

## 映画推薦対話を具体例とした話者内部状態の 推定による対話管理

児玉 貴志<sup>†</sup>・田中リベカ<sup>†</sup>・黒橋 禎夫<sup>†</sup>

人間と機械の間の新たなインターフェースとして知的な対話システムの実現が期待されている。知的な対話システムは対話中の話者の内部状態を推測し、その結果に応じて適切に応答を変更する必要がある。本研究では映画推薦対話を具体例として、対話中の話者内部状態のモデル化とその話者内部状態を踏まえて応答を変更する対話システムの構築に取り組む。映画推薦をドメインとした対話システムを構築し、対話収集を行う。収集した対話データの分析に基づき、話者内部状態を話題に関する知識の有無、話題への興味の有無、対話意欲の有無の3つの軸でモデル化する。モデル化した話者内部状態を収集した対話データにアノテートし、これを学習データとして話者内部状態の自動推定を行った結果、高い推定精度を達成した。また、各話者内部状態に応じてシステムの応答を変更するルールを設計する。学習した話者内部状態推定器を用いて各話者内部状態を判定し、対話システムの応答を変更することでシステム発話の自然さが向上することを対話単位での評価と発話単位での評価の両方で確認した。

キーワード：対話システム、話者内部状態、知識、興味、対話意欲

## Dialogue Management by Estimating User's Internal State Using the Movie Recommendation Dialogue

TAKASHI KODAMA<sup>†</sup>, RIBEKA TANAKA<sup>†</sup> and SADAO KUROHASHI<sup>†</sup>

Intelligent dialogue systems are expected to be a new interface between humans and machines. An ideal intelligent dialogue system should estimate the user's internal states and incorporate the estimation results into its response appropriately. In this paper, we focus on the movie recommendation dialogues and propose a dialogue system that considers the user's internal state. First, we build a movie recommendation dialogue system and collect dialogue data. Based on the analysis of the collected dialogue data, we model and annotate the user's internal states in three aspects: knowledge, interest, and engagement. Second, we train the user's internal state estimators on the dialogue corpus with the annotations of the user's internal states. The trained estimator achieved high accuracy on the annotated corpus. Further, we design a set of rules that modify the system's responses according to each user's internal state. We confirmed that the response modifications based on the results of

---

<sup>†</sup> 京都大学 大学院情報学研究所, Graduate School of Informatics, Kyoto University

the user's internal state estimator improve the naturalness of the system utterances in both dialogue evaluation and utterance evaluation.

**Key Words:** *Dialogue System, User's Internal State, Knowledge, Interest, Engagement*

## 1 はじめに

近年深層学習を利用した対話システムが注目を集めており、大規模データを活用する対話研究が活発に行われている (Adiwardana et al. 2020; Smith et al. 2020). 現在主流の対話システムは大規模データをもとにして入力発話に対して尤もらしい応答を出力するように学習を行う。しかし、同じ入力発話であっても相手の発話意図によって応答すべき内容は異なる。つまり、対話システムは話者の意図を理解・解釈した上でその解釈した結果に応じて発話を行うべきであると考えられる。本研究では話者の意図理解に向けて、対話中の話者内部状態を取り扱い、以下の2つの課題に取り組む。

- (1) 話者内部状態のモデル化
- (2) 話者内部状態を踏まえた応答変更

一般に現実の対話ではたとえ雑談であっても「対話を通して情報を伝達する」や「相手に何らかの影響を与える」などの一定の目的がある。そしてその一貫した目的のもと、複数ターンのやり取りの中で情報が授受されることが自然である。つまり、話者内部状態のモデル化の際には複数ターンのやり取りがあり、かつ対話の目的がその中で一貫している対話データが必要になる。ただ、人間同士の対話は互いの意図が交錯するため、内部状態の分析・モデル化の題材にするのは難しい。一方、近年の深層学習による対話システムと人間の対話は一問一答レベルでは適切な応答ができるものの、複数ターンの対話になると多くの課題が表出する上、その対話には一貫した目的が存在しない場合がほとんどである。

そうした背景から本研究では、映画推薦をドメインとするルールベースの対話システムを構築し、対話システムと人間との目的の一貫した複数ターンの対話データの収集を行う。本研究の全体の流れを図1に示す。収集した対話データの分析に基づいて、対話中の話者内部状態を以下の3つの軸でモデル化する。

- **知識**：話題に関する知識があるかどうか
- **興味**：話題への興味があるかどうか
- **対話意欲**：対話に対して積極的に参加しているかどうか

話者の知識や興味を理解することで適切な情報提供や話題変更を行える。また、対話意欲を考慮することで、例えばユーザの対話意欲が高い場合は聞き役に徹するようにするなど、対話システムが適切な振る舞いができるようになる。

モデル化した話者内部状態を収集した対話データにアノテートし、これを学習データとして

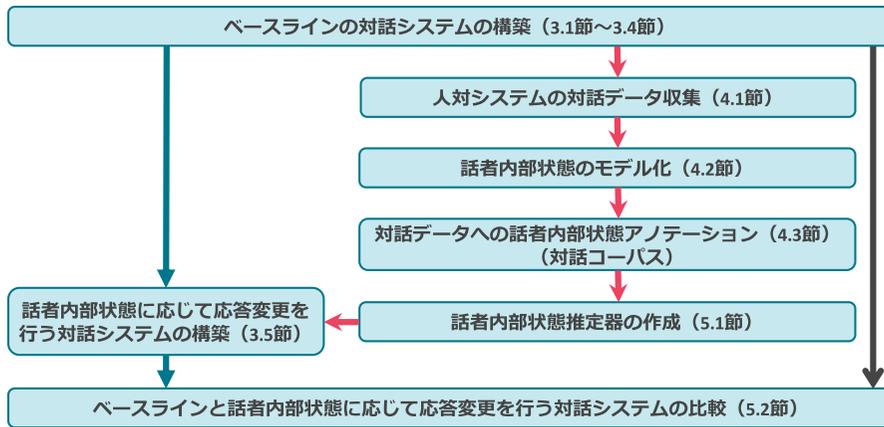


図 1 本研究の流れ

話者内部状態の自動推定を行った。その結果、いずれの内部状態においても 7 値スケールで ±1 のずれを許した場合、約 80~85% と高い推定精度を達成した。

また、話者内部状態の自動推定結果に応じた対話システムの応答の変更にも取り組む。具体的には、知識、興味、対話意欲のそれぞれについてその有無に応じた応答変更のルールを追加で用意する。話者内部状態を自動で推定し、その推定結果に応じて応答変更を行った場合にシステム発話の自然さが向上することを、対話単位での評価と発話単位での評価の両方で確認した。本研究の貢献は以下の 2 点である。

- 発話ごとにユーザ内部状態（知識、興味、対話意欲）を付与した 1 万発話規模のテキスト対話コーパスを構築した。
- ユーザ内部状態を自動推定し、その推定結果に応じてシステム発話を変更することの有効性を実証的に示した。

## 2 関連研究

対話においては、発話自体の内容を理解するのは勿論のこと、それを発話した話者の状態を理解することが重要である。ここでは発話の背後にある話者の内面の状態を「話者内部状態」と呼び、知識・興味・対話意欲の 3 つに注目する。

関連する概念として、対話行為 (Stolcke et al. 2000) がある。対話行為は、対話中の発話を「質問」「情報提供」「約束」といった心理的・社会的な何らかの効果を引き起こす行為として分類したものであり (中野 他 2015)、対話における話者の意図や目的の理解に広く用いられている (Bunt et al. 2012)。対話行為が発話のもたらす効果に注目しているのに対し、本研究で取り扱う話者内部状態は、その行為を選択するに至った話者の、より原始的な内面の状態である。

