

複数人議論における議論構造の推定

姫野 拓未[†] 嶋田 和孝[†]

[†]九州工業大学大学院 情報工学府 〒820-8502 福岡県飯塚市川津 680-4

E-mail: †{t_himeno,shimada}@pluto.ai.kyutech.ac.jp

あらまし 研究室や会議室では、会議が日々行われている。そこで行われた議論内容を会議に参加していない人に共有するためには、議事録が不可欠である。議事録を議論構造として表すことができれば、発話間の関係を捉えることができ、議論全体の流れや論点を把握することができる。今までに、私たちは議論構造の中でも発話間の関係を教師あり学習モデルを用いて推定した。しかし、議論構造の構築において関係のない発話ペアを多く抽出しており、議論構造の特徴を考慮できていない。そこで、本研究では議論構造の特徴に着目して精度向上を目的とした手法を提案する。まず、使用したデータセットについて説明する。次に、議論構造の特徴に着目した選択手法を提案する。そして、以前の手法と本研究で提案した手法の推定精度の比較を行い、提案した手法の有効性を確認する。実験の結果、提案した手法が以前の手法よりも良い精度であることを確認した。

キーワード 複数人議論, 議論マイニング, 議論構造, 構造推定

Predicting Discussion Structures in Multi-Party Conversations

Takumi HIMENO[†] and Kazutaka SHIMADA[†]

[†] Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

E-mail: †{t_himeno,shimada}@pluto.ai.kyutech.ac.jp

Abstract Conversations are often held in laboratories and companies. To grasp the content of a discussion, a summary is important for people who could not attend the conversation. If the summary is illustrated as an argument structure, it is helpful to grasp the essentials of the conversation immediately. We have predicted a link between nodes that consist of some utterances using models based on supervised learnings in our previous work. However, the models could not consider the feature of an argument structure because they predicted many not-links between nodes for the construction of an argument structure. In this paper, we propose a method focusing on the feature of an argument structure. First, we explain a multi-party conversational corpus. Next, we propose a method for selecting edges based on the feature of an argument structure. Finally, we report the result of an experiment in which we compared the previous method and the proposed method. The result showed that the proposed method was more accurate than the previous method.

Key words Multi-Party Conversation, Argument Mining, Argument Structure, Structure Prediction

1. はじめに

研究室や会社では、新しい研究のアイデアや経営戦略を打ち出すために、会議が日々行われている。そこで行われた会議内容を会議に参加していない人に共有するためには、議事録が不可欠である。会議に参加していない人が議事録を読み返すことは議論全体の論点を素早く把握することには適しているが、会議で結論に達するまでの議論の流れを把握することは難しい。そこで、会議の議事録を木構造(議論構造)として表すことができれば、発話間の関係を捉えることができ、結論に達するまで議論の流れや議論全体の論点を素早く把握することを支援でき

ると考えられる。

文章を構造化するタスクの一つとして議論マイニングがある[1]。議論マイニングとは、例えば小論文のような文章を入力文とし、主張を表す文とその主張を支持したり反論したりする文を自動的に判別するタスクである。議論マイニングを構成する要素として要素抽出、要素分類、関係抽出、関係分類の4つのサブタスクがある。要素抽出は、文章中の論理要素に関係する文や節から成り立つ論理要素を抽出するタスクである。要素分類は、抽出された論理要素に対して主張や前提のような役割のラベルを割り当てるタスクである。さらに、関係抽出は抽出された2つの論理要素に対して関係の有無を推定するタスクで

