

# 文脈を考慮した半教師あり学習による対話行為推定

石島 孝俊 白井 清昭

北陸先端科学技術大学院大学 先端科学技術研究科  
{takatoshi.ishijima,kshirai}@jaist.ac.jp

## 概要

本研究では発話の先行文脈を考慮して対話行為を推定する半教師あり学習の手法を提案する。一般に発話の対話行為は文脈に依存するが、教師あり学習によって文脈を考慮したモデルを構築する手法は提案されていたものの、教師なし・半教師あり学習において文脈を考慮することの有効性は十分に検証されていなかった。提案手法では、先行する発話の文脈情報を反映した発話の埋め込みを作成し、クラスターリングやk-NN法に基づいて少数のラベル付き発話からラベルなし発話に対話行為を伝播させる。評価実験の結果、文脈を考慮することで対話行為推定の正解率が向上することを確認した。

## 1 はじめに

対話行為推定とは、対話における発話の対話行為を分類するタスクであり、その実現は対話システムにおける重要な要素技術の一つである。対話行為を推定する際には、発話そのものではなく、その文脈、すなわちその発話に先行する発話の情報が重要な手がかりとなる。多くの対話行為推定の先行研究では、このような文脈情報を考慮したモデルが提案されている。しかし、対話行為のラベルが付与されていないコーパスを用いる教師なし学習や、少量のラベル付きデータと大量のラベルなしデータの両方を用いる半教師あり学習では、文脈の情報はそれほど考慮されていなかった。

本研究では、少量の対話行為ラベル付きデータしか存在しないことを仮定し、発話の文脈を考慮して対話行為を推定する手法を提案する。ラベル付き発話とラベルなし発話が与えられたとき、ラベルなし発話の対話行為をそれと類似したラベル付き発話の対話行為から推定し、この処理を漸進的に繰り返す。また、発話間の類似度を計算する際、先行文脈を考慮するいくつかの方法を探究する。JAIST タグ

付き自由対話コーパス [1] を用いた評価実験を行い、提案手法の有効性を検証する。

## 2 関連研究

対話行為推定の研究は教師あり機械学習に基づく手法が主流である。福岡と白井は、対話行為のクラス毎に最適な特徴のセットを自動的に獲得し、各対話行為の二値分類器を組み合わせて発話の対話行為を推定した [2]。Liu らは、CNN(Convolutional Neural Network) を用いて発話をベクトル表現に変換し、その上に別の CNN 層や RNN(Recurrent Neural Network) 層を重ねることで発話の文脈情報を組み込むモデルを提案した [3]。実験では文脈を考慮することによって対話行為推定の正解率が有意に改善することを示した。Bothe らも文脈を取り入れた対話行為推定の手法を提案している [4]。LSTM(Long Short-Term Memory) を用いて発話の埋め込み表現を獲得し、先行する発話の埋め込み表現と分類対象の発話の埋め込み表現を RNN に入力することで、文脈を考慮した対話行為推定モデルを実現した。

発話以外の情報を手がかりに対話行為を推定する研究も行われている。Maraev らは対話行為を推定するための手がかりとして「笑い」に注目し、「笑い」を対話行為推定モデルに組み込むことの効果を実験的に検証した [5]。He らは、話者交替の情報を「話者ターン埋め込み」として表現し、発話の埋め込みとマージすることで、対話行為推定の性能が向上することを示した [6]。

大規模なラベル付きコーパスを必要としない対話行為推定の研究も行われている。Yang らは、半教師あり学習による対話行為推定手法として、少量の対話行為タグが付与された発話と付与されていない発話を同時にクラスターリングすることで未知の発話の対話行為を推定する手法を提案した [7]。ただし、彼らの手法では発話の先行文脈は考慮されていない。

### 3 提案手法

#### 3.1 問題設定

対話行為のラベルが付与された発話の集合を  $L$ , 付与されていない発話の集合を  $U$  とする. 最初は  $L$  はごく少量の発話から構成されるとし, これをシード発話と呼ぶ.  $L$  と  $U$  を入力とし,  $U$  における全ての発話の対話行為を推定する.

