデル	Reference	単一モデル	
		MAE	Corr
Tr	NIPS, 2017	1.082	0.227
Tr×3		1.109	0.219
CMTr	ACL, 2019	1.083	0.190
TPTr	提案手法	1.114	0.228
TPTr×3	提案手法	**1.068**	**0.232**

×3…Transformerブロック(ネットワーク)を3つ並列処理

提案手法を用いたモデルにおいて最高性能

内面状態推定のためのMM機械学習の現状・課題

2．学習に悪影響 なサンプルの影響

- （例） アノテーション一致率が低い

⇒ **パラメータ最適化（モデル訓練）を阻害するサンプルの影響を軽減**

Yuki Hirano, Shogo Okada, Kazunori Komatani: Recognizing Social Signals with Weakly Supervised Multitask Learning for Multimodal dialogue Systems. ACM International Conference on Multimodal Interaction (ICMI) 2021, pp.141-149

内面状態推定における機械学習の課題

内面状態ラベルの不確実性

- （第三者ラベル）アノテータ間でのスコアの不一致
- （本人ラベル）異なる主観性を持つユーザによるスコア比較が困難

⇒ **機械学習におけるパラメータ最適化に影響を及ぼす**

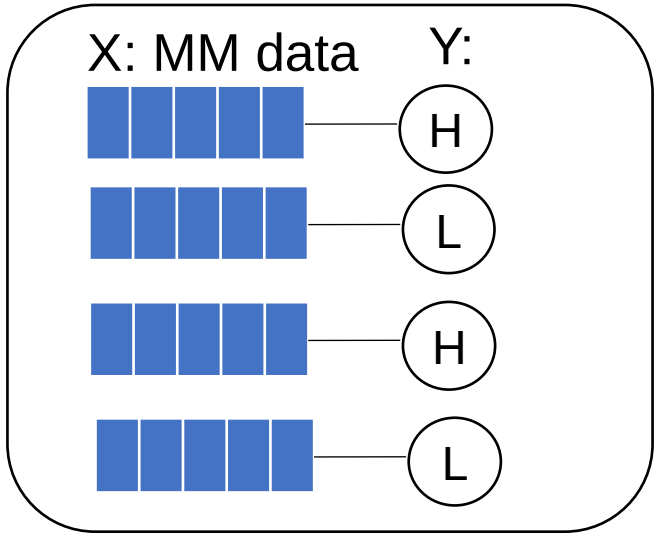
内面状態推定における機械学習の課題の解決策

内面状態ラベルの不確実性

- （第三者ラベル）アノテータ間でのスコアの不一致
- （本人ラベル）異なる主観性を持つユーザによるスコア比較が困難

（解決策） Weakly supervised learning (弱教師付き)の導入 [Zhou+2020] [Jiang+2018][Han+2018]