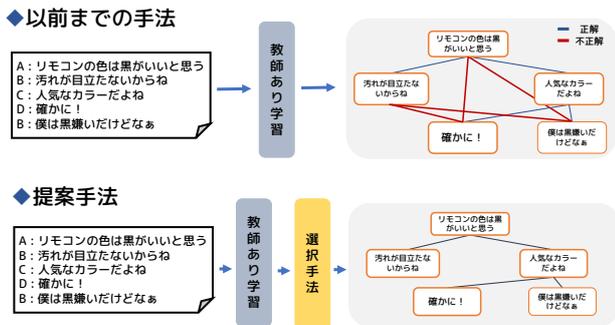


図1 本研究の概要図

ある。最後に、関係分類は関係抽出で得られた論理要素のペアを対象に、主張に対する支持や反論のような関係を持つ役割のラベルを割り当てるタスクである。

これまで複数人議論コーパスを対象に、議論構造の中でも発話間の関係の有無の推定を行う関係抽出を教師あり学習モデルを用いて推定した[2]。しかし、最も良かったモデルの推定結果で再現率は8割程度であるが精度は4割程度しかなく、議論構造において関係のない発話ペアを多く抽出している。この問題点として、子ノードに対して親ノードが一つのみであるという議論構造の特徴が考慮されていないことが挙げられる。そこで、本研究では教師あり学習モデルの出力結果に議論構造の特徴を考慮した選択手法を追加することで精度向上を目指す。本研究の概要図を図1に示す。以前の研究では、図1の上部のように複数人議論の任意の発話ペアを教師あり学習モデルに入力し関係抽出を行うことを目的としてモデルの構築を行った。図1のようにその関係抽出モデルのみでは過剰にエッジを張る傾向がある。本論文では教師あり学習モデルの出力にエッジの選択手法を利用することで、議論構造の特徴を考慮し正解に近い議論構造を構築する。まず、本研究に利用した複数人議論コーパスについて説明する。次に、議論構造の特徴に着目した2つの選択手法について説明する。そして選択手法と以前までの手法との精度比較やそれぞれの手法で構築した議論構造の比較を行う。

2. 関連研究

近年、議論マイニングの研究が盛んに行われている。議論マイニングは、文章を構造化することに適しており、文書要約[3],[4]や自動採点[5]、論文執筆支援[6],[7]、情報検索[8]など数多くの自然言語処理のタスクがある。このタスクの中でも、学生が書いたエッセイに対して議論マイニングに取り組み、要素間の関係抽出に取り組んでいるタスクがある[9]。しかし、対象としている文章が書き言葉であるエッセイでは、文章の形が主張の直後に前提の文があるように決まった形で形式化されており、自由な発言で形式化されていない複数人議論とは特徴が異なる。さらに、関係抽出のタスク後にエッセイの議論構造の構築に着目していない。本研究では、複数人議論特有の特徴に着目したモデルの利用や複数人議論の議論構造に着目したエッジの選択手法を利用し、議論構造の構築を行う。

また、議論構造のエッジの選択手法としてグラフの最短経路問題で解く手法が存在する[10],[11]。最短経路問題とは、重み付きグラフの与えられた2つのノード間を結ぶ経路の中で、重みが最小の経路を求める最適化問題である。最短経路問題を複数人議論の議論構造に用いるためには、議論中の最初の発言となる親ノードから複数存在する末端ノードまでの全体のパスに適応する必要がある。しかし、複数人議論の特徴である子ノードに対して親ノードが一つのみ存在する条件を満たさない場合が存在するため適していない。そのため、本研究では複数人議論の特徴を考慮するために局所的な発話ペアについてエッジの選択手法を提案する。

さらに、議論の可視化に着目した研究も多く存在する[12],[13]。これらの研究では議論中における話者の感情や議論の論点を長所と短所ついて可視化を行っている。しかし、本研究の目的とした結論に達するまでの流れをこれらの可視化から確認することはできない。そこで本研究では、議論構造を構築することにより議論の結論に達するまでの流れを可視化することを目的として実験を行う。

3. データセット

複数人議論の議論構造を推定するためには、発話間の関係などの議論構造の情報がアノテーションされた複数人議論コーパスが必要となる。そこで、本研究では複数人議論コーパスのひとつであるAMI corpusを用いる[14]。AMI corpusはあらかじめ課題が与えられている状態で行われているシナリオ会議を収録したコーパスであり、書き起こしデータが公開されている。議論の設定として、架空の家電企業に勤める異なる役職の4人の従業員が、市場に出回っているユーザーに不親切で魅力の無い古いデザインのテレビリモコンの代わりとなる新しいテレビリモコンを開発するという議論であり、計4回行われる。AMI corpusの各発話には話者IDや時間情報に加えて、談話行為タグが付与されている。談話行為タグとは、各発話がどのような機能を持つかを示すタグである。何らかの情報を提供している発話に付与される「Inform」や相槌を表す発話に付与される「Backchannel」になど計15種類の談話行為タグが付与されている。

