

令和 6 年 6 月 5 日現在

機関番号：17104

研究種目：基盤研究(C) (一般)

研究期間：2020～2023

課題番号：20K12110

研究課題名(和文) グループディスカッションの自動評価のためのマルチモーダル情報理解

研究課題名(英文) Automatic evaluation of group discussion based on multi-modal interpretation

研究代表者

嶋田 和孝 (Shimada, Kazutaka)

九州工業大学・大学院情報工学研究院・教授

研究者番号：50346863

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：教育において、ディベートなどを対象とした場合は、その議論を評価する必要がある。本研究課題では、議論の質評価のためのデータセットを構築し、そのデータにマルチモーダルな評価値推定モデルを適用し、その有効性を検証した。また、作成したデータだけではなく、既存のデータセットを対象とし、議論分析に関係する様々な要素技術についても、手法の提案および評価を行った。本研究課題で作成したデータセットは、Kyutech Debateコーパス Ver.1およびVer.2、以前の科研で作成したKyutechコーパスのVer.3の3つは、いずれもWebで公開をしている。

研究成果の学術的意義や社会的意義

本申請課題で作成したデータは基本的にすべて無償で申請者のWebページに公開している。これらのデータは関連研究者が自由に利用することができ、学術的な意義がある。本申請課題で対象としている議論の評価は人間でさえも評価がぶれ、公平性などの様々な問題が生じる。この問題に対して、機械による客観的な評価が可能であれば、一定の意義がある。これは、ディベートや小論文などの自動評価という観点で社会的な意義がある。

研究成果の概要(英文)：In education, active learning, such as PBL, in which multiple people discuss issues for which there are no clear answers, has gained importance in recent years. We have studied the methods for understanding multi-party discussions and the summarisation techniques. On the other hand, evaluating debate discussions is necessary in education. However, it is not easy for evaluators to assess the quality and content of debates. In this research project, we created a dataset for evaluating the quality of debate and proposed a multimodal estimation model for estimating it. We also proposed and evaluated methods for various elemental techniques related to discussion analysis, not only on the created data but also on existing datasets (AMI corpus and our corpus created in the past). In this research project, we created three datasets, and the datasets are available on the web.

研究分野：自然言語処理

キーワード：マルチモーダル 自然言語処理 議論分析

1. 研究開始当初の背景

教育において、明確な答えのない課題について、複数人で議論する PBL のような能動的な学習が近年重要視されている。また、入試などでも単なる知識量を測る筆記試験のみならず、課題解決型のグループディスカッションが導入されるケースが目立ってきた。このような背景のもと、本研究では 4 人の参加者が 2 組に分かれディベートを行う場面を主な対象とし、そのディベートの自動評価などについての研究を行った。また、ディベートのみならず、既存の 4 人一組議論における議論内容の理解や分析、議論などに関連する種々の言語現象についての分析、実装、評価などを行った。

2. 研究の目的

我々は、これまでに複数人議論における場の理解や議論の要約技術などについて研究してきた。一方で、ディベートなどを対象とした場合は、その議論を評価する必要がある。しかし、議論の質や内容を評価者が評価することは容易ではない。また、評価者によっては評価の軸が異なり、公平に評価することが難しい場合もある。これらの問題を解決するためにマルチモーダル情報理解に基づく議論対話の自動評価システムの構築を目的とする。



図 1. Kyutech Debate コーパスの環境

3. 研究の方法

研究の方針としては、(1) 対象となるディベートデータ (コーパス) の作成、(2) そのコーパスを利用した分析やモデルの構築、(3) 分析モデルなどを支える自然言語処理の要素技術の検証、の大きく 3 つの柱がある。

(1) コーパス作成

対話内容を理解するシステムの開発には、その分析のための対話コーパスが不可欠である。さらに、海外には AMI コーパスのような自由に利用できるコーパスが存在するが、日本語を対象とした自由に利用できるコーパスは多くない。我々は 4 人一組による議論を収録した Kyutech コーパスを作成した経験がある。この作成の知見に基づき、本課題ではディベートを対象とした Kyutech Debate コーパスを作成した。

Kyutech Debate コーパスの議論環境を図 1 に示す。コーパスは 2 人一組に分かれたディベート対話とその後の合意形成対話の 2 つが含まれており、合計で 10 対話 (ディベート 5 対話・合意形成 5 対話) が含まれている。このデータに対して、(1a) データの書き起こしと評価値のラベリングと (1b) 発話意図ラベルの付与を行い、Web 上で公開している。

評価値のラベリングは、一定のトピック単位でまとめたセグメントと呼ばれる単位で行われており、評価値推定のための音声データの分析値や OpenPose および OpenFace を用いた画像解析結果も要望があれば公開できるよう準備されている。