Supervised learning

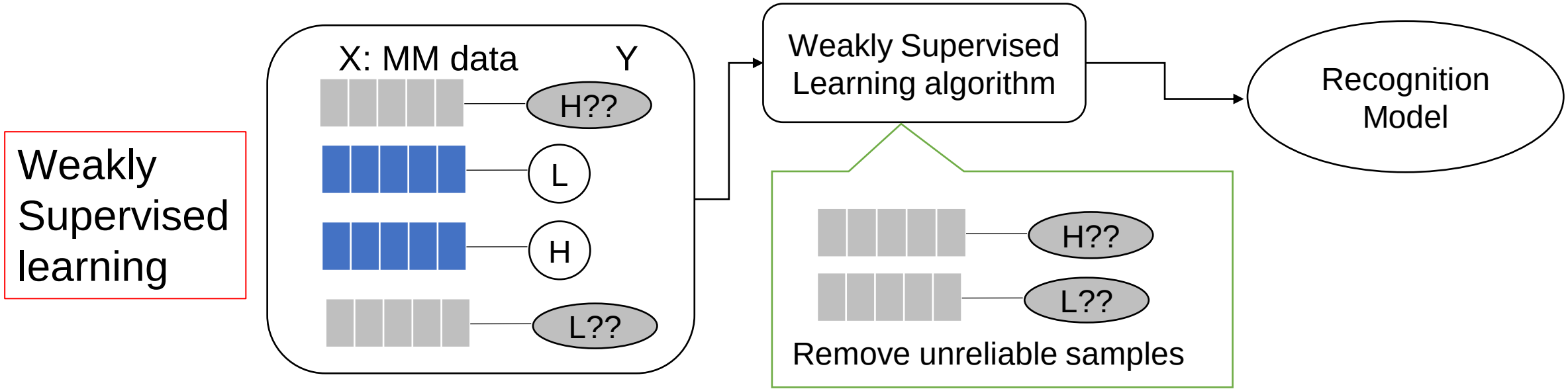


内面状態推定における機械学習の課題の解決策

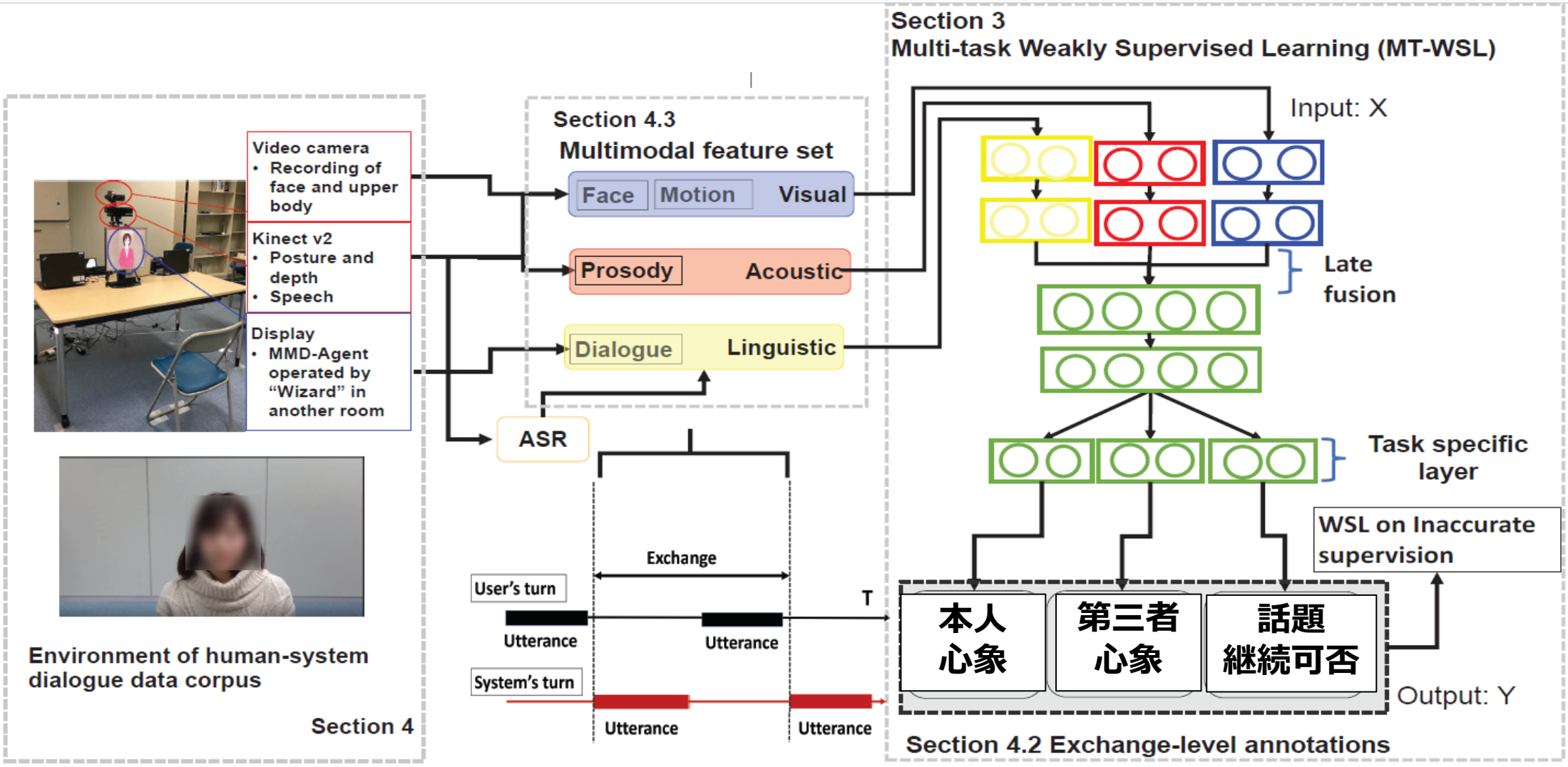
内面状態ラベルの不確実性

- （第三者ラベル）アノテータ間でのスコアの不一致
- （本人ラベル）異なる主観性を持つユーザによるスコア比較が困難

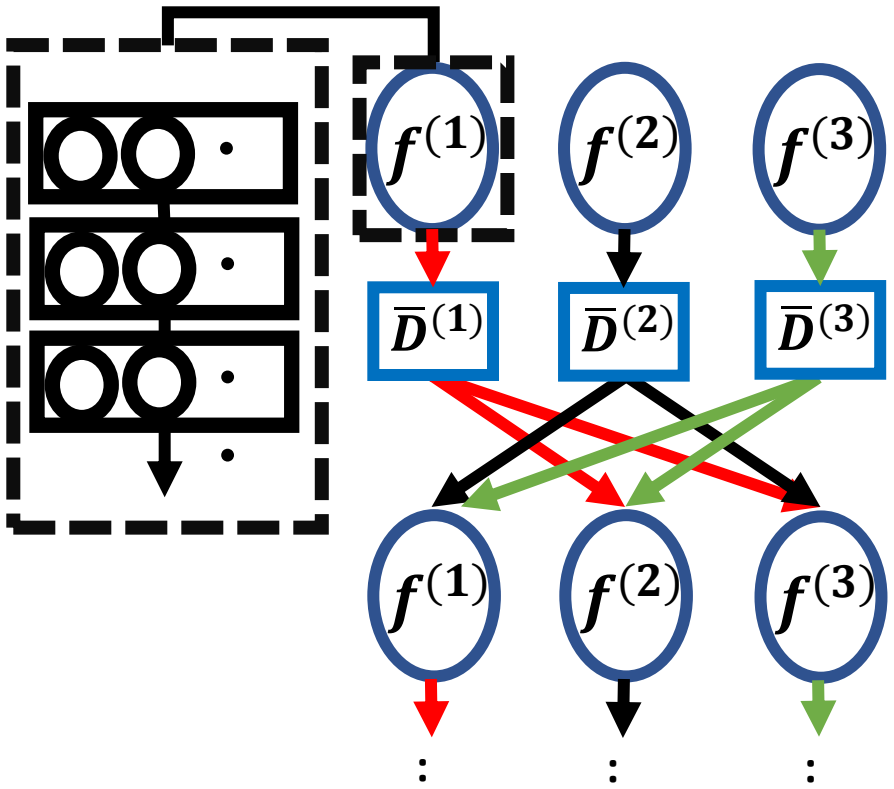
（解決策） Weakly supervised learning (弱教師付き)の導入 [Zhou+2020] [Jiang+2018][Han+2018]



弱教師付き学習に基づく内面推定



提案手法： 弱教師付き学習
Tri-teaching



基本的な戦略：

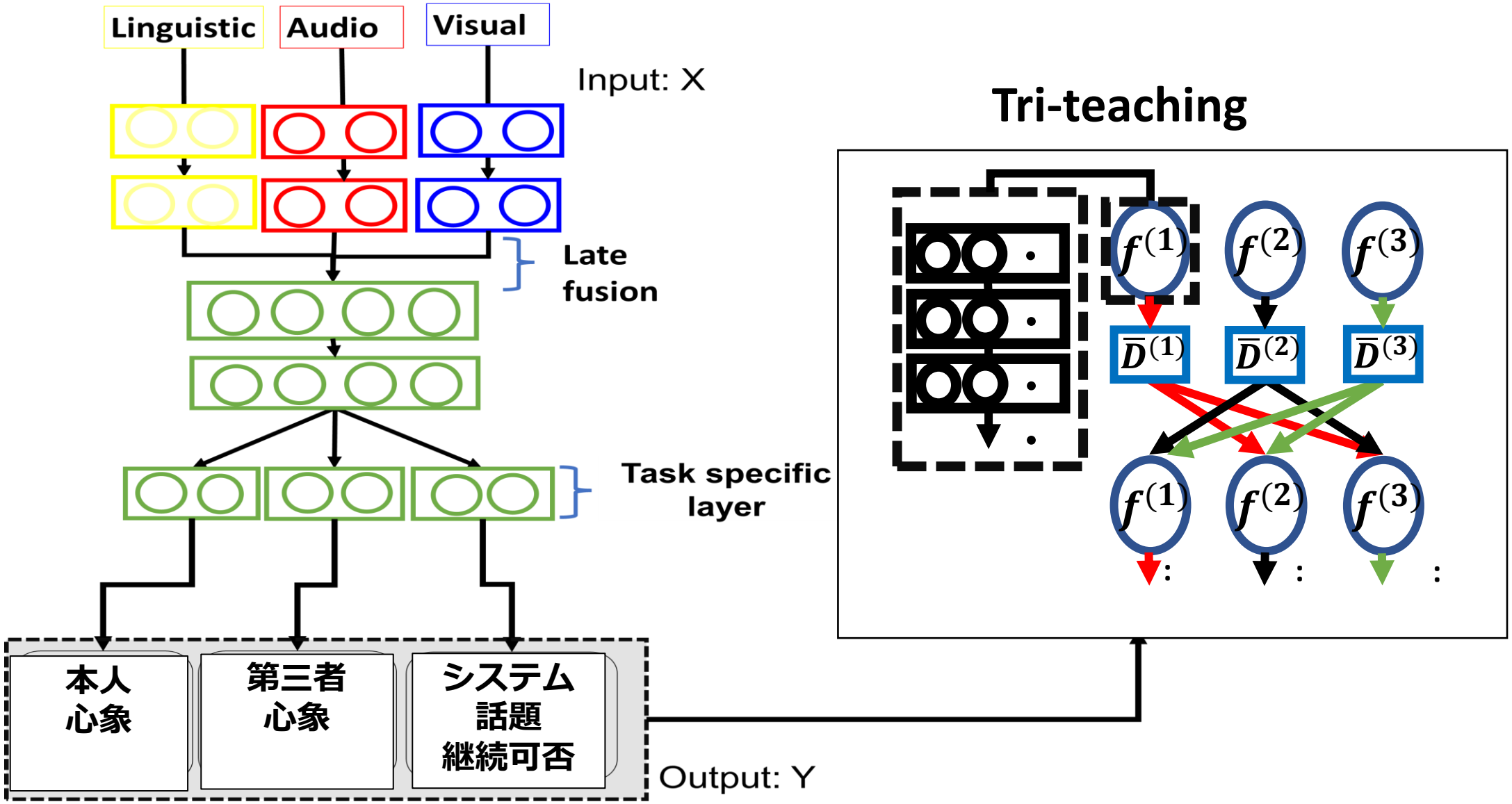
- 複数の同一モデルを用いてLossを減らせるサンプルを探索
- **仮説:** 学習を阻害するサンプルのLossを減らすのは難しい
- 小さいLossのサンプルをお互いのモデルで交換して学習

拡張：

- 2つから3つのモデルでのアンサンブルに拡張

Bo Han, et al, 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In Proc. Conference on Neural Information Processing Systems (NeurIPS). 8527–8537.