また本研究では、議論構造を推定するために、Twente Argument Schema (TAS)を用いる[15]。TASはシナリオ会議の議論で生じる議論構造を明らかにするために作成されたアノテーションスキーマである。TASにおける議論構造は2つの要素から成り立っている。ある話者の発話全体、また発話の一部、複数の発話から成り立っているノードと、そのノード間を結ぶエッジから成り立っている。また、ノードがどのような役割であるかを表すUnit Labelが付与されている。肯定的な主張を表す「Statement」や回答に対して選択肢を与える質問を表す「A/B Issue」など計5種類のUnit Labelが付与されている。さらに、TASでは一つの会議(ダイアログ)の中である議題が挙げられてから、その議題が結論に達した時やその議題を途中で中断した

(注1) : <http://groups.inf.ed.ac.uk/ami/download/>

り放棄したりするまでをディスカッションと定義されている。各ディスカッションは一つのダイアログの中に一つまたは複数存在し、各ノードは一つのディスカッションに複数存在する。また、各ノードには一つの Unit Label が付与されている。TAS を用いた関係抽出の例の詳細は以前の研究の論文に示している [2]。

4. 提案手法

本節では、発話間の関係の有無を推定した結果に対して精度向上を目的とした提案手法について説明する。4.1 節では、先行研究で使用した教師あり学習モデルである SVM と新たに用いた BERT の 2 つのベースモデルについて説明する。4.2 節では、2 つの教師あり学習モデルが推定した結果を用いて議論構造の特徴を考慮した 2 つの手法を提案する。

4.1 ベースモデル

本節では、発話間の関係の有無を推定する 2 種類のモデルについて説明する。

4.1.1 SVM

SVM による発話間の関係の有無の推定を行うモデルについて説明する。モデルの概要図を図 2 に示す。モデルの入力として、ディスカッションの内の任意のノードペアである $node_i = (w_1, w_2, \dots, w_k)$, $node_j = (w_1, w_2, \dots, w_l)$ と以前の研究で提案した 8 種類の素性を入力とする。ここで w はノード内の単語を表しており、 k と l はノードの単語長を表している。まず、ノードペアについて説明する。ノード内の単語を Google が公開している Word2Vec² を用いて得られた単語ベクトル $V_i = (v_1, v_2, \dots, v_l)$, $V_j = (v_1, v_2, \dots, v_l)$ に変換する。ここで v_n は w_n の単語ベクトル、 V はノードベクトルを表す。このとき、NLTK によって選択されているストップワードを除く。そして、ノード内の単語ベクトルの和 V_{node_n} を以下の式から算出する。

$$V_{node_n} = \sum_{x=1}^m v_x$$

ここで v_x は $node_n$ の単語ベクトルを表し、 m はノードの単語長を表している。次に提案素性について説明する。以前の研究 [2] では提案した複数人議論特有の素性を機械学習モデルに利用することで素性が有効であることを確認している。そこで、本研究においても提案素性を SVM に用いる。8 種類の提案素性は以下の通りである。

- ノードペアの単語数
- ノードペアの共通単語
- 話者情報
- 時間情報
- ノード間の距離
- 談話行為タグ
- Unit Label
- ノードペアの極性値