感情は対話研究において積極的に活用が検討されてきた話者内部状態であり, 発話からの感情の推定や, 感情を踏まえた発話生成の研究が盛んに行われている。徳久 他 (2009) は Web から獲得したコーパスを用いて発話から「嬉しい」「楽しい」などの 10 種類の感情の推定を行った。感情を踏まえた発話生成の研究では, Tweet 中の顔文字を感情のアノテーションとみなすことで感情を反映した発話生成を行う手法 (Zhou and Wang 2018) や, 感情の明示的表現と暗示的表現に着目し, 特定の感情を発話に反映する手法が提案されている (Song et al. 2019)。

一方で, 話者の発話だけからでは陽には読み取れない情報に着目したものとして, ペルソナ (Li et al. 2016; Zhang et al. 2018) がある。ペルソナは話者の年齢や性別といった個人的な背景やそれに基づく話し方などを指す。ユーザごとに ID 番号を割り振ることで話者の違いを学習する手法 (Li et al. 2016) や, ユーザの属性を短文数文程度にまとめることでペルソナのモデル化を行ったものがある (Zhang et al. 2018)。

このような感情とペルソナの活用は主に対話システム側の応答を豊かにすることを目的としており, 本研究の話者の意図理解とは目的が異なる。

ユーザの知識・興味・対話意欲について着目した研究もなされている。宮崎 他 (2017) はコールセンタの音声対話を具体例として知識量推定に有効な言語的・対話的特徴量について調査を行っている。またその調査に基づき, 対話全体から話者の知識量を推定し, 知識の少ない話者の対話を抽出する手法を提案している。

興味については, マルチモーダル情報を利用して興味を推定する研究がある (富増, 荒木 2016; 松本, 荒木 2018; 西本 他 2018)。また, Wizard-of-Oz 法を用いて収集したマルチモーダル対話に, 話題への興味の有無をアノテートしたデータについて分析した研究もある (荒木 他 2017; 駒谷 他 2018)。テキスト対話からユーザの興味を推定する研究としては, 人間同士の雑談対話において, 発話から決められたトピックについての興味の度合いを推定した研究がある (Inaba and Takahashi 2018)。Inaba and Takahashi (2018) は対話終了後に各被験者に 24 種類のトピックの興味の度合いを選択させることで 1 つの対話と 24 種類の各トピックの興味の度合いの紐付けを行っている。

対話意欲では千葉 他 (2015) がマルチモーダル対話データを分析し, 提示された質問や話題に対する話者の対話全体での対話意欲の有無を 70~80% で推定できることを示した。Ishihara et al. (2018) は同様にマルチモーダル情報を用いてインタビュー対話から話者の対話意欲の有無を推定した。

これらの先行研究では知識, 興味, 対話意欲のいずれか 1 つの推定に取り組んでいるが, 本研究では 3 つの話者内部状態を同時に扱う。また, これらの内部状態は対話の中で刻々と変化していくと考えられるため, 対話のターンごとに推定を行う。そして, その推定結果によって応答を適切に変更することで, ユーザの内部状態を解釈して自然な応答を行える対話システムを構築する。

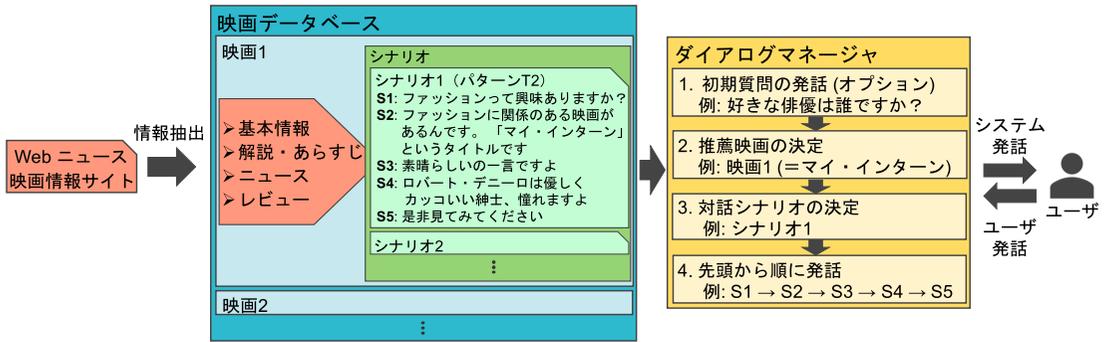


図 2 映画推薦対話システムの概要図

### 3 映画推薦対話システムの構成

話者内部状態の観察・分析のためには一貫した目的を持って複数ターンにわたって対話を続ける対話システムが必要となる。そこで映画推薦を目的としたルールベースの対話システムを構築した (図 2)。

本システムでは、いくつかのテンプレートを用意し、予め映画ごとにシステム発話の内容を作成しておく。この発話集をシナリオと呼ぶ。映画データベースはこのシナリオと、簡単な質問応答の際に必要な映画情報を保持する。対話管理はダイアログマネージャが担い、必要に応じて映画データベースに格納された映画情報を参照しながら対話を進める。本システムによる対話例を例 (1) に示す。

- (1) S1: ファッションって興味ありますか?
- U1: 興味ありますよ
- S2: ファッションに関係のある映画があるんです。「マイ・インターン」というタイトルです。
- U2: タイトルは聞いたことあるような気がします。
- S3: 素晴らしいの一言ですよ。
- U3: そうなんですね。
- S4: ロバート・デニーロは優しくカッコいい紳士、憧れますよ。
- U4: それはいいですね
- S5: 是非見てみてください

#### 3.1 対話戦略

現在の技術レベルでもある程度自然で意味のある映画推薦対話が可能となるように、本シス

テムは以下の3つの対話戦略に基づき動作する。

(1) **システム主導** システムはユーザ発話に対する簡単な応答をしつつ、基本的にはシナリオ通りに主導的に対話を進める。

(2) **間接的話題導入** システムは対話のメインの話題である映画のタイトルをいきなり言うのではなく、以下の3つのパターンのいずれかを用いてその周辺の話題から対話を始める（下線部は映画によって異なる。）

**T1：最近のエンタメニュース**

（例）俳優の染谷将太と女優の菊地凛子夫妻に第2子が誕生したことが11日までにわかったと話題です

**T2：映画のテーマ**

（例）飛行機って興味ありますか？

**T3：映画情報（俳優の名前など）**

（例）細田守監督をご存知ですか？

(3) **推薦ポイント** 推薦する映画を褒める文（推薦ポイント）を複数回発話することで、ユーザがその映画を見たいと思えるように念押しする。

## 3.2 映画データベース

映画データベースには213作品の基本情報、解説・あらすじ、ニュース情報、レビュー、シナリオがそれぞれ格納されている。基本情報、解説・あらすじ、レビューは映画レビューサイト<sup>1</sup>からWebスクレイピングで取得した。基本情報は映画のタイトル、キャスト、スタッフ、ジャンルなどの情報である。レビューは、映画レビューサイトのユーザが映画の感想や評価などを書き込んだ文章であり、各映画につき高評価順に300件ずつ取得した。ニュース情報は、エンタメとスポーツのニュースをニュースサイトからWebスクレイピングで取得した。このニュース情報は3.1節で述べたパターンT1のシナリオの作成に用いる。

## 3.3 対話シナリオ

シナリオは対話戦略に沿って映画ごとに複数作成した。シナリオの前半部分では、間接的に話題を導入し、推薦映画を提示する。間接的話題導入の発話は、3.1節で述べた3つのパターンT1～T3に基づき、映画データベース中の情報を用いて生成する。