マルチタスク弱教師付き学習



クロスコーパスでの推定精度の比較

異なるコーパスを訓練とテストに用いたより現実的な実験設定

Hazumi1902は，より広い年代のユーザとの対話を収録

[%]	Hazumi1712 → Hazumi1902		Hazumi1902 → Hazumi1712	
	第三者心象	話題継続	第三者心象	話題継続
ベースライン	63.2	56.2	56.5	51.9
MT-CL-pre [Lotfian+ 2019]	64.1	56.4	56.7	52.4
MT-CL-agree [Lotfian+ 2019]	63.0	55.5	56.3	51.6
Co-teaching (2つのモデル)	63.6	58.6	56.8	52.2
Tri-teaching (3つのモデル)	**66.8**	**63.6**	**57.2**	52.2

CL-pre[Lotfian+ 2019]: 小さいLossのサンプルから先に学習するカリキュラム学習

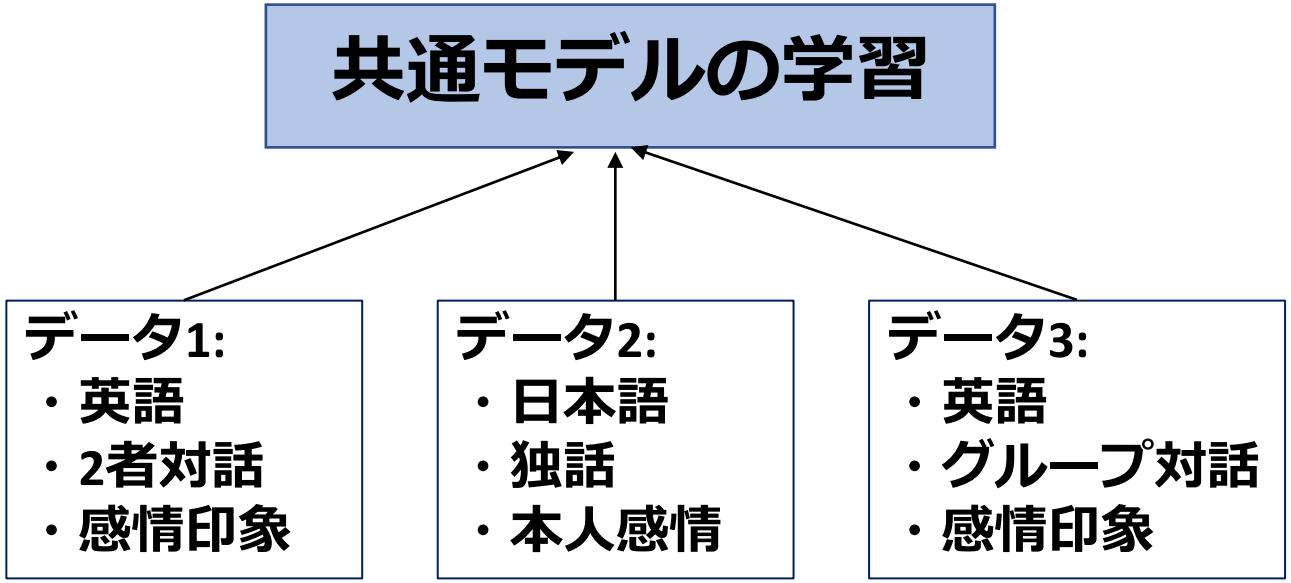
CL-agree[Lotfian+ 2019]: アノテーション一致率の高いサンプルから先に学習するカリキュラム学習

内面状態推定のためのMM機械学習の現状・課題

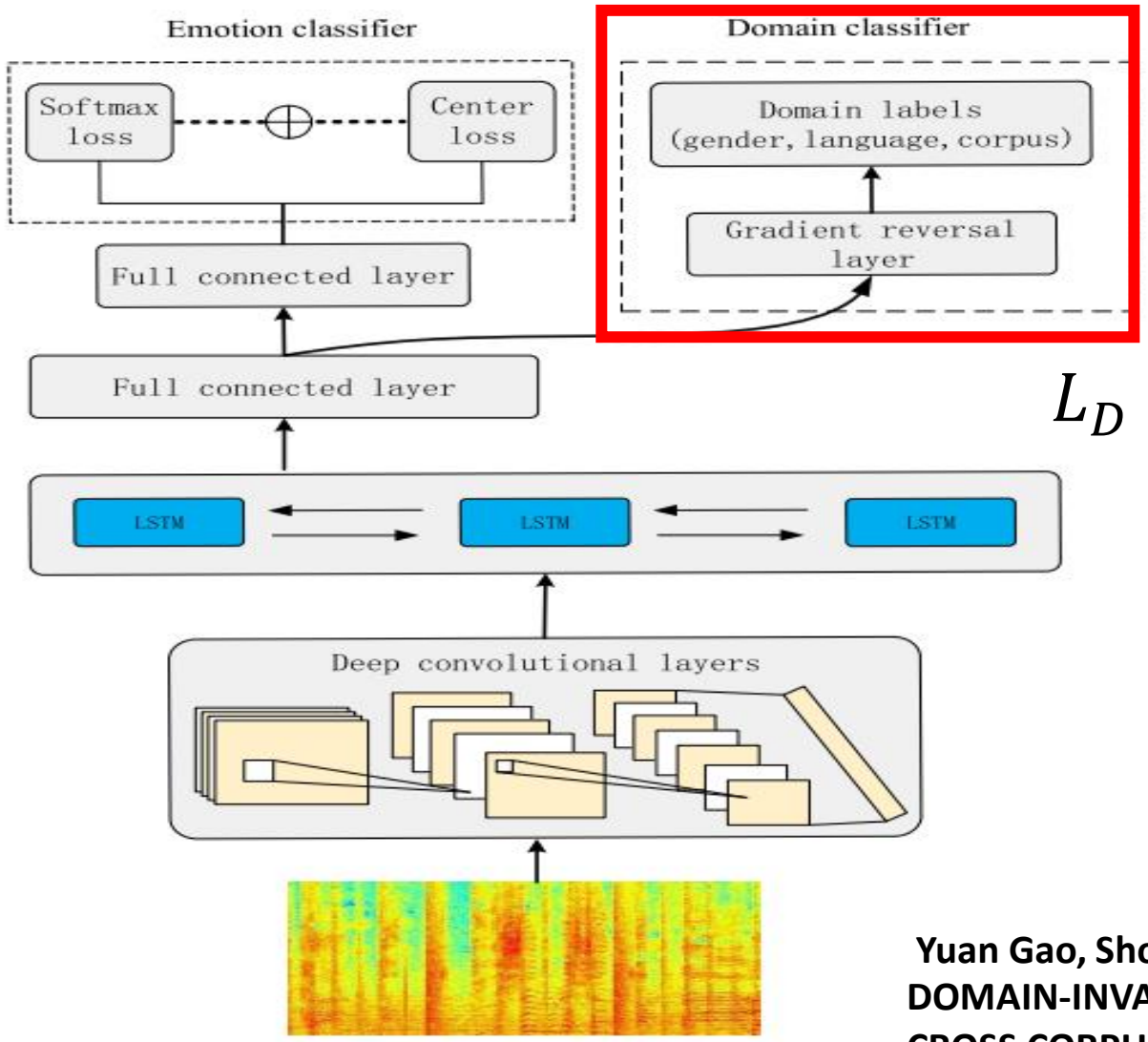
３．訓練サンプルの少量問題

- （例）アノテーション収集コスト，マルチモーダルデータの収集コスト

⇒　**複数のデータセットから内面状態推定に共通する特徴量を学習**



音声感情認識における主目的変数以外の変数の効果の軽減 ：Domain Adversarial Neural Network



話者，コーパス，言語を
識別しないように
Lossを制御

$$L_D = L_e - \lambda(\beta L_s + (1-\beta)L_c)$$

Yuan Gao, Shogo Okada, et al.
DOMAIN-INVARIANT FEATURE LEARNING FOR IMPROVING CROSS CORPUS SPEECH EMOTION RECOGNITION, IEEE ICASSP 2022

クロスコーパスでの音声感情推定精度の比較

- Baseline: CNN + LSTM
- DANN_1: 話者とコーパスを識別しないよう感情のみ識別学習
- DANN_2: ジェンダーと言語を識別しないよう感情のみ識別学習
- S: Softmax loss function
- C: Center loss function