最後に二つのノードベクトルと提案素性を連結したものを SVM に入力し、ノード間の関係の有無を推定する。

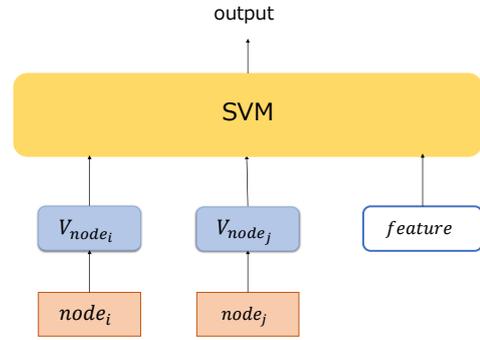


図 2 SVM を用いたモデルの概要図

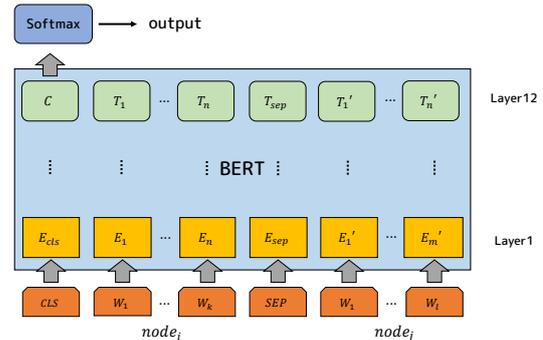


図 3 BERT を用いたモデルの概要図

4.1.2 BERT

Devlin ら [16] が提案した BERT による発話間の関係の有無の推定を行うモデルについて説明する。BERT は大規模なテキストコーパスを用いて事前学習を行った後、タスクごとに fine-tuning する汎用言語モデルである。含意関係認識や感情分析、質問応答などの様々な自然言語処理のタスクにおいて、BERT は高い性能を発揮することが知られている [17], [18]。その BERT モデルの概要図を図 3 に示す。BERT の入力には 4.1.1 節と同様にディスカッションの内の任意のノードペア $node_i, node_j$ と入力の先頭に分類トークン (CLS)、ノードペアの間に二つのノードを区別するためのトークン (SEP) を用いる。出力には BERT の 12 層目の CLS トークンを Softmax 関数に入力し、Softmax 関数から出力された値を利用して発話間の関係の有無を推定する。

4.2 選択手法

本節では、4.1 節で説明した教師あり学習モデルの結果を利用して精度向上を目的とした 2 つのエッジの選択手法について説明する。

4.2.1 Bottom-up 手法

4.1 節で説明した 2 つの教師あり学習モデルの出力結果に対して本手法を用いる。図 4 に本手法の概要図を示す。図 4 の推定結果のように関係抽出モデルから発話間の関係を推定する。まず、発話間の関係の有無の推定結果から議論構造を構築する。次に、議論構造の子ノードに対して親ノードが複数存在する場合、ノード間のコストを用いて枝切りを行う。SVM では scikit-learn の decision_function を利用してエッジ間のコストを決定する。BERT では Softmax 関数の正解と予測された確率値と以下の計算式を利用して 2 種類のエッジのコストを決定する。

(注 2) : <http://code.google.com/archive/p/word2vec/>

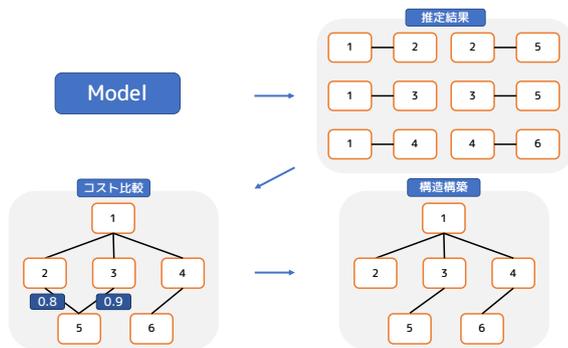


図4 Bottom-up 手法の概要図

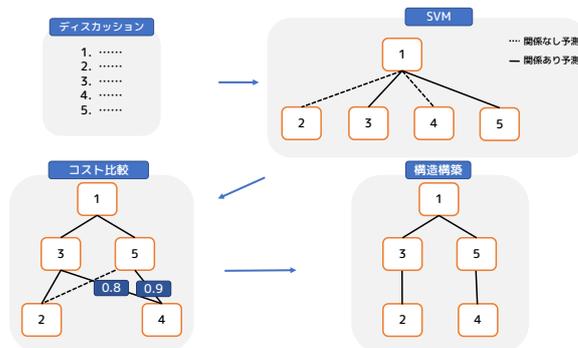


図5 Top-down 手法の概要図

$$Weight = softmax \cdot time \cdot distance$$

ここでの *softmax* は Softmax 関数で正解と予測された確率値, *time* は親ノードの発話終了時刻から子ノードの発話開始時刻までの時間を表し, *distance* は親ノードから子ノードまでに発話されたノード数を表す. その後, それぞれエッジのコストの値で比較を行い, 小さい方のエッジのコストを枝切りする. 具体的には図4のコスト比較のようにエッジ間にコストが設定され小さい方を枝切りする (0.8 vs. 0.9). 最後に, 子ノードに対して親ノードが一つに定まるまで繰り返し, 図4のような議論構造を構築する.