パターンT1では、最近のエンタメニュースの話題で対話を始め、そのニュースの登場人物の出演映画を推薦する。映画データベースのニュース記事の最初の1文を抽出し、キャストの名前が含まれていたなら、「<最近のエンタメニュースの1文目>というニュースが話題になって

<sup>1</sup> <https://movies.yahoo.co.jp/>

いますね」と発話する。

パターン **T2** では、映画のテーマを用いて「<テーマ>って興味ありますか?」といった文で対話を始める。映画のテーマは映画データベースの解説・あらすじからキャスト・スタッフ等の人名を取り除いた上で tf-idf のスコアが最も高い名詞を選択する。ただし、一般的な語がテーマとして使用されるのを防ぐために、tf-idf の最高値が 0.35 を下回った場合には、その映画についてはパターン **T2** のシナリオは作成しない。

パターン **T3** では、映画データベースの基本情報に含まれている主要キャスト 2 名と監督の名前を用いて「<人名>さんってご存知ですか?」といった発話から対話を始め、その人物に関連する映画を推薦する。

シナリオ後半部分では、各映画のレビューから抽出した 2 つの推薦ポイント文を発話する。まず、レビューの文章を文単位で分割する。次にレビュー文は口調が統一されていないため、Juman++辞書<sup>2</sup>を用いて語尾を活用させ、丁寧口調（「です」または「ます」）に変換したのち、語尾に情報提供の口調の「よ」を追加する。さらに、映画を褒めている、かつある程度の長さの文を選ぶ。具体的には、まず独自に選んだ映画に関連するポジティブな単語（「傑作」、「面白い」など）50 語の単語ベクトルの和と各文の単語ベクトルの和のコサイン類似度を計算し、類似度と各文の文字数をかけ合わせたスコアを用いてスコア上位 100 文を抽出する。これを各映画について行い、各シナリオでは 100 文の中からランダムに 2 文を用いる。なお、単語ベクトルは Web テキスト約 98 億文で学習した word2vec を用いた。

最後に視聴を促す発話を以下の 5 つからランダムに選択する。

- 是非見てみてください
- 是非ご覧下さい
- 良かったら見てみてはいかがですか?
- 面白い作品なのでぜひ見て頂きたいです
- きっと楽しめると思いますよ

### 3.4 ダイアログマネージャ

ダイアログマネージャは、映画データベースの情報を参照しつつ、推薦映画の決定やシナリオの選択、シナリオに基づく発話を行う。

推薦映画の決定はユーザの好みを尋ねてからその返答に応じて決定する方式とランダムに決定する方式の 2 つを用意し、その割合を 8:2 とした。ユーザの好みを尋ねる際に用いる質問（＝初期質問）は以下からランダムに選択する。

- 好きな俳優は誰ですか?

<sup>2</sup> <https://github.com/ku-nlp/jumandic-grammar>

- 好きな女優は誰ですか？
- 好きな監督は誰ですか？
- 好きなジャンルは何ですか？
- 邦画と洋画どちらが好きですか？

初期質問をする場合には、例えばユーザが好きな女優を挙げたら、その女優が出演する映画の中から推薦映画を選ぶ。その後、用意されているシナリオから1つ選択し、そのシナリオに基づいて発話を行う。ユーザの発話からの人物名の抽出は正規表現によるマッチングを用いる。ユーザが「いない」等と答えた場合や、初期質問をしない場合には、推薦映画をランダムに決定する。

推薦映画の主演キャスト、他のキャスト、監督の名前やジャンルなどの簡単な質問をユーザがした場合には、映画情報データベースを参照してその応答を通常のシステム発話の前に挿入する。質問であるかどうかの判定にはNTTコミュニケーションズが提供するCOTOHA API<sup>3</sup>を用いた。COTOHA APIでは文タイプを平叙文、疑問文、命令の中から判定することができる。本研究ではユーザ発話の文タイプが疑問文である発話を質問とみなし、質問の種類の判定は正規表現ベースでマッチングを行う。

### 3.5 話者内部状態による応答変更

本節では話者内部状態推定器の推定結果に応じてシステムの応答を変更するルールについて説明する。話者内部状態推定器は以下の3つの手続きにより作成する。

- (1) 3.1節から3.4節までで述べた要素によって構成された映画推薦対話システムで対話を収集（詳細は4.1節）
- (2) 収集した対話に知識、興味、対話意欲のアノテーションを行い対話コーパスを構築（詳細は4.3節）
- (3) 構築した対話コーパスを用いて話者内部状態推定器を学習（詳細は5.1節）

本システムは予めシステム発話の内容が決められたシナリオに基づいて対話を進める。そのため、各話者内部状態の対象（「何に対して知識があるのか」、「何に対して興味がないのか」など）を前のシステムの発話からある程度想定することができる。そこで知識・興味・対話意欲の有無とそのそれぞれの対象に応じて適切な応答を予め用意しておく。表1に話者内部状態タイプと各応答変更の概要を示す。話者内部状態推定器により各話者内部状態を推定し、用意した適切な応答に変更することでシステム発話がより自然な応答となることを目指す。

#### 3.5.1 知識の有無による応答変更

知識の有無による応答変更には以下の4タイプを用意する。応答変更の判断基準となるユー

<sup>3</sup> <https://api.ce-cotoha.com/contents/index.html>

話者内部状態タイプ	応答変更概要
1. 人物の知識がない	人物についての簡単な説明を追加
2. 映画の知識がない	映画についての情報（公開年度）を追加
3. 映画の知識がある	発話の語尾を同意口調に変更
4. 映画の深い知識がある	発話をユーザがその映画の視聴経験があると想定した発話に変更
5. ニュースに興味がない	補足の発話を追加
6. テーマに興味がない	発話をユーザの好みを問う質問に変更
7. 人物に興味がない	発話をユーザの好み（キャスト・監督）を問う質問に変更
8. 対話意欲がない	発話を控えめな口調に変更

表 1 話者内部状態による応答変更一覧

ザ発話には隅付き括弧内にそのユーザ発話の知識の有無を記す。

1. 人物の知識がない 間接的話題導入のパターン **T3** の発話の直後のユーザ発話から知識がないことがわかる場合、ユーザはその人物を知らないと考えられる。この場合はその人物についての簡単な説明文を次の発話の前に挿入する。説明文はその人物についての Wikipedia 記事の第一文目を活用する。Wikipedia 記事の取得には MediaWiki API<sup>4</sup>を使用した。具体例を例 (2) に示す。  $S_{before}$  と  $S_{after}$  はそれぞれ応答変更をしなかった場合と応答変更をした場合のシステム発話を指す。また、主な変更点を  $S_{after}$  の下線部に示す。

- (2) S: ジョージ・ルーカスさんをご存知ですか？  
 U: 良く分かりません。【知識なし】  
 $S_{before}$ : ジョージ・ルーカスさんが監督をされている映画があります。「スター・ウォーズ」というタイトルです。  
 $S_{after}$ : ジョージ・ルーカスさんはアメリカの映画監督、映画プロデューサー、脚本家です。ジョージ・ルーカスさんが監督をされている映画があります。「スター・ウォーズ」というタイトルです。

2. 映画の知識がない 映画のタイトルを提示した直後のユーザ発話から知識がないことがわかる場合、ユーザはその映画を知らないと考えられる。この場合、例 (3) のようにその映画についての情報を次の発話の前に補足する。本研究では補足する情報として映画の公開年度を使用した。

- (3) S: 好きな女優は誰ですか？  
 U: 綾瀬はるかさんです  
 S: では綾瀬はるかさんが出演している映画を紹介しますね

<sup>4</sup> [https://www.mediawiki.org/wiki/API:Main\\_page/ja](https://www.mediawiki.org/wiki/API:Main_page/ja)

U: お願いします

S: 「本能寺ホテル」という映画です

U: 知らないです。【知識なし】

*S<sub>before</sub>*: 最後の京都の現在の町と過去が交差する景色は、人の営みと思いは綿々と繋がっていること、そして過去の人々のお陰で現在の平和があることを示唆していますよ

*S<sub>after</sub>*: 2017年に公開された作品です。最後の京都の現在の町と過去が交差する景色は、人の営みと思いは綿々と繋がっていること、そして過去の人々のお陰で現在の平和があることを示唆していますよ

**3. 映画の知識がある** 映画のタイトル提示および1つ目の推薦発話の直後のユーザ発話に知識がある場合、ユーザがその映画を知っていると考えられるため、その次の発話の語尾を情報提供の口調の「よ」から同意口調の「よね」に変更する(例(4))。映画のタイトル提示直後と1つ目の推薦発話直後の応答変更はそれぞれ独立に知識の有無を判定して行われる。

(4) S: 好きな俳優は誰ですか?