		Arousal 覚醒度(興奮-リラックス)				**Valence** 感情価(快-不快)			
Model	Loss	MSP	SAVEE	Emodb	Average	MSP	SAVEE	Emodb	Average
Baseline	S	59.72	73.75	67.35	66.94	**59.52**	54.79	**49.73**	54.68
DANN_1	S	60.51	74.58	66.05	67.05	59.19	56.15	47.53	54.29
DANN_2	S	**63.57**	74.79	69.58	69.31	57.57	57.08	49.05	54.57
DANN_2	S+C	62.97	**75.2**	**71.64**	**69.94**	57.26	**58.12**	49.46	**54.95**

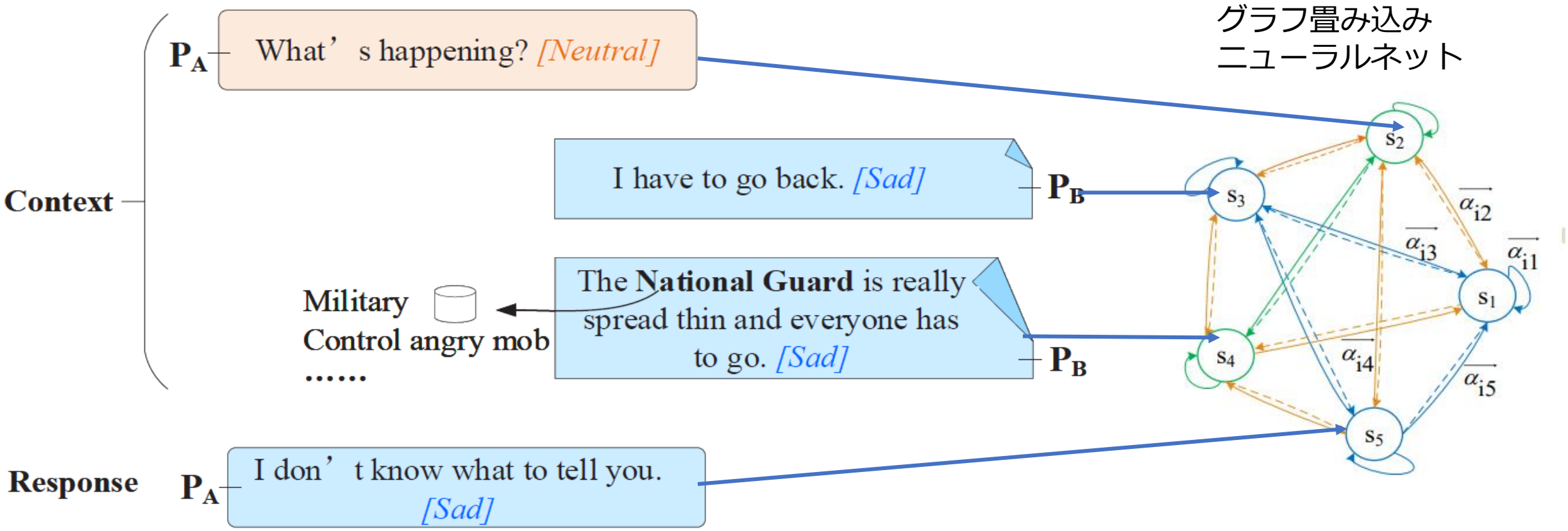
内面状態推定のためのMM機械学習の現状・課題

４．話し言葉における情報の欠損

- （例）省略された単語に重要な感情の手がかりがある（かも）．．．

⇒　**外部知識（知識データベース）を援用して，関連知識の極性を利用**

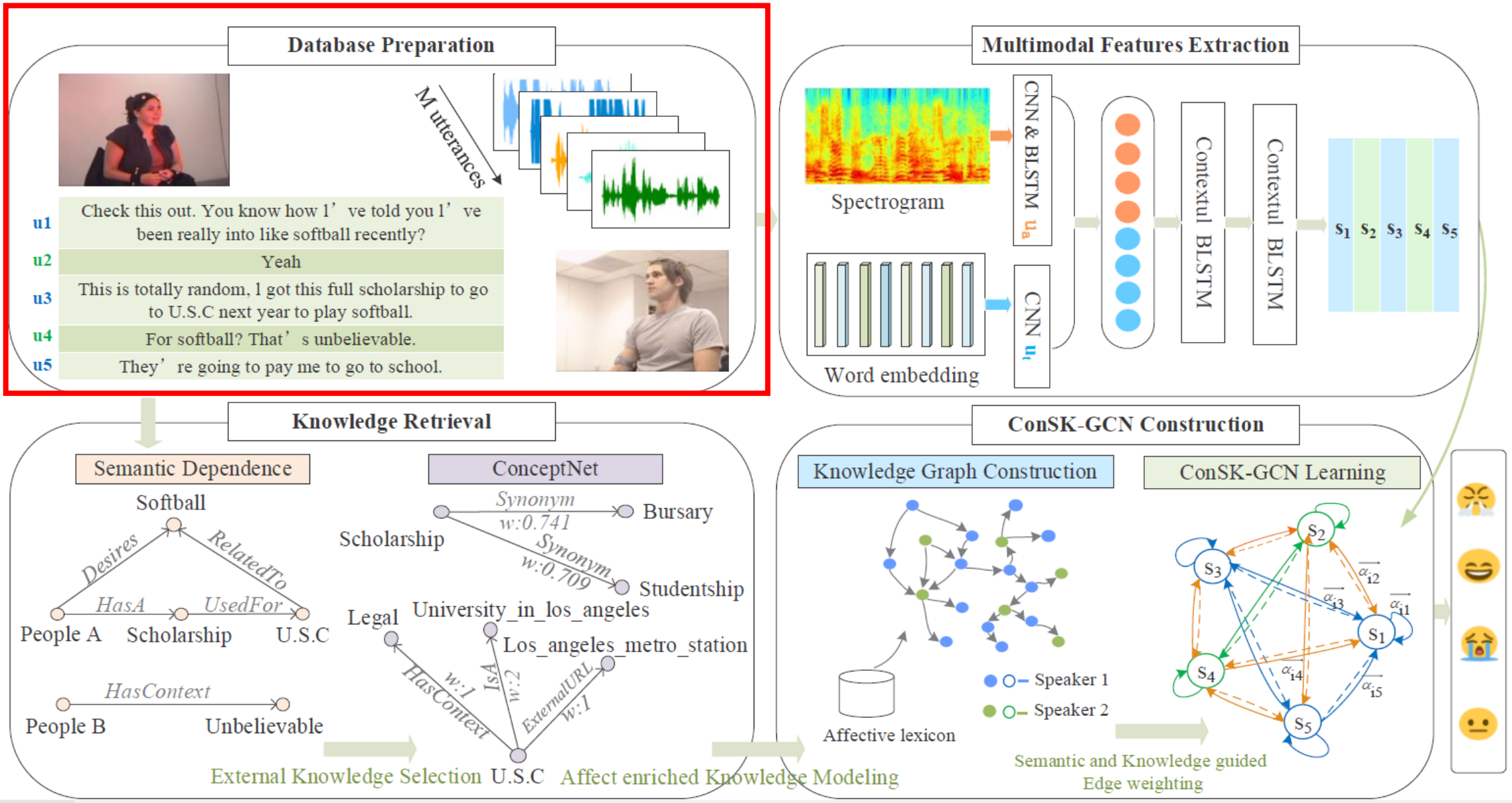
外部知識の利用に基づく感情推定



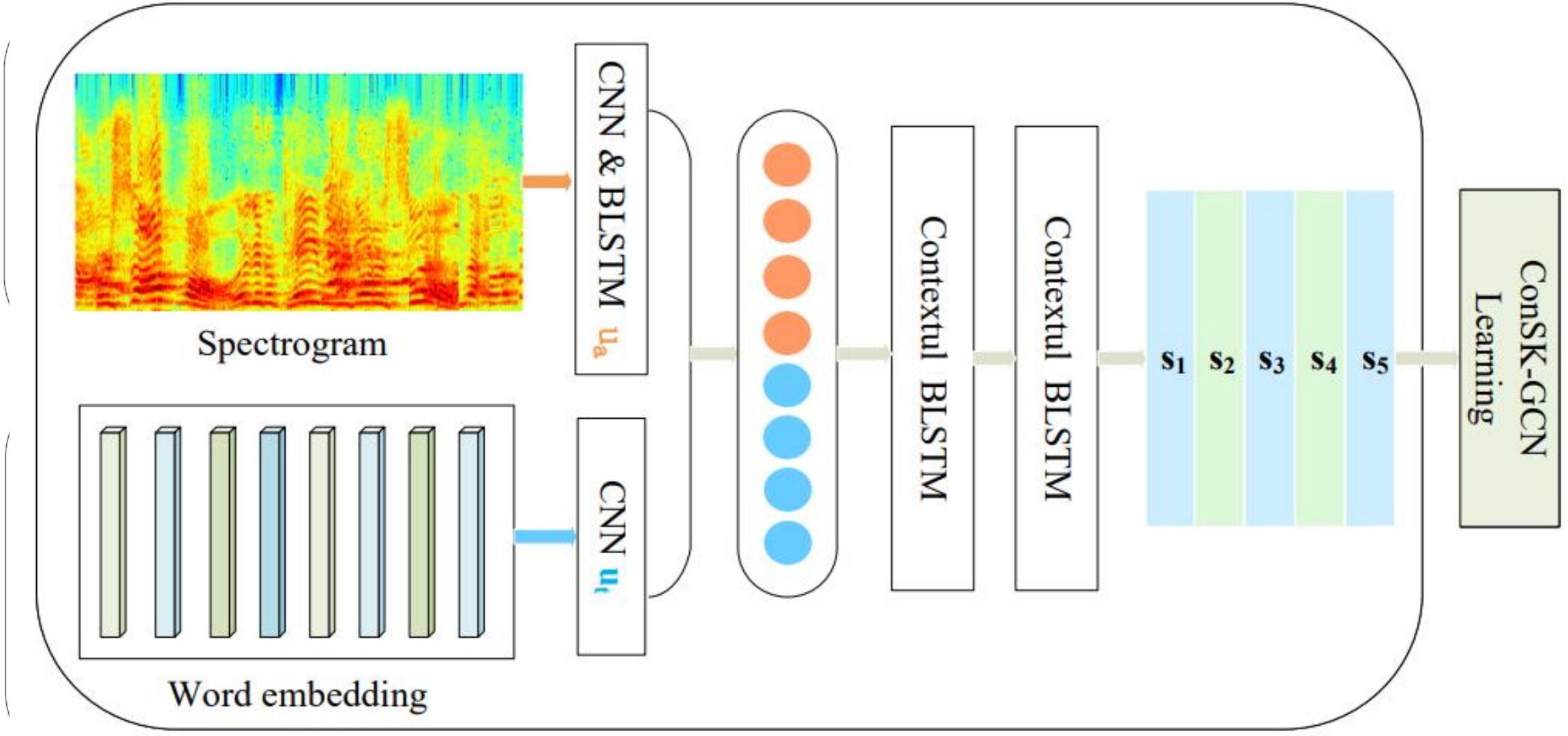
Y. Fu, S. Okada, *et al*., CONSK-GCN: Conversational Semantic-and Knowledge-Oriented Graph Convolutional Network for Multimodal Emotion Recognition, IEEE ICME, pp.1-6, 2021

Y. Fu, S. Okada, *et al*., "Context- and Knowledge-Aware Graph Convolutional Network for Multimodal Emotion Recognition," in *IEEE MultiMedia*, doi: 10.1109/MMUL.2022.3173430.