4.2.2 Top-down 手法

4.1 節で説明した SVM が発話間の関係を推定する際に本手法を用いる. 図5に本手法の概要図を示す. まず, 図5のようにディスカッション中で最も発話開始時刻の早いノードを議論構造の親ノードと仮定し, その親ノードと他のノードとの関係の有無を推定する. 次に, 仮定した親ノードと関係があると推定されたノードを次の親ノードと仮定し, 推定されていない残りのノードとの関係の有無を推定する. 一つの子ノードに対して複数の親ノードが関係があると推定された場合, 図5のコスト比較のようにノード間のコストを用いて枝切りを行う (0.8 vs. 0.9). エッジのコストは scikit-learn の `decision_function` を利用して決定し, エッジの比較を行いコストの小さい方のエッジを枝切りする. 全てのノードが関係を持つまで SVM の推定を繰り返し, 図5のような議論構造を構築する. Top-down 手法は全てのノードの組み合わせを考慮する必要があるため Bottom-up 手法と比べて計算コストが大きい. 同様に, SVM と比較して BERT では推定に時間がかかるため本研究において Top-down 手法は SVM にだけ適用する.

5. 実験と分析

本節では, 4.1 節で述べたモデルのパラメータについて説明する. 次に, 3. 節で述べた複数人議論コーパスを用いた実験データについて説明する. 最後に, 2つのモデルで発話間の関係を推定した結果と 4.2 節で説明したエッジの選択手法を利用した結果を比較する. また, 二つのモデルの実験結果に対して考察を行う.

5.1 モデルのパラメータ

2つのモデルで設定したパラメータについて説明する.

表1 各データのノードペアの分布

	ダイアログ	ディスカッション	関係あり	関係なし
訓練データ	84	201	4002	41127
開発データ	4	13	235	1822
評価データ	4	12	242	2009

SVM のパラメータではカーネルに `rbf` を利用し, コストパラメータを 100 と設定した.

BERT の事前学習モデルには BERT-Base を利用し, 前処理として英単語を全て小文字化したものを利用した. また訓練データのバッチサイズは 16, 開発データと評価データのバッチサイズを 8 とし, エポック数を 3, ロス関数には `cross-entropy` を利用した. さらに最適関数には Adam [19] を利用し, 学習率を $2e-5$ とした.

5.2 実験データ

実験データは, AMI corpus のダイアログ 92 対話から作成された 219 個のディスカッションを用いて実験を行った. 本実験では, 各ディスカッションでノードは全て抽出されており, ノードに付与されている Unit Label は分類されていると仮定した. 訓練データ, 開発データ, 評価データに AMI corpus をダイアログ単位で 84 対話, 4 対話, 4 対話に分割した. また, それぞれのディスカッションの数は 201 個, 13 個, 12 個となっている. ディスカッション内の 2 つノードを総当たりでノードペアを作成した. 訓練データ, 開発データ, 評価データのダイアログの数, ディスカッションの数, ノードペアの数の分布を表 1 に示す. 表 1 より, 訓練データのノードペアの総数 45129 に対し関係ありが 4002 ペア, 関係なしが 41127 ペアとデータに偏りがある. そこで, SVM と BERT に訓練する際はデータを関係ありのノードペアの数に合わせて, 訓練データの関係なしのノードペアから無作為に 4002 ペア抽出し, 関係ありの 4002 ペアと均衡なデータに成形して学習させた.

5.3 実験結果

エッジの選択手法を用いなかったモデルと用いたモデルで, 発話間の関係の有無の推定結果を表 2, 表 3 に示す.