#### 3.2 発話埋め込み

提案手法では発話をベクトルで表現する. 以下, これを発話埋め込みと呼ぶ. 本研究では, 事前学習した Bidirectional Encoder Representations from Transformers(BERT)[8, 9] を用いる. 発話を BERT に入力し, [CLS] トークンに対する埋め込みを発話埋め込みとする.

事前学習済み BERT を対話ドメインに適応させるため, 対話コーパスを用いた再事前学習を行う. 具体的には, 表 1 に示す 3 つの対話コーパスを使用し, Masked Language Modeling タスクによって BERT を再事前学習する.

表 1 BERT の再事前学習に用いた対話コーパス

コーパス	発話数
JAIST タグ付き自由対話コーパス [1]	96,962 発話
BTSJ 日本語自然会話コーパス [10]	178,619 発話
日本語話題別会話コーパス:J-TOCC[11]	158,139 発話

発話埋め込みを作成する際, その発話の先行発話の情報を文脈として組み込む. 分類対象の発話を  $u_0$ , それより前の  $C$  件の発話を  $u_{-1}, \dots, u_{-C}$  とおく. また, BERT によって得られる発話  $u_i$  の埋め込みを  $\mathbf{e}_i$  と記す. 以下の 5 通りの方法で発話埋め込みを作成する.

**ave** 対象発話と先行文脈の発話の埋め込みの平均ベクトル.

$$\mathbf{e}_{ave} = \frac{1}{C+1} \sum_{i=0}^C \mathbf{e}_{-i} \quad (1)$$

**mp** 対象発話と先行文脈の発話の埋め込みの Max Pooling によって得られたベクトル.

$$\mathbf{e}_{mp} = \max\_pooling(\mathbf{e}_0, \dots, \mathbf{e}_{-C}) \quad (2)$$

**mp-sep** 対象発話のベクトルと, 先行文脈の発話埋め込みの Max Pooling によって得られたベクトルを連結したベクトル.

$$\mathbf{e}_{mp-sep} = \mathbf{e}_0 \oplus \max\_pooling(\mathbf{e}_{-1}, \dots, \mathbf{e}_{-C}) \quad (3)$$

**w-ave** 対象発話と先行文脈の発話の埋め込みの重み付き平均ベクトル. 重みは対象発話との距離によって決める. 1 つ前の発話に対する重みを 0.9 とし, それより前の発話については 0.1 ずつ重みを小さくしていく.

$$\mathbf{e}_{w-ave} = \frac{1}{C+1} \left( \sum_{i=0}^C (1 - 0.1 \times i) \mathbf{e}_{-i} \right) \quad (4)$$

**w-ave-sep** 対象発話のベクトルと, 先行文脈の発話埋め込みの重み付き平均ベクトルを連結したベクトル. 1 つ前の発話に対する重みは 1 とし, それより前の発話は 0.1 ずつ重みを小さくする.

$$\mathbf{e}_{w-ave-sep} = \mathbf{e}_0 \oplus \left( \frac{1}{C} \sum_{i=1}^C (1 - 0.1 \times (i - 1)) \mathbf{e}_{-i} \right) \quad (5)$$

#### 3.3 対話行為の推定

本項では発話の対話行為を推定する 4 つの手法について述べる.

##### 3.3.1 クラスタリングに基づく手法

まず, 発話集合  $L \cup U$  に対してクラスタリングを行う. 次に, ラベルなし発話  $u \in U$  に対し, その対話行為ラベルを以下の手続きで付与する.

1.  $u$  が属するクラスタの中に  $L$  に属する発話が存在する場合, それらの発話の対話行為クラスの多数決によって付与する対話行為を決める. 最多の対話行為クラスが複数あるときは, その中からランダムに選択する.
2. それ以外の場合,  $u$  が属するクラスタの重心と最も近いラベル付き発話を探し, その対話行為を  $u$  に付与する.