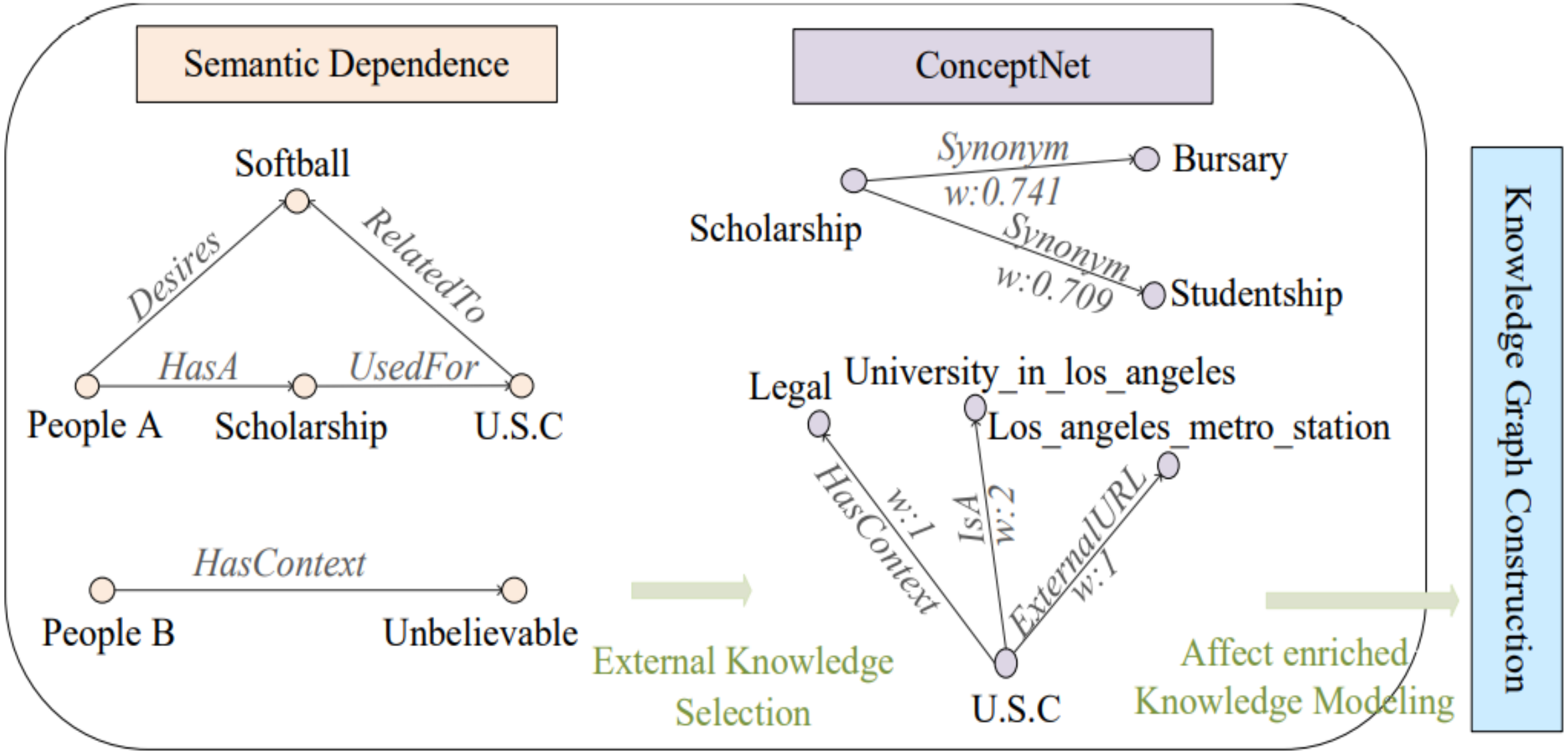
外部知識の利用に基づく感情推定



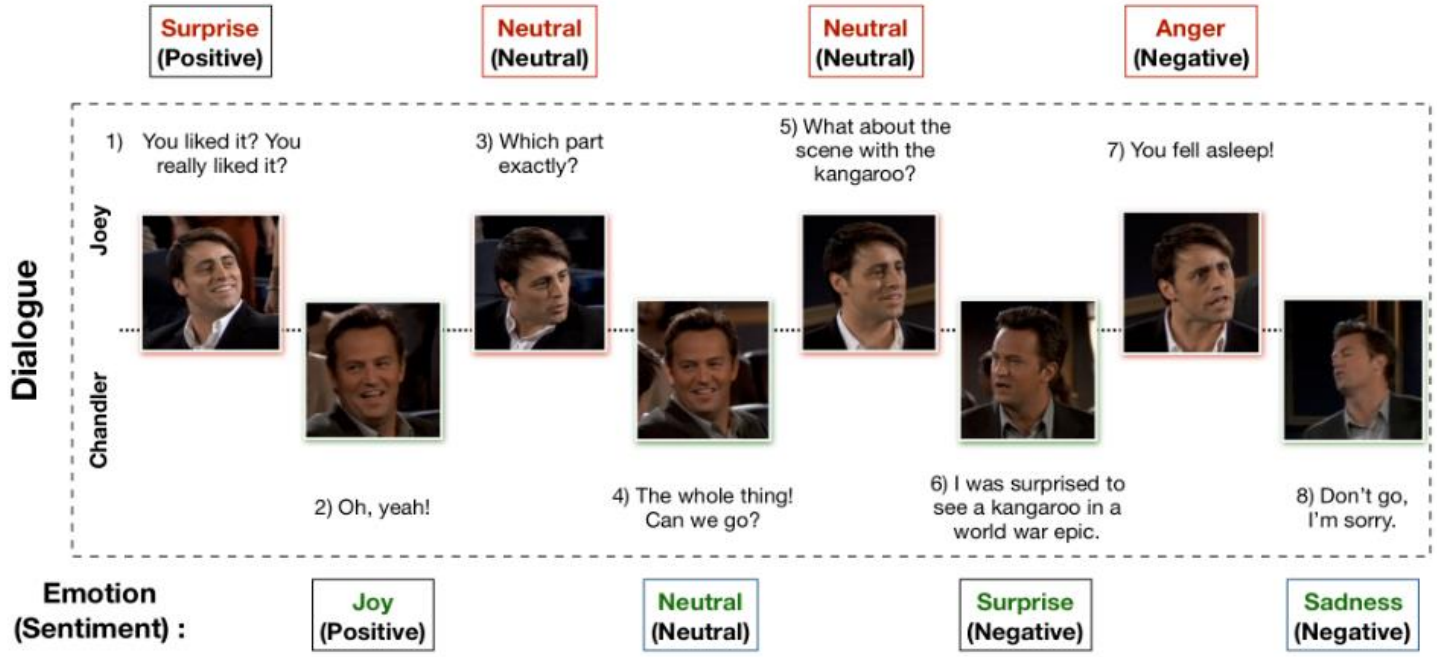
外部知識の利用に基づく感情推定



外部知識の利用に基づく感情推定



MELD: Multimodal Emotion Lines Dataset を用いた評価



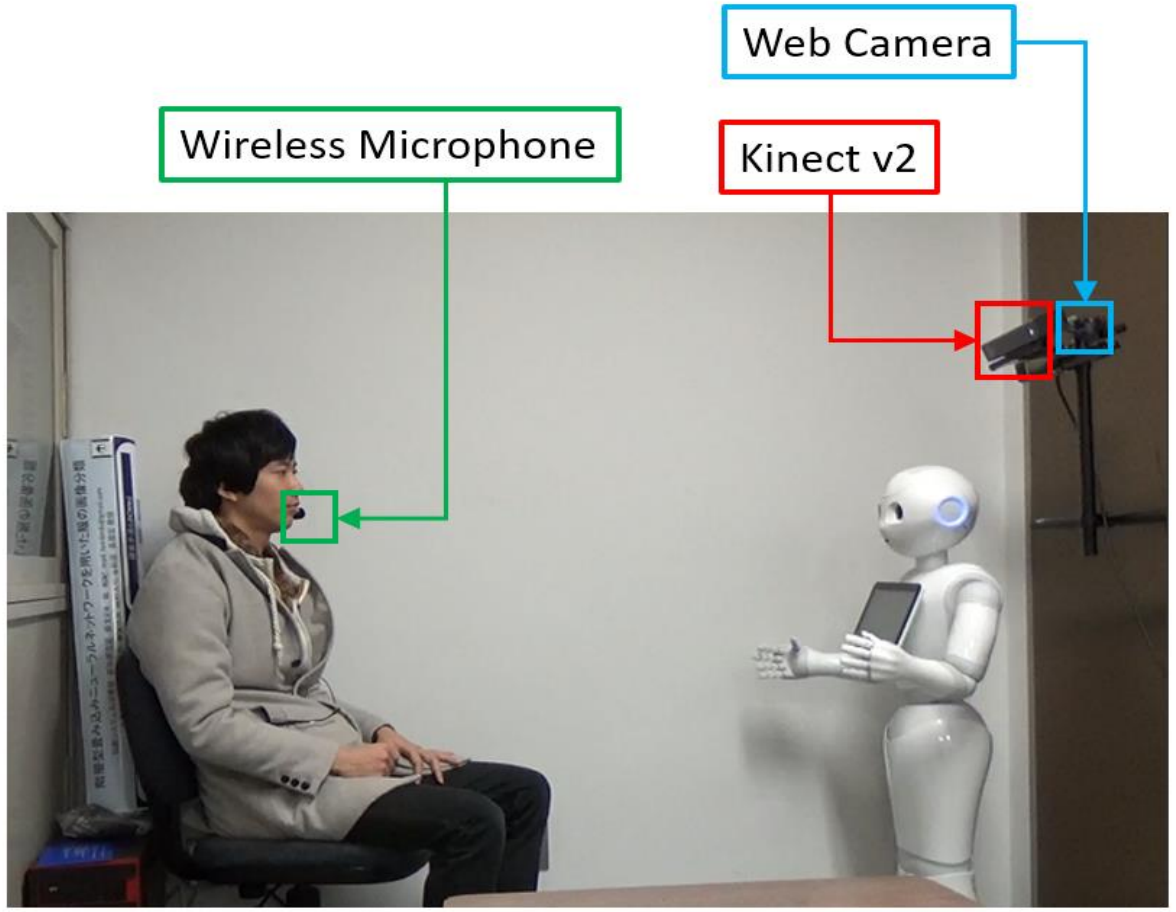
S. Poria, D. Hazarika, N. Majumder, G. Naik, R. Mihalcea, E. Cambria. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversation. (2018)

Models	Neutral	Anger	Disgust	Joy	Surprise	Sadness	Fear	W-F1
CNN [4]	67.3	12.2	0.0	32.6	45.1	19.6	0.0	45.5
LSTMs [5]	67.6	12.3	0.0	36.0	45.7	17.2	0.0	46.0
bc-LSTM [6]	77.0	38.9	0.0	45.8	47.3	0.0	0.0	54.3
DialogueRNN [9]	73.7	41.5	0.0	47.6	44.9	23.4	5.4	55.1
DialogueGCN [10]	-	-	-	-	-	-	-	58.1
ConS-GCN	77.0	50.3	**2.9**	58.8	59.1	35.8	0.0	62.0
ConK-GCN	**80.0**	51.6	0.0	56.3	58.1	35.1	**13.7**	61.9
ConSK-GCN (Ours)	78.1	**54.1**	0.0	**61.1**	**61.0**	**36.9**	10.5	**63.8**