表 2, 表 3 よりどちらのモデルにおいても選択手法を用いなかった結果 (SVM のみ, BERT のみ) より用いた結果の方が関係ありの精度が高くなっている. また, SVM の選択手法の中では Bottom-up 手法が有効であることが F 値からわかる. BERT ではノード間の時間情報とノード間の距離情報を用いたエッジ

表2 SVMによる推定結果

Model	関係あり			関係なし		
	P	R	F1	P	R	F1
SVMのみ	0.38	0.84	0.53	0.98	0.84	0.90
+ Bottom-up	0.68	0.63	0.66	0.96	0.97	0.96
+ Top-down	0.58	0.47	0.48	0.94	0.94	0.94

表3 BERTによる推定結果

Model	関係あり			関係なし		
	P	R	F1	P	R	F1
BERTのみ	0.15	0.65	0.24	0.93	0.54	0.68
+ Bottom-upsoftmax	0.40	0.31	0.35	0.92	0.95	0.93
+ Bottom-upweight	0.56	0.32	0.41	0.92	0.97	0.95

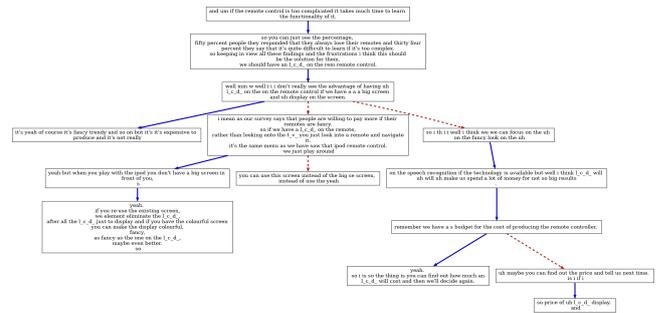


図8 SVMに Bottom-up 手法を用いた推定結果による議論構造

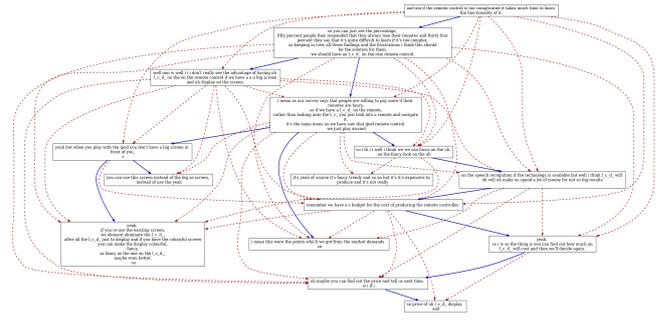


図9 BERTの推定結果による議論構造

図6 正しい議論構造

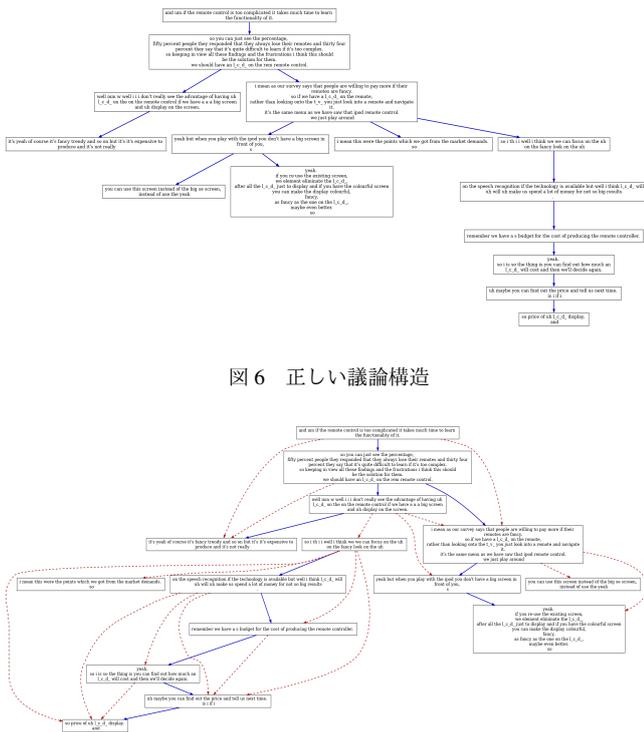


図7 SVMの推定結果による議論構造

のコストを利用した選択手法 (Bottom-upweight) が有効であることを確認できた。

さらに、選択手法で有効であった SVM の Bottom-up 手法と BERT の Bottom-upweight 手法を利用して議論構造を構築する。1つのディスカッションに着目して正しい議論構造と2つモデルに対して選択手法を利用した場合と利用しなかった場合の議論構造を構築した結果を図6~図10に示す。図6~図10中の実線は正解エッジを示し、点線は不正解エッジを示す。図7と図9より選択手法を用いていない議論構造では不正解エッジが目立つ結果になっているが、図8と図10より選択手法を用いている議論構造では不正解エッジが少なく、図6の正しい議論構造に近い形になっていることが確認できる。