クラスタリングアルゴリズムとして  $k$ -means を使用する.  $k$ -means ではクラスタ数をあらかじめ設定する必要がある. 本手法では同じクラスタに属するラベル付き発話から対話行為を伝播するため, クラスタは同じ対話行為の発話で構成されることが要求される. すなわち, ここでのクラスタリングでは Purity(クラスタ内で最多の対話行為クラスが占める割合)が高いことが求められる. 一方, ある対話行為の発話は必ずしも 1 つのクラスタとしてまとまる必要はない. そのため, サイズは小さいが Purity の高いクラスタを構築するために, クラスタ数は比較的大きい値に設定する. 具体的には, クラスタ数  $k$  を平均クラスタサイズが 5 になるように設定する ( $k \approx \frac{|L \cup U|}{5}$ ). 以下, この手法を DAC-CL と記す.

### 3.3.2 クラスタリングとブートストラップに基づく手法

前述の手法 DAC-CL にブートストラップの手法を適用し，ラベル付き発話を漸進的に増大させる．3.3.1 で述べた手続きにおいて，対話行為の決定に用いた最も近いラベル付き発話との距離を求める．その距離が閾値  $T_c$  以下のとき， $u$  の対話行為を決定し，それを  $U$  から  $R$  に移す．以上の操作を何回か繰り返す．最初のサイクルでは  $U$  のうち 2% の発話を  $R$  に移すように  $T_c$  を決定し，以降のサイクルではその  $T_c$  をそのまま用いる．

距離が近いラベル付き発話によって推定した対話行為は，その信頼度が高いと考えられる．図 1 に示すように，信頼度が高いものから順に対話行為を決定することで対話行為推定の性能を向上させることを狙う．以下，この手法を DAC-CL-B と記す．

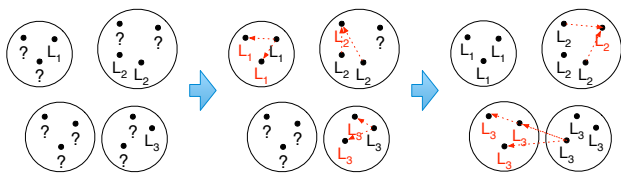


図 1 クラスタリングとブートストラップに基づく半教師あり対話行為推定

### 3.3.3 k-NN に基づく手法

ラベルなし発話  $u \in U$  に対し，ラベル付き発話集合  $L$  の中から最近傍の  $k$  個の発話を選択し，それに付与された対話行為ラベルの多数決によって対話行為を付与する．最多の対話行為ラベルが複数あるときはその中からランダムに選択する．以下，この手法を DAC-kNN と記す．

### 3.3.4 k-NN とブートストラップに基づく手法

k-NN に基づく手法でもブートストラップ的にラベル付き発話を漸進的に増大させることを試みる．最近傍の  $k$  個のラベル付き発話のうち最多の対話行為を持つ発話との距離の平均値が閾値  $T_k$  以下のとき， $u$  の対話行為を決定し，それを  $U$  から  $R$  に移す．この操作を繰り返す．閾値  $T_k$  は，手法 DAC-CL-B と同様に，最初のサイクルで  $U$  のうち 2% の発話を  $R$  に移すように決定する．以下，この手法を DAC-kNN-B と記す．

## 4 評価実験

### 4.1 実験設定

JAIST タグ付き自由対話コーパス [1, 12] を用いて提案手法の評価実験を行う．同コーパスは人間同士の雑談を書き起した発話に対して対話行為などのタグが付与されている．対話行為タグのうち「自己開示」がおおよそ 8 割を占めるが，その定義は広く，『自分の好みや弱点をさらけ出す発話』といった狭義の自己開示ではない発話に対しても付与されている．本研究ではその定義を見直し，「自己開示」の対話行為タグが付与された発話を再分類した．結果として表 2 に示す 21 個の対話行為クラスを設定した．それぞれの定義と発話例を付録 A に示す．

表 2 対話行為クラスの定義と発話数

伝達-意見	152	あいさつ	3	確認	9
伝達-体験	88	感謝	2	要求	2
伝達-他者体験	37	同意	45	フィラー	2
伝達-伝聞	33	非同意	11	質問 (YesNo)	6
伝達-情報	57	独り言	9	質問 (What)	7
感想	53	自己開示	2	応答 (YesNo)	2
感情表出	30	あいづち	4	応答 (平叙)	23