U: 藤原竜也

S: では藤原竜也さんが出演している映画を紹介しますね

U: はい、お願いします

S: 「僕だけがいない街」という映画です

U: はい、DVDで鑑賞しました【知識あり】

*S<sub>before</sub>*: 謎解きは勘のいい方ならすぐに分かるレベルですが、本作は物語のベースに暖かいメッセージがこめられていて感動しますよ

*S<sub>after</sub>*: 謎解きは勘のいい方ならすぐに分かるレベルですが、本作は物語のベースに暖かいメッセージがこめられていて感動しますよね

**4. 映画の深い知識がある** 例(5)に示すように映画のタイトル提示および2つの推薦発話の直後の全てで知識がある場合、ユーザはその映画を既に見たことがあると想定し、最後の視聴を促す発話をユーザに視聴経験があると想定した発話に変更する。変更する発話は以下の2つからランダムに選択する。

- また見てみてはいかがですか
- 是非もう一度見てみてください

(5) S: 好きな女優は誰ですか?

U: 薬師丸ひろ子

- S: じゃあ薬師丸ひろろ子さんが出演している作品を紹介しますね  
 U: よろしくお願ひします  
 S: 「コーヒーが冷めないうちに」というタイトルです  
 U: よく知っていますね【知識あり】  
 S: 設定のせいもあるかもしれないけど、やっぱり、本物の役者さんって言うのは、凄いのだと、改めて感じますよね  
 U: 年季を感じますね【知識あり】  
 S: 数ちゃんと新谷君の心が近づいていって結ばれる軸となるエピソードに、ほっと心が温まりますよね  
 U: 私もです【知識あり】  
 S<sub>before</sub>: 面白い作品なのでぜひ見て頂きたいです  
 S<sub>after</sub>: 是非もう一度見てみてください

### 3.5.2 興味の有無による応答変更

興味の有無による応答変更には以下の3タイプがある。応答変更の判断基準となるユーザ発話には隅付き括弧内にそのユーザ発話の興味の有無を記す。

5. ニュースに興味がない 間接的話題導入のパターン **T1** の発話の直後に興味がない場合、提示したニュースに興味がないと考えられる。ただ、ニュースに興味がない場合でも映画に興味を示すことは考えられるため、そのまま映画を推薦する。その際、ユーザの発話を無視したという印象を薄くするため、例(6)に示すように以下のいずれかの発話からランダムに選択した発話を次の発話の前に挿入する。

- 結構評判みたいですよ。
- 結構話題になっているみたいなのですが、

- (6) S: 俳優の木村拓哉と歌手の工藤静香の長女・C o c o m i が26日、自身のインスタグラムを更新し、久々に外出したことをつづったと話題となっています  
 U: そうなん【興味なし】  
 S<sub>before</sub>: 木村拓哉さん、「無限の住人」という映画に出演されています  
 S<sub>after</sub>: 結構評判みたいですよ。その木村拓哉さん、「無限の住人」という映画に出演されています

6. テーマに興味がない 間接的話題導入のパターン **T2** の発話の直後に興味がない場合、その映画のテーマに興味がないと考えられる。映画のテーマに興味がない場合、映画についても興味を示さないことが想定されるため、推薦する映画を変更する。その際にユーザの好みを把握するため、例(7)のように初期質問を行う。この初期質問は3.4節で述べた5つからランダムに

選択する.

- (7) S: タイムトラベルって興味ありますか?  
 U: いえ、あまり興味ありません【興味なし】  
*S<sub>before</sub>*: タイムトラベルに関係のある映画があります。「アバウト・タイム愛おしい時間について」というタイトルです  
*S<sub>after</sub>*: そうなんですね。では好きな映画監督は誰ですか?

7. 人物に興味がない 間接的話題導入のパターン **T3** の発話の直後に興味がない場合, その映画の監督または主要キャストに興味がないと考えられるため, 初期質問を行って推薦する映画を変更する. このとき, 例 (8) のように女優 (俳優/監督) の名前から対話を始めた場合は好きな女優 (俳優/監督) を尋ねる初期質問を行う.

- (8) S: サンドラ・ブロックさんをご存知ですか?  
 U: 知っていますが、あまり興味はないです。【興味なし】  
*S<sub>before</sub>*: サンドラ・ブロックさんが声優をしている映画があります。タイトルは「ミニオンズ」です。  
*S<sub>after</sub>*: そうなんですね。では好きな女優は誰ですか?

### 3.5.3 対話意欲の有無による応答変更

対話意欲の有無による応答変更は以下の1タイプである. 応答変更の判断基準となるユーザ発話には隅付き括弧内にそのユーザ発話の対話意欲の有無を記す.

8. 対話意欲がない 2回目の推薦発話の直後に対話意欲がない場合, これまでの推薦の効果薄いと考えられるため, 視聴を促す発話を控えめな言い回しである以下のいずれかの発話からランダムに選択して変更する. 例 (9) に具体例を示す.

- 騙されたと思って見てみてはいかがですか
- 見てみたら意外と面白いかもしれませんよ

- (9) S: ヘイリー・アトウェルさんをご存知ですか?  
 U: しらないな。  
 S: ヘイリー・アトウェルさんはイングランド生まれの女優です。ヘイリー・アトウェルさんが出演されている映画があります。タイトルは「プーと大人になった僕」です。  
 U: そうなんですね。  
 S: はい、プーだけじゃなくて、イーヨーやピグレット、他のキャラたちも、声に魅

力があるし、ユアンマクレガーのイギリス人らしいイントネーションも、原語ならではの雰囲気ですよ

U: へー

S: クリストファーロビンに邪魔者扱いされ、悲しそうな顔から、最後は絆を取り戻せて幸せそうな表情に変わるプーさんがキュートすぎてたまりませんよ

U: そうですか。【対話意欲なし】

*S<sub>before</sub>*: 是非見てみてください

*S<sub>after</sub>*: 見てみたら意外と面白いかもしれませんよ

## 4 コーパス構築

まず、話者内部状態による応答変更を行わないシステムで対話を収集し (4.1 節)、話者内部状態のモデル化を行う (4.2 節)、そして、収集した対話にモデル化した話者内部状態のアノテーションを行う (4.3 節)。

### 4.1 学習・分析用対話収集

本システムを用いてクラウドソーシング<sup>5</sup>で対話を収集した。ワーカーには事前に、213 作品の中から対話的に映画を薦めるシステムとの対話であることを伝えている。対話はシステム側から開始し、ワーカーが返答するという順に進める。最後はシステム側の発話で対話を終了する。

対話システムは外部サイトとして用意し、ワーカーはクラウドソーシングのタスク画面内の対話システムの URL をクリックし、外部サイトで対話を行う。また、対話終了後にクラウドソーシングのタスク画面上で以下の 5 段階評価のアンケートに回答する。

(1) 説得度: 推薦された映画を見たくなくなりましたか?