5.4 考察

表2の結果から考察する。表2では Bottom-up 手法が最も関係ありの F 値が高い。1.節で述べたように SVM のみのモデル

図10 BERTに Bottom-up 手法のコストに weight

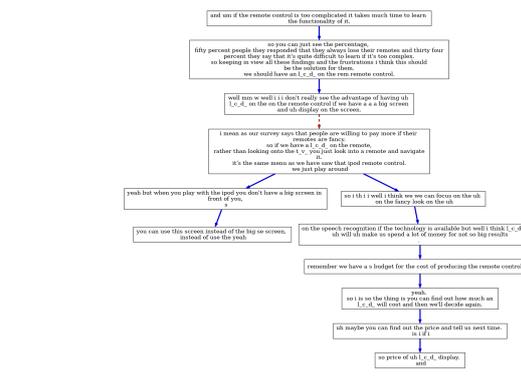


図10 BERTに Bottom-up 手法のコストに weight を用いた推定結果による議論構造

では、精度が低く再現率が高いという特徴がありノード間の関係がないペアを多く推定されている。そのため、子ノードに対して親ノードが一つのみ関係があるという議論構造の特徴とは異なる結果になっていることが多い。Bottom-up 手法を用いることで議論構造の特徴に基づく構造に近くなることから精度が向上したと考えられる。また、Top-down 手法でも選択手法を用いなかった結果と比較して精度が向上している。しかし、関係ありの F 値では選択手法を用いなかった結果より低くなった。Top-down 手法は Bottom-up 手法と同様に議論構造の特徴に基づいてエッジの選択を行っている。しかし、エッジの選択の過程において Bottom-up 手法とは異なり、親ノードから他のノードに対して逐次的に関係の有無の推定を行うため発話関係を保持することはできても議論構造全体の関係は保持できていない。そのため、Bottom-up 手法と比較して精度は高いが再現率は低くなっている。

次に、表3の結果から考察する。表3より、BERTのみのモデルでは、関係ありの再現率と関係なしの精度が高く、関係ありの精度と関係なしの再現率が低い。言い換えると、図9でわかるようにBERTの出力結果から議論構造を構築した際にエッジの数が多くなり、その中でも関係のない発話ペアが多くなっている。一方で、表3より二種類の選択手法を用いることで、関係ありの精度と関係なしの再現率を向上できている。特に、Bottom-up_{weight}手法の精度が向上している。以前の研究[2]で発話間の関係の有無の推定においてノード間の時間情報とノード間の距離情報が有効な素性であることが確認されている。本研究においても、ノード間の時間情報と距離情報を利用したコスト設定で有効であることがわかった。

6. おわりに

本研究では、複数人議論コーパスを対象に議論構造の構築を行った。複数人議論コーパスにはAMI corpusを利用した。次に、以前の研究での問題点である精度が低いことを解決するためにエッジの選択手法の提案を行い、発話間の関係の有無の推定を行うモデルに対して、エッジの選択手法の利用したモデルと利用しないモデルで比較実験を行った。その結果、提案手法の精度が高くなることを確認できた。また、議論構造を構築した際にも選択手法を利用したモデルのエッジの総数が少なくなり、不正解エッジが少なくなっていることで正しい議論構造に近い形になっていることが確認できた。