JAIST タグ付き自由対話コーパスにおける 1 つの対話を選び，その発話で「自己開示」のタグが付与されていた 884 個の発話に対して，著者 2 名が表 2 のいずれかの対話行為タグを割り当てた．両者のタグ付けの一致率は 0.624， $\kappa$  係数は 0.560 であった．一致しない発話の対話行為タグは著者 2 名の合議により決定した．表 2 中の数値はタグ付けされた発話数である．

シード発話  $L$  として，それぞれの対話行為クラス毎に 3 つの発話をランダムに選択した．3 つ以下の発話しかない対話行為クラスについてはその全てをシード発話とした．それ以外の発話をラベルなし発話集合  $U$  とし，提案手法によってその対話行為を推定した．評価指標は  $U$  に対する対話行為推定の正解率とした．また，シード発話の選択を 4 回繰り返し，その正解率の平均を求めた．文脈を考慮しない手法をベースラインとし，提案手法と比較した．

### 4.2 結果と考察

クラスタリングに基づく手法 DAC-CL の評価結果を表 3 に示す．行は文脈を考慮した発話埋め込みの作成方法 (3.2 項)，列は先行文脈の長さを表す．提案手法はほとんどのケースでベースライン (文脈を考慮しない発話埋め込みを用いる手法) の正解率 (0.093) を上回った．一番良い組み合わせは，重



み付き平均で発話埋め込みを作成し (**w-ave**), 文脈長が  $C = 1$  のときで, そのときの正解率は 0.133 であった. ただし, 先行文脈の長さの優劣については, はっきりした傾向は読み取れなかった.

表 3 手法 DAC-CL の正解率

	$C = 1$	$C = 2$	$C = 3$	$C = 4$	$C = 5$
<b>ave</b>	0.125	0.114	0.101	0.111	0.129
<b>mp</b>	0.125	0.099	0.102	0.122	0.120
<b>mp-sep</b>	0.115	0.096	0.090	0.108	0.104
<b>w-ave</b>	<b>0.133</b>	0.120	0.101	0.131	0.124
<b>w-ave-sep</b>	0.115	0.103	0.105	0.094	0.118
文脈なし	0.093				

クラスタリングとブートストラップに基づく手法 DAC-CL-B の評価結果を表 4 に示す. 発話埋め込みは表 3 で一番良い結果が得られた **w-ave** を用いている. 反復学習の途中では全ての発話について対話行為が決定されていないが, その時点で得られたラベル付き発話  $L$  を基に手法 DAC-CL で全ての発話の対話行為を推定し, その正解率を示している. 「0 回目」は最初のシード発話のみを用いたときの結果である.  $C = 2$  と  $C = 3$  のときでは反復操作によって正解率が向上した. しかし, その効果はそれほど大きくなく, また他の文脈長では正解率は低下した.

表 4 手法 DAC-CL-B(**w-ave**) の正解率

反復回数	0 回目	1 回目	2 回目
$C = 1$	<b>0.133</b>	0.107	0.104
$C = 2$	0.120	<b>0.122</b>	<b>0.122</b>
$C = 3$	0.101	<b>0.111</b>	0.106
$C = 4$	<b>0.131</b>	0.116	0.117
$C = 5$	<b>0.124</b>	0.114	0.115

$k$ -NN に基づく手法 DAC-kNN の評価結果を表 3 に示す.  $k$  は  $k$ -NN 法におけるパラメタ,  $C$  は先行文脈の長さを表す. 5 種類の発話埋め込みを用いて実験を行ったが, 紙面の都合上, 一番結果の良かった **w-ave-sep** の結果のみを示す. 太字は各行において最も正解率が高いことを示す. 文脈を考慮することにより, おおむね対話行為推定の正解率が改善されている.  $C = 3$  のときに一番正解率が高くなることが多く, またこのときは全ての  $k$  について文脈なし ( $C = 0$ ) の正解率を上回る. 最高の正解率は 0.233 であり, 表 3 に示したクラスタリングに基づく手法の最高の正解率よりも高かった.