Agenda

- 内面状態の推定技術の背景と周辺
- 内面状態の推定における課題と研究紹介
- **内面状態を推定することの効用とは？**

発話意欲を推定して適応するインタビューロボット



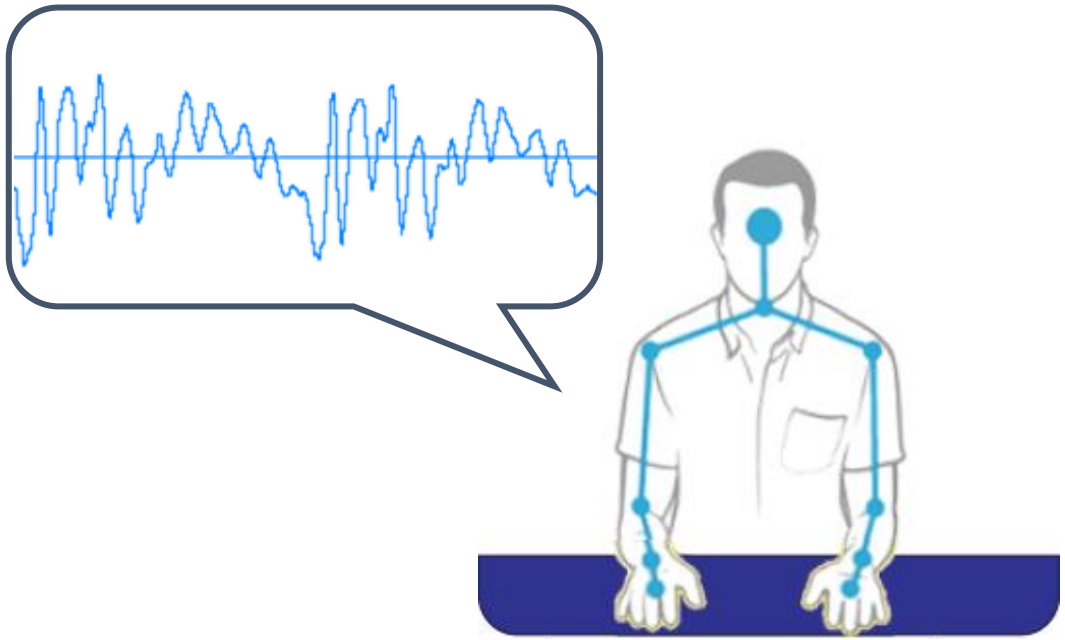
- 相手の**話す際の振る舞い**
  (身振りや音声)から発話意欲を推定
- 推定発話意欲から次の質問を
  適応的に選択

Fuminori Nagasawa, Shogo Okada, Takuya Ishihara, and Katsumi Nitta,
Adaptive Interview Strategy Based on Interviewee's Speaking Willingness Recognition for Interview Robots, under review.