{5:見たい, 4:どちらかと言えば見たい, 3:どちらとも言えない, 2:どちらかと言えば見たくない, 1:見たくない}

(2) 流れの自然さ: 対話の流れは自然でしたか?

{5:自然だった, 4:どちらかと言えば自然だった, 3:どちらとも言えない, 2:どちらかと言えば不自然だった, 1:不自然だった}

外部サイト上でワーカー固有の ID を表示し、クラウドソーシングの回答欄に入力してもらうことで対話とアンケート結果の紐付けを行う。収集した対話例を表 2 の「対話」列に、その統計情報を表 3 に示す。全部で 1,060 対話を収集した。また、アンケート結果を表 4 に示す。説得度について 51.7%、流れの自然さについて 60.3%の対話が 4 以上の評価を獲得した。この結

<sup>5</sup> <http://crowdsourcing.yahoo.co.jp/>

対話	話者内部状態		
	知識	興味	対話意欲
S: お姫さまって興味ありますか?			
U: 興味あります	2	3	3
S: お姫さまに関係のある映画があります。「今夜、ロマンス劇場で」というタイトルです			
U: どんなストーリーですか?	-3	3	3
S: 内容は詳しく言えませんが、泣ける映画です。リアリズムを徹底的に求めるのではなく、綾瀬はるかへのセリフと音楽を楽しむのが良いと思いますよ			
U: 音楽興味ありですね	0	3	3
S: 悲劇ではないけど切ない結末だなと思いきや、白黒からフルカラーへの反転で大団円ですよ			
U: 素敵な展開ですね	-2	3	3
S: ぜひ見てみてください			

表 2 対話コーパス例 (対話コーパスは対話を収集したのち、別途話者内部状態をアノテートすることで作成。)

			システム	ワーカー
対話数	1,060	発話数	6,154	5,094
シナリオ数	836	異なり発話数	4,840	2,485
平均ターン数	10.6	形態素数	163,347	20,279
参加ワーカー数	432	異なり形態素数	5,123	1,786

表 3 収集対話の統計情報 (参加ワーカー数は Yahoo!クラウドソーシングの会員 ID で区別して計算。形態素分割には Juman++ (Tolmachev et al. 2020) を使用)

スコア	説得度 (%)	流れの自然さ (%)
5	16.0	23.0
4	35.7	37.3
3	29.2	15.7
2	10.8	15.9
1	8.4	8.1

表 4 学習・分析用対話のアンケート結果

果から本システムが映画推薦という対話の目的を十分に達成しながら、ある程度自然な対話を実現できていることを確認できた。

## 4.2 話者内部状態のモデル化

前節で収集した対話を用いて、より自然な対話を実現するために必要な要素について分析を

行った。その結果、話者の「知識」・「興味」・「対話意欲」の3つの内部状態を把握し、それぞれの内部状態の有無に応じて適切な応答を行うことが対話の自然さの向上につながると考えた。先行研究においては、知識、興味、対話意欲のそれぞれの重要性は示唆されているものの(宮崎他 2017; Inaba and Takahashi 2018; 千葉 他 2015)、各々を独立に取り扱っている。ただ、これらの内部状態を統合的に考慮することで、話者の内部状態をより豊かに表現でき、より詳細な話者の意図理解につながると考えられる。

また、本研究で提案する話者内部状態モデルは、それぞれの内部状態が旅行や音楽などの複数の話題があるインタビュー対話(千葉 他 2015)や雑談対話(Inaba and Takahashi 2018)などで取り扱われていることから、今回の映画推薦のドメインに依存せず、雑談を含む他のドメインにおいても汎用的に活用できると考えられる。

本研究で用いる各話者内部状態の定義は以下の通りとする。

- 知識：話題に関する知識があるかどうか
- 興味：話題への興味があるかどうか
- 対話意欲：対話に対して積極的に参加しているかどうか

なお本研究では、システムの発話によってユーザの内部状態は刻々と変化すると考えられるため、ユーザが発話するごとにその内部状態を考慮する。

### 4.3 話者内部状態アノテーション

クラウドソーシングを用いて、収集した対話の全てのユーザ発話に話者内部状態をアノテートする。アノテータは対話履歴を見ながら、各ユーザ発話から読み取れる知識・興味・対話意欲の有無をアノテートする。話者内部状態のアノテーションはあり/なしの2値(Yoshino and Kawahara 2015; Inoue et al. 2018)、もしくはあり/どちらでもない/なしの3値(荒木 他 2017)で行われることが多いが、本研究では有無を断定できない発話があることを考慮し、3値でのアノテーションを行う。アノテータは以下に示す3段階の選択肢(それぞれが1点、0点、-1点に対応)から設問に対して最も適切なものを選択する。

設問： 発話から読み取れる内部状態を下記の選択肢から選んでください。

- 知識：{1:知識あり, 0:特に読み取れない, -1:知識なし}
- 興味：{1:興味あり, 0:特に読み取れない, -1:興味なし}
- 対話意欲：{1:対話への関与が積極的, 0:どちらとも言えない, -1:対話への関与が消極的}

アノテーションは各発話に対して3人ずつで行う。アノテーションには458名のアノテータが参加した。スコアの集約では各アノテータの多数決を採るという方法もあるが、今回取り扱う話者内部状態では「知識がありそう」などのように、3段階の中にさらにそれぞれの中間的な状態があると考えられるため、3人のアノテータのスコアを合計した3から-3までの7段階の

スコアをアノテーションとする。アノテーションした対話コーパスの例を表 2 に、各内部状態のスコアの分布を表 5 に示す。興味と対話意欲に関してはスコアが高い傾向にあったが、知識に関してはほぼ一様に分布している。

本研究で扱う話者内部状態のアノテーションにはアノテータの主観が介入すると考えられる。そこで、アノテータ間の一致率を測ることでアノテーションの信頼性を検証する<sup>6</sup>。ここではアノテータが各発話に対して 3 人であることと、アノテーションの尺度水準 (Stevens 1946) が順序尺度とみなせることから Krippendorff's  $\alpha$  (Krippendorff 2004) を指標として用いる。Krippendorff's  $\alpha$  は 2 人以上のアノテータ間の一致度を求めるための指標で、スコア間距離 (= 不一致度) を尺度水準によって変更することができるため高い汎用性を持つ。本研究では順序尺度のスコア間距離を用いて Krippendorff's  $\alpha$  を計算する。

全データ (*Full* と呼ぶ) でのアノテータ間の一致率を表 6 の *Full* 列に示す。いずれの内部状態についても  $\alpha$  は 0.40 程度であった。一般に社会学では  $\alpha$  が 0.80 を超える場合に信頼性を持ったデータであると結論付けることができる。しかし言語処理分野においては、マルチモーダル対話データに興味の有無をアノテーションした先行研究 (駒谷 他 2018) で、尺度水準を間隔尺度としているため単純比較はできないものの、 $\alpha$  が 0.50 程度となっている。このアノテーションは専門家によって行われており、不特定多数の人間が参加できるクラウドソーシングによるアノテーションより品質が高いと考えられる。一方, Chollet et al. (2016) は母語話者と非母語話者のプレゼンテーション能力についてクラウドソーシングによるアノテーションを行ってお