$k$ -NN とブートストラップに基づく手法 DAC-kNN-B の評価結果を図 2 に示す. 横軸は反復回数, 縦軸は正解率である. 反復学習の各ステップで, その時点でのラベル付き発話  $L$  を基に手法 DAC-kNN で全ての発話の対話行為を推定したときの正解率を示し

表 5 手法 DAC-kNN(**w-ave-sep**) の正解率

	$C = 0$	$C = 1$	$C = 2$	$C = 3$	$C = 4$	$C = 5$
$k = 1$	0.123	0.129	<b>0.140</b>	0.128	0.117	0.127
$k = 2$	0.178	<b>0.194</b>	0.182	<b>0.194</b>	0.181	<b>0.194</b>
$k = 3$	0.205	0.205	0.211	<b>0.221</b>	0.214	0.220
$k = 4$	0.214	0.210	<b>0.228</b>	0.221	0.224	0.225
$k = 5$	0.211	0.209	0.217	<b>0.233</b>	0.225	0.218
$k = 6$	0.209	0.206	0.210	<b>0.233</b>	0.220	0.207
$k = 7$	0.205	0.196	0.211	<b>0.229</b>	0.219	0.212

ている. 先行文脈の長さは, 表 5 で一番高い正解率が得られた  $C = 3$  と設定している.  $k = 2$  の場合を除き, 反復学習を進めるにつれて正解率が向上する傾向は見られなかった. また, シード発話によっては正解率が 0.30 程度になることもあるが, 4 回の試行の平均では  $k = 2$  の場合を除いて手法 DAC-kNN(「0 回目」の結果)を上回ることはなかった.

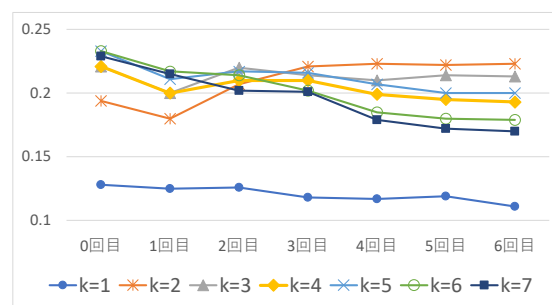


図 2 手法 DAC-kNN-B(**w-ave-sep**,  $C = 3$ ) の評価

考察をまとめると, 発話埋め込みに文脈の情報を組み込むことで正解率が向上することが確認された.  $k$ -NN 法に基づく手法の方がクラスタリングに基づく手法よりも正解率が高かった. どちらの手法でもブートストラップ法によってラベル付き発話の量を漸進的に増加させることの効果は見られなかった. また, 正解率自体も最高で 0.233 と低かった. 分類対象とする対話行為のクラスが 21 と多いこと, ラベル付き発話の量が少ない半教師あり学習の設定での実験であることなどが要因として考えられる.

## 5 おわりに

本論文では, 文脈を考慮した対話行為推定の半教師あり学習手法を提案した. 評価実験により発話埋め込みに文脈を反映させることの有効性を確認した. しかしながら, 対話行為の正解率は最高でも 0.233 であり, 改善の余地が大きい. 主な今後の課題として, 大規模言語モデルなどの外部知識を利用して, 対話行為推定の正解率を改善することが挙げられる.