今後の課題としては、先行研究で提案したモデルの改良が挙げられる。Bottom-up手法を用いるとしてもエッジの候補がどちらも正しくなければ、選択手法で改善することはできない。そのため、正しい議論構造を構築することも不可能である。一方で正しいエッジを多く推定できれば選択手法によって正しい議論構造に近づけることができる。これら理由から発話間の関係の有無の推定において再現率がより高くなるようなモデルの構築が今後の課題として挙げられる。また、構築した議論構造から多くの情報を取得できるようにするため、ノードのUnit Labelを推定する要素抽出や、エッジのラベルを推定する関係分類に取り組むことが挙げられる。要素抽出や関係分類に取り組むことによって、議論構造を確認すれば発話がどのような意図で話されているかだったり、ある主張に対してどのような発言意図があるかの理解支援に役立つと考えられる。

謝辞

本研究は科研費20K12110の助成を受けたものです。

文献

[1] C. Stab and I. Gurevych, "Parsing argumentation structures in persuasive essays," *Computational Linguistics*, vol.43, no.3, pp.619–659, 2017.

[2] T. Himeno and K. Shimada, "Relation identification using dialogical features in multi-party conversation," In *Proceedings of the 8th International Symposium on Applied Engineering and Sciences*, pp.1–11, 2020.

[3] E. Barker and R. Gaizauskas, "Summarizing multi-party argumentative conversations in reader comment on news," *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pp.12–20, 2016.

[4] A. Peldszus, "Towards segment-based recognition of argumentation

structure in short texts," *Proceedings of the First Workshop on Argumentation Mining*, pp.88–97, 2014.

[5] D. Ghosh, A. Khanam, Y. Han, and S. Muresan, "Coarse-grained argumentation features for scoring persuasive essays," *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp.549–554, 2016.

[6] C. Stab and I. Gurevych, "Training argumentation skills with argumentative writing support," *Proc. SEMDIAL 2017 (SaarDial) Workshop on the Semantics and Pragmatics of Dialogue*, pp.166–167, 2017.

[7] H. Nguyen and D. Litman, "Context-aware argumentative relation mining," *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp.1127–1137, 2016.

[8] C. Stab, T. Miller, B. Schiller, P. Rai, and I. Gurevych, "Cross-topic argument mining from heterogeneous sources," *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp.3664–3674, Association for Computational Linguistics, Brussels, Belgium, Oct. Nov. 2018.

[9] C. Stab and I. Gurevych, "Identifying argumentative discourse structures in persuasive essays," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.46–56, 2014.

[10] E.W. Dijkstra, et al., "A note on two problems in connexion with graphs," *Numerische mathematik*, vol.1, no.1, pp.269–271, 1959.

[11] H.N. Gabow and R.E. Tarjan, "A linear-time algorithm for a special case of disjoint set union," *Journal of computer and system sciences*, vol.30, no.2, pp.209–221, 1985.

[12] J. Chamberlain, U. Kruschwitz, and O. Hoebler, "Scalable visualisation of sentiment and stance," *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pp.4181–4185, 2018.

[13] L. Lugini, C. Olshefski, R. Singh, D. Litman, and A. Godley, "Discussion tracker: Supporting teacher learning about students' collaborative argumentation in high school classrooms," *Proceedings of the 28th International Conference on Computational Linguistics: System Demonstrations*, pp.53–58, International Committee on Computational Linguistics (ICCL), Barcelona, Spain (Online), Dec. 2020. <https://www.aclweb.org/anthology/2020.coling-demos.10>

[14] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, et al., "The ami meeting corpus: A pre-announcement," *International workshop on machine learning for multimodal interaction* Springer, pp.28–39 2005.

[15] R. Rienks, D. Heylen, and E. van der Weijden, "Argument diagramming of meeting conversations," *Multimodal Multiparty Meeting Processing, Workshop at the 7th International Conference on Multimodal Interfaces*, pp.85–92, 2005.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, pp.1–16, 2018.

[17] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, "SentiLARE: Sentiment-aware language representation learning with linguistic knowledge," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.6975–6988, Association for Computational Linguistics, Online, Nov. 2020.

[18] D. Liu, Y. Gong, J. Fu, Y. Yan, J. Chen, D. Jiang, J. Lv, and N. Duan, "RikiNet: Reading Wikipedia pages for natural question answering," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp.6762–6771, Online, July 2020.

[19] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *The International Conference on Learning Representations (ICLR)*, pp.1–15, 2015.