スコア	知識	興味	対話意欲
3	13.4% (684)	20.2% (1,030)	20.4% (1,039)
2	15.3% (781)	22.2% (1,130)	19.7% (1,006)
1	15.6% (793)	18.2% (929)	17.8% (906)
0	14.4% (735)	13.3% (680)	14.1% (716)
-1	15.8% (807)	11.4% (579)	12.0% (609)
-2	14.0% (711)	8.7% (443)	9.5% (486)
-3	11.4% (583)	5.9% (303)	6.5% (332)

表 5 話者内部状態の分布 (括弧内は発話数)

内部状態	<i>Full</i>	<i>Filtered</i>
知識	0.41	0.67
興味	0.40	0.59
対話意欲	0.35	0.63

表 6 アノテータ間の一致率

<sup>6</sup> 458 名のアノテータがアノテーションを行ったため、各アノテータの、他のアノテータとのアノテーションの一致率を評価した (付録 A)。その結果、極端に信頼性が低いアノテータはごく少数であったため、付与したアノテーションはすべてそのまま使用した。

り、言語処理の主観評価タスクでは  $\alpha$  は 0.40 程度で妥当であると報告している。したがって、クラウドソーシングによってアノテーションを行った本研究の  $\alpha$  が 0.40 程度であるのは妥当であると考えられる。

また、アノテータ間でスコアが  $-1$  と  $1$  とに割れている発話をコーパスから取り除いてフィルタリングしたデータを *Filtered* とする。*Filtered* のアノテータ間の一致率を表 6 の *Filtered* 列に示す。*Filtered* データを用いることでアノテータ間の一致率の高い、品質の良いデータのみから学習できるという利点がある。その一方で、アノテータ間で意見が割れる難しい事例を取り除くことにもなるため、そうした事例が実際に出現した際に対応が出来ない可能性がある。そこで、*Full* と *Filtered* によって学習した推定器の精度を実験で比較し、フィルタリングの効果について検証する。

## 5 実験

構築した対話コーパスを利用して、話者内部状態を推定しながらその結果に応じて応答を変更する対話システムを構築する。まず、コーパスから話者内部状態の推定を行う推定器を学習する。次にこの推定器をこれまでに構築した対話システムに組み込み、話者内部状態に応じて応答を変更できるようにする。最後に、構築した応答変更機能付きの対話システムで対話を収集し、その質の評価を行う。

### 5.1 話者内部状態推定

#### 5.1.1 話者内部状態推定器の作成

話者内部状態の推定には BERT (Devlin et al. 2019) を用いる。BERT は Transformer (Vaswani et al. 2017) をベースとし、大規模な生コーパスで事前学習した後、各タスクで fine-tuning することで自然言語処理の様々なタスクで SOTA を達成している。本研究では、NICT BERT 日本語 Pre-trained モデル BPE あり<sup>7</sup>を利用する。この BERT モデルは日本語 Wikipedia 全文に対して、半角を全角に正規化したのち、形態素分割および subword 分割を行い、110 万ステップ pre-training したものである。

話者内部状態推定推定器を図 3 に示す。それぞれの話者内部状態の有無は、「あり」「なし」といったラベルで予測するより、「知識がありそう」などの中間状態を表現しやすい連続値のスコアで予測する方が良くと考え、分類モデルではなく回帰モデルで推定を行う。推定対象のユーザ発話とそれ以前の対話履歴を入力として対象のユーザ発話の話者内部状態推定スコアを出力する。具体的には、推定対象のユーザ発話の先頭と末尾にそれぞれ [CLS] トークンと [SEP] トー

<sup>7</sup> <https://alaginrc.nict.go.jp/nict-bert/index.html>

クンを挿入し, 対話履歴はより新しいものを左から順に最大 512 トークンまで入力する. また, 各システム発話, ユーザ発話の前にはそれぞれ分離トークン [S], [U] を追加する. 出力は [CLS] トークンに対応する 768 次元の隠れ状態ベクトルを 1 次元に線形変換を行った実数値が推定スコアとなる. 3 を上回る推定スコアは 3 に, -3 を下回るスコアは -3 に変換する. 損失関数には二乗誤差関数を採用し, 推定スコアと正解スコアの差が小さくなるように学習を行う. 話者内部状態推定器は知識, 興味, 対話意欲のそれぞれ個別に学習を行い, 合計で 3 つ作成する.

### 5.1.2 実験設定

対話コーパス 1,060 対話を学習データ : 開発データ : テストデータ = 8 : 1 : 1 の割合で分割する. 各データの発話数を表 7 に示す. *Filtered* は *Full* と比べて約 8 割程度のデータ量であった. 評価指標は以下の 2 つを用いる.

- **Acc:** 推定スコアと正解スコアの差が  $\pm 0.5$  以下である割合
- **Broad Acc:** 推定スコアと正解スコアの差が  $\pm 1.5$  以下である割合

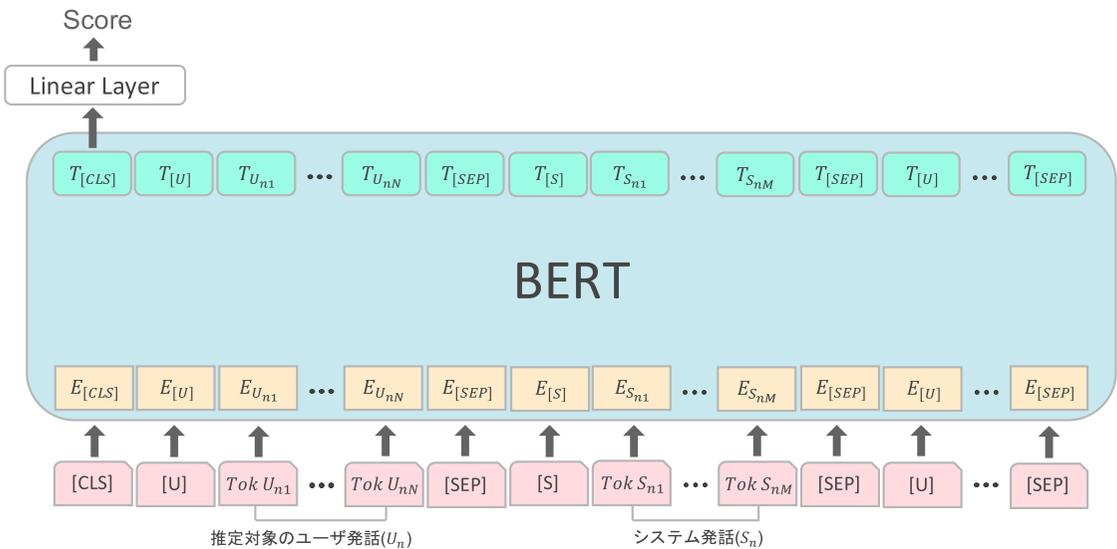


図 3 話者内部状態推定器 ([S], [U] はそれぞれシステム発話, ユーザ発話の分離トークンを指す)

内部状態	<i>Full</i>	<i>Filtered</i>
知識	5,094 (4,082/511/501)	4,073 (3,266/410/397)
興味	5,094 (4,082/511/501)	4,292 (3,424/432/436)
対話意欲	5,094 (4,082/511/501)	3,926 (3,134/396/396)

表 7 *Full* および *Filtered* の発話数. 括弧内の数字はそれぞれ学習データ, 開発データ, テストデータの発話数を表す.

Acc は 7 値分類の正解率に, Broad Acc は 7 値分類においてスコア  $\pm 1$  まで誤差を許容した正解率に対応している.

開発データを使用して, Devlin et al. (2019) の設定に従い, 下記の条件で fine-tuning のハイパーパラメータチューニングを行う. 開発データでの Acc が最も良い推定器を最終的な推定器として使用する. なおドロップアウト率は実験を通して 0.1 で固定する. その他のパラメータについては, NICT BERT 日本語 Pre-trained モデル BPE ありのデフォルトの値を用いる.