## 参考文献

- [1] Kiyooki Shirai and Tomotaka Fukuoka. JAIST annotated corpus of free conversation. In **Proceedings of the Eleventh Edition of the Language Resources and Evaluation Conference**, pp. 741–748, 2018.
- [2] 福岡知隆, 白井清昭. 対話行為に固有の特徴を考慮した自由対話システムにおける対話行為推定. 自然言語処理, Vol. 24, No. 4, pp. 523–546, 2017.
- [3] Yang Liu, Kun Han, Zhao Tan, and Yun Lei. Using context information for dialog act classification in DNN framework. In **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 2170–2178, 2017.
- [4] Chandrakant Bothe, Cornelius Weber, Sven Magg, and Stefan Wermter. A context-based approach for dialogue act recognition using simple recurrent neural networks. In **Proceedings of the Eleventh International Conference on Language Resources and Evaluation**, pp. 1952–1957, 2018.
- [5] Vladislav Maraev, Bill Noble, Chiara Mazzocconi, and Christine Howes. Dialogue act classification is a laughing matter. In **Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue**, 2021.
- [6] Zihao He, Leili Tavabi, Kristina Lerman, and Mohammad Soleymani. Speaker turn modeling for dialogue act classification. In **Findings of the Association for Computational Linguistics: EMNLP 2021**, pp. 2150–2157, 2021.
- [7] Xiaohao Yang, Jia Liu, Zhenfeng Chen, and Weilan Wu. Semisupervised learning of dialogue acts using sentence similarity based on word embeddings. In **Proceedings of International Conference on Audio, Language and Image Processing**, pp. 882–886, 2014.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 4171–4186, 2019.
- [9] BERT base Japanese (IPA dictionary, whole word masking enabled) – Hugging Face, (2024-1 閲覧). <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking/>.
- [10] 宇佐美 まゆみ (監修). BTSJ 1000 人日本語自然会話コーパス. 科研基盤研究 (A) 「語用論的分析のための日本語 1000 人自然会話コーパスの構築とその多角的な研究」(研究代表者: 宇佐美 まゆみ), 2023.
- [11] 中俣尚己, 太田陽子, 加藤恵梨, 澤田浩子, 清水由貴子, 森篤嗣. 日本語話題別会話コーパス: J-TOCC. 計量国語学, Vol. 33, No. 1, pp. 11–21, 2021.
- [12] 言語資源協会. JAIST タグ付き自由対話コーパス, (2024 年 1 月閲覧). <https://www.gsk.or.jp/catalog/gsk2017-b>.

## A 対話行為クラスの定義

実験に用いた対話行為クラスの定義と発話例を表6に示す。発話例における( )内は前の発話である。

表6 対話行為クラスの定義と発話例

対話行為	定義	発話例
伝達-意見	自分の意見、見解、意思、意向を伝える発話	(お鍋セット、いいねー。)うーん、家族がたくさんいる人は良いかもしれないけどさー。
伝達-体験	自分が体験したこと、体験する予定のこと、今の心理状態を伝える発話	あの一、あそこに行ったよ、この間。
伝達-他者体験	自分が直接見たり聞いたりした他者のことを伝える発話	あいた、あいたって、しょっちゅうやってるよ。
伝達-伝聞	人から聞いた状況(自分はその場にはいない)を伝える発話	いつも、なんか、こんなふうにして寝ますって言ってたよ。
伝達-情報	客観的な事実(知識、常識なども含む)を伝える発話	受験生も一番多かったよ、Bが。
感想	(原則として相手の発話や周りの状況を受けて)自分の感想を表す発話	(チョコレートとかねー、ちょっと食べたかったね。)チョコレートかー、ええなー。
感情表出	自身の感情を表現するような発話	かわいそうだねー。
自己開示	自分の好みや弱点をさらけ出す発話	こっちはこっちでねー、小さい人にあこがれてんだから。
あいさつ	一般的なあいさつを行っている発話	先生、こんにちは。
感謝	相手にお礼の気持ちを伝えている発話	ありがとう。
同意	相手の発言に納得や理解を示している発話	うん、ほんとに。
非同意	相手の発言に対して同意できないことを示している発話	(じゃ、友達呼んで、お鍋パーティすればいい。)うーん、そうだけでも。
独り言	相手に話しかけているわけではないつぶやきのような発話	よししょ。
あいづち	相手の発話に対して、話の続きを促し、また、発話者が相手の話を聞いていることを示す短い発話	そうだよね。
確認	発話者がすでに知っている、またはそうだと思っていることに対して、相手にそれが正しいかを確認する発話	(すごい混んでたよね。)まじですか。
要求	相手に対して、具体的な行動を支持もしくは依頼する発話	ちゃんとナビしてな。
質問(YesNo)	「はい」「いいえ」などで答えられる質問をしている発話	知ってる?
質問(What)	平叙文で答える必要がある質問をしている発話	何で行くの?
応答(YesNo)	質問に対する「はい」「いいえ」などの短い肯定、否定の発話	「うん。」 「そう。」
応答(平叙)	質問に対して平叙文で答えている応答文に相当する発話	うーん、覚えてない。
フィラー	言い淀みなど発話の合間に挟み込む短い発話	「えーと。」 「あの一。」