発話意欲推定 & インタビュー戦略切替

音声（声の韻律など）
姿勢（身振り手振りなど骨格情報）

入力変数



意欲推定

150サンプル程度でRFを訓練

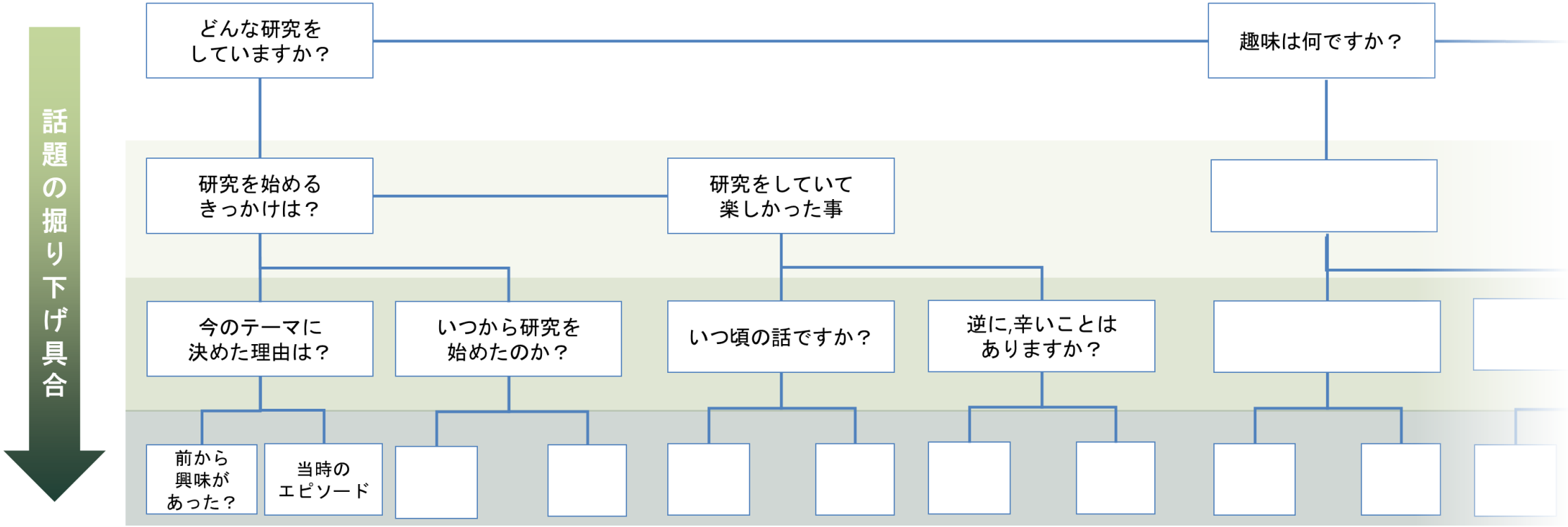
72.79%で発話意欲を推定

推定発話意欲

出力変数

戦略切替

推定した発話意欲によるインタビュー戦略切替

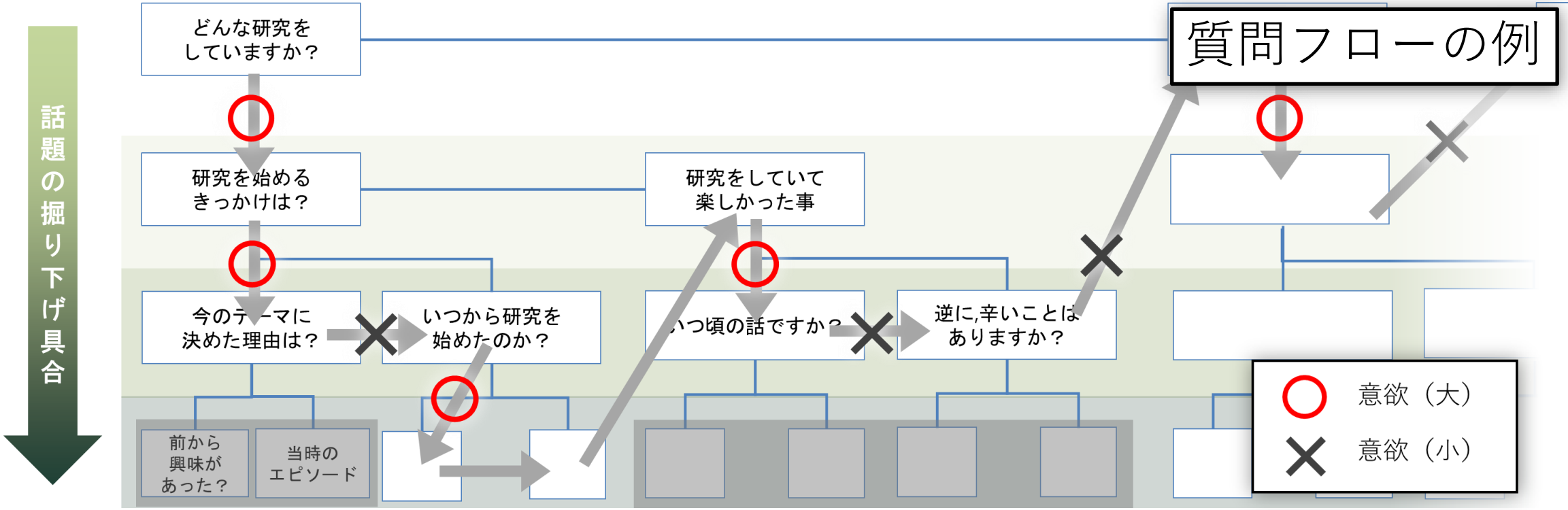


インタビュー中に意欲推定を行い，それに応じて次の質問を選択

- 意欲(大)　→　より掘り下げた質問をする
- 意欲(小)　→　違う話題に切り替える

など…

推定した発話意欲によるインタビュー戦略切替



インタビュー中に意欲推定を行い，それに応じて次の質問を選択

- 意欲(大)　→　より掘り下げた質問をする
- 意欲(小)　→　違う話題に切り替える

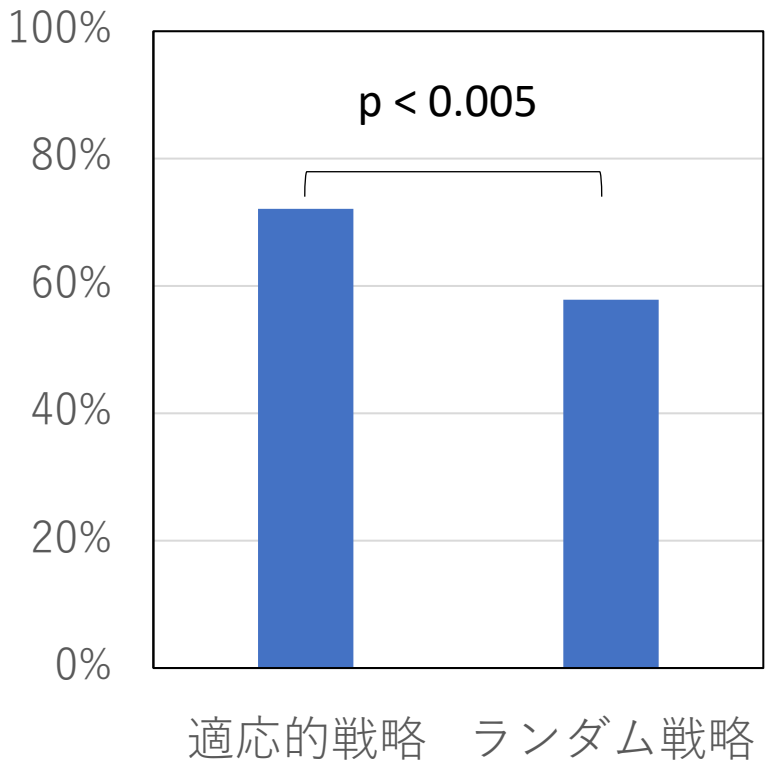
など…

インタビュー対話による評価実験

- 30人の被験者(20～60代の一般男女)との対話実験
- 二つの話題(スポーツ・趣味・勉強・研究・仕事・育児の中から選択)について，インタビュー対話を実施
- ランダム戦略　(意欲推定なし)と適応戦略　(意欲推定あり)で比較
- 片方の話題で意欲推定戦略を使用
  - 順番はランダム
- 意欲推定以外のモジュールはWoZで制御（発話開始・終了）
- 対話終了後
  - アンケートによる主観評価
  - 本人による意欲あり区間のアノテーション(正解データの作成)