- バッチサイズ: 16, 32
- 学習率 (AdamW (Loshchilov and Hutter 2019)):  $5e-5, 3e-5, 2e-5$
- エポック数: 2, 3, 4

### 5.1.3 結果

各話者内部状態の推定結果を表 8 に示す. 学Full-テFull は学習・テストともに Full を用いる推定器で, いずれの内部状態においても Acc は 30%, Broad Acc は 70~80%程度であった. 表 5 より知識, 興味, 対話意欲のマジョリティベースラインがそれぞれ 15.6%, 22.2%, 20.4%であることを考慮すると, 各内部状態を妥当に推定できているといえる. 次に Full データで学習し, Filtered のテストデータでテストを行った学Full-テFiltered では学Full-テFull と比較して Acc, Broad Acc とともに正解率が向上している. これは Filtered がアノテーションの一致率が高い, つまり人間にとっても各話者内部状態の有無を認知しやすいデータとなっていて, 推定がしやすくなったためであると考えられる.

さらに, Filtered データで学習を行った学Filtered-テFull と学Filtered-テFiltered はそれぞれ学Full-テFull, 学Full-テFiltered と比べて, 興味については Acc, Broad Acc が多少下がっているものの, 知識・対話意欲については向上している. フィルタリングにより学習データ量が減少しているにも関わらず, 複数の内部状態で精度が向上していることから, データのフィルタリングは内部状態推定の精度向上に効果があることが分かる. 興味については, 元々知識や対話意欲と比べて推定精度が高いためフィルタリングによる効果が小さかったと推察される. また, 表 6 より, フィルタリング後の興味のアノテータ一致率が知識や対話意欲と比べて低かったことも精度が上がらなかった原因であると思われる.

内部状態	学Full-テFull		学Full-テFiltered		学Filtered-テFull		学Filtered-テFiltered	
	Acc	Broad Acc	Acc	Broad Acc	Acc	Broad Acc	Acc	Broad Acc
知識	29.1	73.7	31.5	76.1	30.9	74.5	36.3	80.6
興味	32.9	82.0	36.0	85.8	31.5	79.0	33.5	83.0
対話意欲	28.3	72.5	32.1	78.5	30.7	75.8	36.4	84.8

表 8 ユーザ内部状態の推定結果. セル内の数字はパーセント表記.

### 5.1.4 分析

図 4 に推定スコアと正解スコアの絶対誤差をシナリオのターン別にまとめる。例えば「タイトル提示」行の場合、システムがタイトルを提示した直後のユーザ発話を推定対象の発話としている。また、初期質問の有無により推薦映画の決定方法が異なるため、以下のように分割して考察する。

初期質問あり (成功) : 初期質問を行って、ユーザの好みに応じて推薦映画を決めた場合

初期質問あり (失敗) : 初期質問を行ったが、システムがランダムに推薦映画を決めた場合 (ユーザの好みに合致する推薦映画が映画データベースになかった場合など)

初期質問なし : 初期質問を行わず、システムがランダムに推薦映画を決めた場合

図 4 より「初期質問あり (成功)」の場合は知識、興味、対話意欲のいずれについても誤差が対話全体にわたって小さい。興味、対話意欲に関しては正解スコアの平均が 2 ポイント前後であり、「初期質問あり (成功)」は高い興味と対話意欲の獲得につながると考えられる。また知識についても「初期質問あり (失敗)」や「初期質問なし」と比較すると正解スコアの平均が 1 ポイント近く高くなっている。これは好きな俳優や監督などについての映画はある程度ユーザにも知識があるためと考えられる。このように平均的に各内部状態のスコアが高くなっているため、推定器も推測がしやすくなり、誤差が小さくなっていると推察できる。

一方、知識の「初期質問あり (失敗)」と「初期質問なし」は対話が進むにつれて推定誤差が大きくなっている。例 (10) に実際の対話例を示す。最後のユーザ発話  $U_4$  に対して正解スコアは  $-2$ 、推定スコアは  $2.22$  であった。この  $U_4$  のように陽に話者内部状態が読み取れない発話の場合は文脈の解釈が重要になる。その文脈からは、 $U_2$  よりこのユーザは推薦された映画は見えていないことが分かり、 $U_4$  の直前のシステム発話  $S_4$  では映画の中のシーンの説明をしていることが分かる。つまり見えていない映画のシーンに関しては知識がないだろうと推定することができるが、この推定には知識の高度な構造化が必要である。この構造化に向けた知識の推定対象の詳細化は今後の研究課題である。

	知識			興味			対話意欲		
	初期質問あり (成功)	初期質問あり (失敗)	初期質問なし	初期質問あり (成功)	初期質問あり (失敗)	初期質問なし	初期質問あり (成功)	初期質問あり (失敗)	初期質問なし
話題導入	0.93 [8] (0.00, 1.43)	0.89 [55] (-0.35, 4.67)	0.92 [20] (-0.30, 4.96)	0.76 [11] (2.09, 0.89)	1.19 [60] (-0.32, 4.02)	0.87 [25] (0.04, 4.71)	0.78 [12] (1.92, 1.72)	0.94 [54] (-0.06, 4.51)	0.96 [20] (-0.20, 4.80)
タイトル提示	0.72 [13] (1.08, 5.91)	0.97 [45] (-0.40, 4.61)	0.88 [23] (0.43, 6.08)	0.79 [12] (2.00, 0.91)	0.98 [56] (-0.07, 3.67)	1.07 [26] (0.31, 2.22)	0.78 [13] (1.77, 2.53)	1.11 [50] (-0.16, 5.44)	1.11 [20] (0.35, 2.87)
推薦ポイント①	0.69 [11] (0.82, 5.56)	1.19 [49] (-0.41, 4.33)	1.15 [23] (-0.60, 3.73)	0.44 [13] (2.23, 1.36)	1.10 [52] (0.58, 4.05)	0.83 [22] (0.64, 2.34)	0.56 [10] (2.50, 0.28)	0.97 [46] (0.26, 5.31)	0.95 [19] (0.26, 4.54)
推薦ポイント②	0.71 [11] (1.27, 3.22)	1.35 [50] (-0.34, 4.02)	1.44 [20] (-0.65, 4.33)	0.73 [10] (2.50, 0.50)	0.94 [55] (0.71, 4.51)	0.81 [24] (0.79, 2.87)	0.60 [10] (2.50, 0.50)	1.01 [53] (0.53, 4.95)	0.92 [23] (0.74, 4.11)

図 4 話者内部状態推定スコアと正解スコアの絶対誤差の平均値。角括弧内の数字はサンプル数を指す。また、各セル内の丸括弧内の数字はそれぞれサンプルの正解スコアの平均、不偏分散を示す。絶対誤差の値が大きいほど色が濃くなるようにカラースケールで色付けしている。

- (10) S1: 俳優の染谷将太と女優の菊地凜子夫妻に第2子が誕生したことが11日までにわかったと話題です  
U1: おめでたいです  
S2: 染谷将太さん、「バクマン。」という映画に出演されていましたね  
U2: バクマン。は観ていませんが確か漫画原作ですよ。  
S3: 話しも素直に友情、青春、ライバル関係、生い立ち等も分かりやすく展開され好感が持てますよ  
U3: 映画は数年前でしたか。DVDで観てみたいと思います。  
S4: ちょっと分からないんですが、無駄なシーンが一切なく、サクサク進んでテンポがいいので、気持ちいいです  
U4: 期待できます。