(2) 議論分析モデル

議論分析としては、まず、前述の Kyutech Debate コーパスを対象とした評価値推定モデルを実装し、その有効性を検証した。提案手法として、言語、音声、画像のそれぞれを組み合わせたマルチモーダルなモデルを提案した。図 2 に階層型 LSTM モデルに基づく、議論評価値推定モデルを示す。

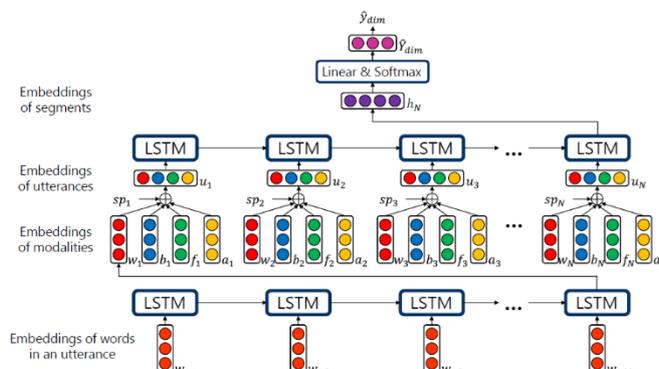


図 2. 階層型 LSTM. 赤が言語特徴、青が身体特徴、緑が顔特徴、黄色が音声特徴を意味する。

議論の分析技術としては、上記の議論評価値推定だけではなく、発話間の推定モデルや対話内での話者の役割（リーダ分析や議論を阻害する参加者の検出など）を推定、分析する技術についても提案した。

さらにより直接的に議論を分析するために、発話内容の可視化に基づく分析モデルを提案し、その有効性を検証した。図3に可視化ツールの例を示す。議論中の発話のトピックをクラスタリングし、話者ごとの発話量などと合わせて可視化することで、どの話者がどのような傾向の発言をしたかなどを定量的に分析ができるようになった。本発表はHCGシンポジウム2021で学生優秀インタラクティブ発表賞を受賞した。

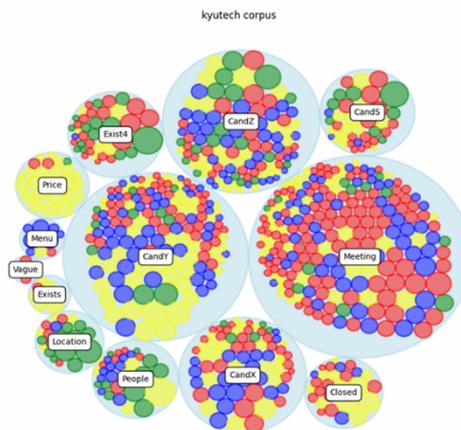


図3. 可視化モデル. 発話のトピック分布や発話量を可視化する。

(3) 関連技術の検証

前項の議論分析は、議論そのものの質や内容の評価であった。一方で、幅広い議論の分析をするためには、議論中に存在する様々な言語現象に対する要素技術が必要となる。本課題では、発話内容については、攻撃性のある発言の検出や嘘の検出を行い、入力部分として音声誤りの検出や非流暢性箇所（フィラーや言い淀み）の検出などについて取り組んだ。また、対象は対話ではなく、エッセイデータであるが、エッセイにおける論理性の判断をする手法の提案も行った。

一例をあげると、音声誤りの検出では、対話には何らかのトピックがあるという仮定を置き、その対話のトピックを表すトピックモデル（LDA）と誤り検出にファインチューニングされたBERTを併用する手法を提案した。図4に概要図を示す。BERTの出力する各トークンの埋め込み表現とLDAの確率分布をベクトルと見立てたものを結合し、系列解析モデルのCRFで判定するモデルを構築した。

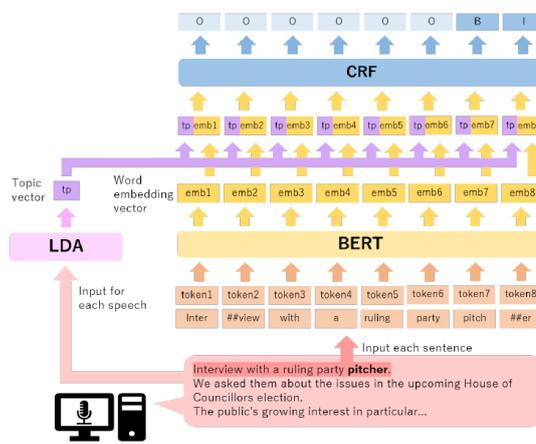


図4. 音声誤り検出にトピック情報を利用するモデル。

4. 研究成果

(1) コーパス作成

コーパス作成に関しては、Kyutech Debate コーパスの公開（Ver1：評価値付与データおよびVer2：発話意図付与データ）をWebで公開している。また、以前の科研費で作成したKyutechコーパスについても拡張を行い、トピック単位の正解要約の作成も行った。Ver3として公開している。