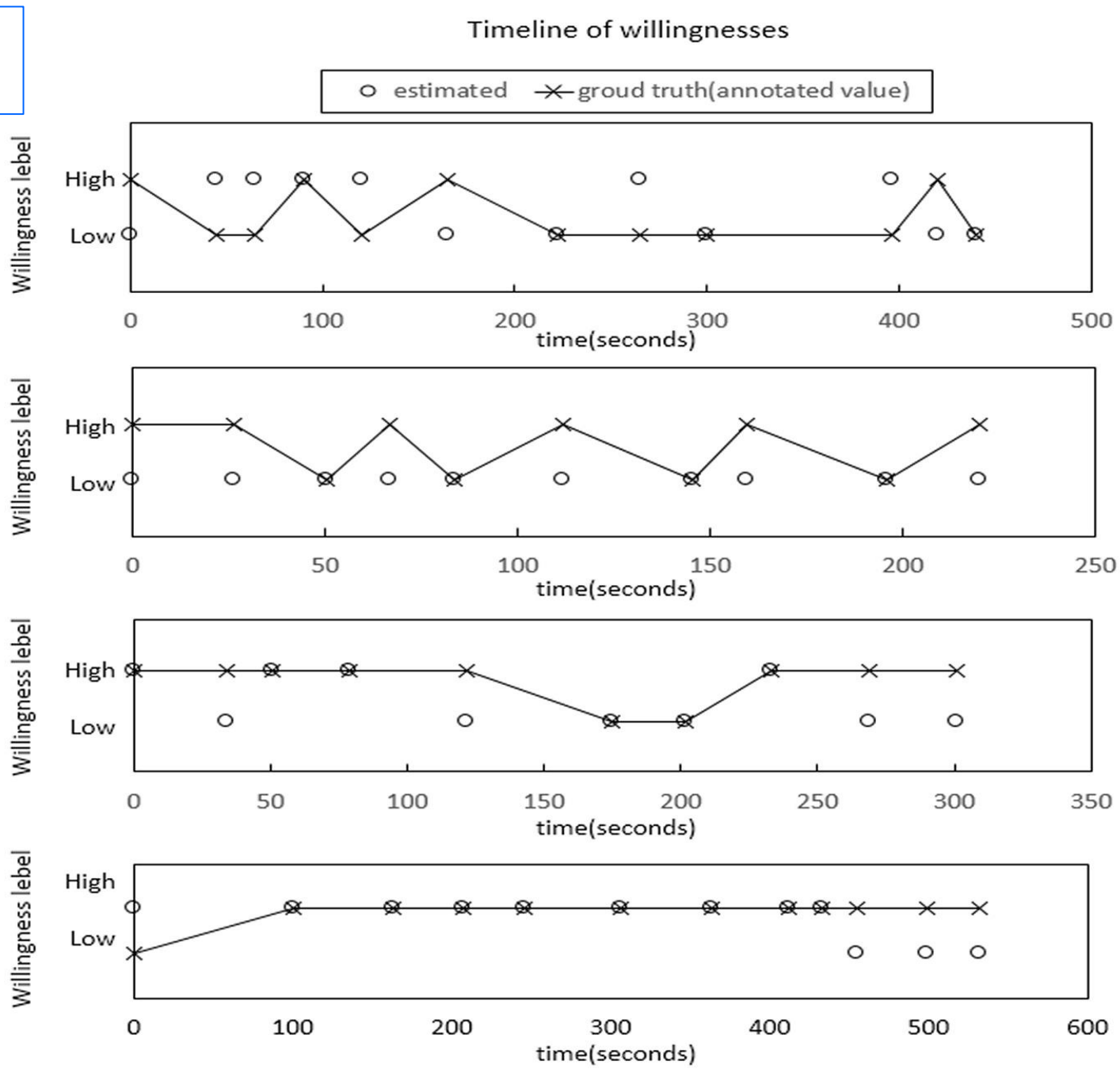
意欲推定の有/無による意欲発話割合の調査

- 対話全体に占める「意欲あり」発話の割合を比較



- 発話意欲ありの場合に，対話全体における意欲的発言割合が向上
- 内面状態の推定精度が高くなくとも，推定モデルが役に立つ一例

推定精度	意欲あり発言割合	傾聴姿勢（↓ good）	不適切質問 (↑good)
36%	33%	4	3
50%	63%	1	4
60%	80%	1	5
66%	92%	2	4



今後の課題

- 個人差・コーパスの差のモデル化
  - 個人差の定式化（感情表現Xの差，自己報告ラベルYの差，写像関数f(x)の差）
  - 共通部分と差異の解明
  - 転移学習モデルの実現
- 内面状態の推定の応用
  - 対話システムによる適応
    - 楽しそうなので，さらに話題を掘り下げる
  - フィードバック
    - ユーザが自信がなさそうな箇所を可視化
- アクティブセンシング
  - 内面状態を発露させるシステム側のインタラクションの探求

謝辞

共同研究者の皆様のご協力に感謝いたします．

研究室学生（紹介した論文の筆頭著者）

- 平野裕貴，堅田俊，Fu Yahui, Gao Yuan，長澤史記

共同研究者

- 駒谷和範 教授（大阪大），中野有紀子 教授（成蹊大），
- 新田克己教授（東工大），西田豊明 教授 (福知山大)，Candy Olivia Mawalim 助教（JAIST）
- 北陸先端大 SSILab，東工大 新田研，京大 西田研の皆様
- 企業共同研究者の皆様