(2) 議論分析モデル

マルチモーダルな議論評価値推定モデルでは、言語的な特徴が明確に有効な「合理性」という評価尺度ではテキスト情報のみでも比較的高い精度が得られていたが、「有効性」という尺度では入力モダリティが音声・画像と増えるほど精度が向上する傾向がみられ、提案手法（H-LSTM）の有効性が示された。表1に実験結果を示す。

他の議論分析タスクにおいて、発話間の関係の有無を識別するモデルでは、広く使われる事前学習モデルBERTと比較して、BERTに事後処理を合わせた二段階手法（表2における+Bottom-up）のほうが高い精度になることを確認した。

(3) 関連技術の検証

研究成果の一例として、エッセイにおける論理性の判断をする手法について取り上げる。この実験では岡山大学が公開している小論文データを対象とし、「論理性」と呼ばれる評価尺度に対しての値推定を行った。手法として、Handcraftedな特徴量（たとえば、小論文中の文字数など）を20種類とBoW特徴量を用意し、それをランダムフォレスト回帰で学習したもの、事前学習モデルBERTで学習したもの、BERTと特徴量を組み合わせたハイブリッドモデルの3種類を比較した。実験結果を表3に示す。QWKと呼ばれる評価尺度で、ハイブリッドモデル（BERT+BoWやBERT+20features+BoW）が高い精度を得ていることがわかる。ここで、人間による評価についてQWKを計算したところ、その値が0.561であり、提案手法は人間による評価よりも良い精度であることが分かり、機械学習モデルによる小論文評価の可能性を見いだすに至った。

表 1. 議論評価値推定モデルの実験結果. Tはテキスト, Bは身体(Body), Fは顔, Aは音声特徴を意味する. 例えば, TFA はテキストと顔情報, 音声情報を組み合わせたもの.

評価軸	モデル	評価性能 (マイクロ平均の平均値)							
		T	TB	TF	TA	TBF	TBA	TFA	TBFA
Re	baseline	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333
	SVM	0.451	0.338	0.337	0.343	0.333	0.317	0.320	0.340
	LSTM	0.387	0.398	0.392	0.380	0.410	0.379	0.388	0.360
	A-LSTM1	0.412	0.392	0.387	0.399	0.371	0.359	0.398	0.398
	A-LSTM2	0.371	0.337	0.400	0.420	0.404	0.374	0.387	0.339
	HLSTM	0.359	0.354	0.459	0.388	0.415	0.391	0.370	0.405
Ef	baseline	0.384	0.384	0.384	0.384	0.384	0.384	0.384	0.384
	SVM	0.459	0.382	0.383	0.436	0.384	0.392	0.406	0.379
	LSTM	0.428	0.478	0.438	0.476	0.467	0.472	0.486	0.435
	A-LSTM1	0.433	0.470	0.426	0.468	0.450	0.396	0.444	0.490
	A-LSTM2	0.452	0.412	0.416	0.436	0.443	0.425	0.401	0.404
	HLSTM	0.459	0.433	0.416	0.379	0.414	0.440	0.431	0.451

表 2. 二段階モデル (+Bottom-up) による発話関係認識モデルの実験結果.

Model	関係あり			関係なし		
	P	R	F1	P	R	F1
<i>BERT</i> のみ	0.15	0.65	0.24	0.93	0.54	0.68
+ <i>Bottom-up</i> _{softmax}	0.40	0.31	0.35	0.92	0.95	0.93
+ <i>Bottom-up</i> _{weight}	0.56	0.32	0.41	0.92	0.97	0.95

表 3. 小論文評価データにおける機械学習モデルの比較.

手法	QWK
ランダムフォレスト回帰 (20 features)	0.496
ランダムフォレスト回帰 (BoW)	0.603
ランダムフォレスト回帰 (20 features +BoW)	0.608
BERT	0.596
BERT +20 features	0.587
BERT +BoW	0.632
BERT +20 features +BoW	0.622

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Shunsuke Yonemitsu and Kazutaka Shimada	4. 巻 8
2. 論文標題 Recognizing a participant behavior in a multi-party conversation: detection of a participant that returns to a discussion that is already over	5. 発行年 2022年
3. 雑誌名 Information Engineering Express	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.52731/iee.v8.i1.635	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計18件（うち招待講演 0件 / うち国際学会 9件）

1. 発表者名 Tsukasa Shiota and Kazutaka Shimada
2. 発表標題 Annotation and multi-modal methods for quality assessment of multi-party discussion
3. 学会等名 Proceedings of PACLIC36（国際学会）
4. 発表年 2022年

1. 発表者名 Zhenming Li and Kazutaka Shimada
2. 発表標題 Combining Pre-Trained Language Models and Features for Offensive Language Detection
3. 学会等名 Proceedings of IIAI AAI 2022-Winter（国際学会）
4. 発表年 2022年

1. 発表者名 中島 寛人, 嶋田 和孝
2. 発表標題 前後文脈を用いた対話文の言い淀み検出
3. 学会等名 電子情報通信学会 NLC研究会
4. 発表年 2022年

1. 発表者名 中島 寛人, 嶋田 和孝
2. 発表標題 対話文における双方向文脈補完を用いた言い淀み検出
3. 学会等名 電子情報通信学会九州支部, 第30回学生会講演会
4. 発表年 2022年

1. 発表者名 川崎 慎乃介, 嶋田 和孝
2. 発表標題 機械学習モデルを用いたKyutechコーパスのトピック分類
3. 学会等名 電子情報通信学会 NLC研究会
4. 発表年 2022年

1. 発表者名 米満 駿甫, 嶋田 和孝
2. 発表標題 複数の議論コーパスを利用した談話行為推定
3. 学会等名 2022年度 人工知能学会全国大会 (第36回)
4. 発表年 2022年

1. 発表者名 Takumi Himeno and Kazutaka Shimada
2. 発表標題 Discussion Structure Prediction Based on Two-step Method
3. 学会等名 Proceedings of Recent Advances in Natural Language Processing, pp. 543-551 (国際学会)
4. 発表年 2021年

1. 発表者名 Yuri Nakayama, Tsukasa Shiota, Kazutaka Shimada
2. 発表標題 Corpus construction for topic-based summarization of multi-party conversation
3. 学会等名 Proceedings of the International Conference on Asian Language Processing (IALP), pp. 229-234 (国際学会)
4. 発表年 2021年

1. 発表者名 西山 空良, 嶋田 和孝
2. 発表標題 議論の分析とファシリテーションのための可視化ツールの構築
3. 学会等名 電子情報通信学会, HCGシンポジウム2021, 1-2-2
4. 発表年 2021年

1. 発表者名 中山 友梨, 塩田 宰, 嶋田 和孝
2. 発表標題 複数人対話におけるトピック単位の要約データの構築とその要約
3. 学会等名 電子情報通信学会, NLC研究会, NLC2021-4, pp. 19-24, 2021
4. 発表年 2021年

1. 発表者名 Tsukasa Shiota, Kouki Honda, Kazutaka Shimada, and Takeshi Saitoh
2. 発表標題 Leader Identification Using Multimodal Information in Multi-party Conversations
3. 学会等名 Proceedings of the International Conference on Asian Language Processing (IALP), P1-2, pp. 7-12 (国際学会)
4. 発表年 2020年

1. 発表者名 Tsukasa Shiota and Kazutaka Shimada
2. 発表標題 The Discussion Corpus toward Argumentation Quality Assessment in Multi-Party Conversation
3. 学会等名 Proceedings of the 9th International Conference on Learning Technologies and Learning Environments (国際学会)
4. 発表年 2020年

1. 発表者名 Taskumi Himeno and Kazutaka Shimada
2. 発表標題 Relation Identification Using Dialogical Features in Multi-Party Conversation
3. 学会等名 Proceedings of the 8th International Symposium on Applied Engineering and Sciences, C-02-02 (国際学会)
4. 発表年 2020年

1. 発表者名 Shohei Takabatake, Kazutaka Shimada, and Takeshi Saitoh
2. 発表標題 Detection of Lying Situations in Liar Corpus
3. 学会等名 Proceedings of the 8th International Symposium on Applied Engineering and Sciences, C-02-04 (国際学会)
4. 発表年 2020年

1. 発表者名 Shunsuke Yonemitsu and Kazutaka Shimada
2. 発表標題 Don't beat a dead horse: Recognizing a person who returns to a done-deal in a multi-party conversation
3. 学会等名 Proceedings of the 8th International Conference on Smart Computing and Artificial Intelligence (国際学会)
4. 発表年 2020年

1. 発表者名 塩田 宰, 嶋田 和孝
2. 発表標題 マルチモーダル情報を用いた複数人議論の品質評価
3. 学会等名 人工知能学会, 第91回 言語・音声理解と対話処理研究会, pp. 116-121
4. 発表年 2021年

1. 発表者名 姫野拓未, 嶋田和孝
2. 発表標題 複数人議論における発話間の関係を対象とした関係分類
3. 学会等名 言語処理学会第27回年次大会, NLP2021, P5-11
4. 発表年 2021年

1. 発表者名 荻野 奈津実, 姫野 拓未, 嶋田 和孝
2. 発表標題 複数人議論における対話的役割分類モデルの比較
3. 学会等名 火の国シンポジウム2021, A1-3
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Kyutechコーパス/Kyutechディベートコーパス
<http://www.pluto.ai.kyutech.ac.jp/~shimada/resources.html>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------