## 5.2 対話および発話単位での評価

### 5.2.1 話者内部状態推定の導入

話者内部状態の推定器を対話システムに導入する。知識・興味・対話意欲のそれぞれに対して有無を判定し、その結果に基づいて3.5節で述べたようにシステムの応答を変更する。総合的な推定精度を重視して今後の実験では *Filtered* データで学習した推定器を用いる。知識・興味・対話意欲の有無の判定は閾値を設け、正の閾値を上回ったらその内部状態について「あり」、負の閾値を下回ったらその内部状態について「なし」と判定する。本研究では、知識・興味については正の閾値を1.5、負の閾値を-1.5に、対話意欲については正の閾値を1.0、負の閾値を-1.0に設定した。また、適用すべき応答変更のルールが複数存在する場合は、推定スコアの絶対値が大きい内部状態での応答変更を優先して行う。

### 5.2.2 対話収集と対話単位での評価

話者内部状態推定による応答変更機能を組み込んだ対話システムで4.1節と同様にクラウドソーシングで299対話を収集した。また、比較対象として応答変更を行わないシステムでも297対話を収集した。この際、この2つのシステムに共通するモジュールを元の対話コーパス構築時から一部変更した。詳細は付録Bに記す。

対話終了後に推薦された映画の視聴経験があるかどうかについてワーカーにアンケートを行ったところ、全体の約67%のワーカーは推薦された映画の視聴経験がなかった。また、映画推薦対話はタスク指向と非タスク指向の両方の性質を合わせ持っていると考えられる。本研究ではシステムの発話の自然さについて特に着目しているが、映画推薦としてのタスクの達成度を下げることなくシステム発話の自然さが向上することが望ましい。そこで、以下に示す3つの5段階リッカート尺度のアンケート（5が最高評価、1が最低評価）を実施した。

指標	応答変更あり	応答変更なし
説得度	3.44	<b>3.48</b>
自然さ	<b>3.46</b>	3.20
満足度	<b>3.34</b>	3.15

表 9 評価実験のアンケート結果. スコアは各ワーカーのスコアの平均値.

- (1) 説得度: 推薦された映画を見たくなくなった
- (2) 自然さ: システムの返答は自然だった
- (3) 満足度: システムからの返答に満足した

アンケート結果を表 9 に示す. 説得度については応答変更ありと応答変更なしの間で大きな差は見られなかった. 一方, 自然さについては応答変更ありの方が応答変更なしのスコアを 0.26 ポイント, 満足度については 0.19 ポイント上回った. スコアの差が大きい自然さと満足度について有意水準を 5%とするウィルコクソンの順位和検定を行ったところ, それぞれ p 値は 0.017 と 0.123 となり, 自然さについて有意にスコアが向上していることが分かった. この結果から, ユーザの内部状態を推定し, その結果に応じて応答を変更することで, 映画推薦のタスクに関して一定の達成度を保ちつつ, システムの発話の自然さを向上させられたことが分かる.

### 5.2.3 発話単位での評価

本システムで構築した対話システムは基本的にシナリオベースであるため, 応答変更が起きている分岐点以外は応答変更ありと応答変更なしのシステムは全く同じ応答をする. そこで収集した評価用対話データからそれぞれの話者内部状態タイプの分岐点において, 変更した応答と変更しなかった応答 (= シナリオ通りの応答) をペアで抜き出し, ペアワイズで比較を行う. 具体的には, 応答変更ありの対話から抜き出す場合は使用したシナリオから応答を変更しなかった場合の発話を取り出して元の変更した応答と比較する. 応答変更なしの対話から抜き出す場合は内部状態推定を行い, 応答の変更が行われる分岐点を見つけ, 変更した応答と変更しなかった元の応答を比較する.

各話者内部状態タイプに対して応答ペアの抽出を行った結果を表 10 に示す. 応答変更あり対話と応答変更なし対話のそれぞれから最大 15 ペアずつ, 合計最大 30 ペアずつサンプルし, 評価に用いる.

発話単位での評価についてもクラウドソーシングで実施する. クラウドソーシングの回答画面を図 5 に示す. ワーカーは対話履歴と, 変更した応答, 変更しなかった応答を見てどちらが自然か選択する. ワーカーにはどちらが変更した応答かどうかは伏せてあり, 「どちらの返事も同程度に自然で選べない」「どちらの返事も同程度に不自然で選べない」の 2 つの選択肢も追加

話者内部状態タイプ	応答ペア数	
	応答変更あり対話	応答変更なし対話
1. 人物の知識がない	58	48
2. 映画の知識がない	121	124
3. 映画の知識がある	201	175
4. 映画の深い知識がある	56	42
5. ニュースに興味がない	5	5
6. テーマに興味がない	19	15
7. 人物に興味がない	2	0
8. 対話意欲がない	57	51
全体	519	460

表 10 各話者内部状態タイプの分岐点の統計情報

以下の対話を読んで、最後のAさんの返事として自然な方を選択肢から選んで下さい。

設問ID:174

【対話】

Aさん: 邦画と洋画どちらが好きですか?

Bさん: 邦画です

Aさん: では邦画を紹介しますね。呪いって興味ありますか?

Bさん: ありません

Aさん: 【選択肢の中から1つ選択して下さい】

---

呪いに関係のある映画があります。「もののけ姫」というタイトルです

そうなんです。では好きな女優は誰ですか?

(どちらの返事も同程度に自然で選べない)

(どちらの返事も同程度に不自然で選べない)

図 5 クラウドソーシングによる発話評価の回答画面

した。1 応答ペアごとに 10 人のワーカーで評価を行う。

発話単位での評価実験の結果を表 11 に示す。全ての話者内部状態タイプで応答変更ありの方が自然さが上回っていることがわかる。この結果からも、提案した話者内部状態を推定し、その結果に応じて応答を変更することでより自然な返答ができる対話システムを構築できたといえる。

### 5.2.4 エラー分析

話者内部状態タイプ 3, 4 については応答変更ありの方がスコアは高かったものの、その効果は他の話者内部状態タイプと比較すると低い。そこで、それぞれに対してその原因を分析する。

話者内部状態タイプ	応答変更あり	応答変更なし	どちらも自然	どちらも不自然
1. 人物の知識がない (30)	<b>240</b>	33	11	16
2. 映画の知識がない (30)	<b>180</b>	44	32	44
3. 映画の知識がある (30)	<b>87</b>	84	50	79
4. 映画の深い知識がある (30)	<b>135</b>	111	11	43
5. ニュースに興味がない (10)	<b>54</b>	11	6	29
6. テーマに興味がない (30)	<b>170</b>	71	6	53
7. 人物に興味がない (2)	<b>11</b>	8	0	1
8. 対話意欲がない (30)	<b>134</b>	100	30	36
全体 (192)	<b>1,011</b>	462	146	301

表 11 各話者内部状態タイプの自然さ. 表内の数字はワーカーの評価の合計を, 話者内部状態タイプの括弧内の数字はサンプル数を指す.

3. 映画の知識がある この話者内部状態タイプでは語尾を情報提供の口調の「よ」から同意口調の「よね」に変更する。「どちらも自然」のスコアがある程度高かったことを判断材料とすると, 語尾の文字を1字変えただけでは大きな印象の差を与えられなかったと推察される. また, 「どちらも不自然」のスコアも「応答変更あり」, 「応答変更なし」と同程度に高かった. これはレビューから抽出してきた推薦ポイント文が文脈と合わなかったためと考えられる. 例 (11) にワーカー 10 名中 6 名が「どちらも不自然」と判定した例を示す. この例では応答を変更する直前にユーザが「もっと前の映画が好きです。」と言っているのにも関わらず, その発話を無視しているためシステムの発話が不自然に感じられたと考えられる. 今後, 文脈を用いて検索ベースで推薦ポイントを選択するなどの改善手法が考えられる.

- (11) S: 好きな映画監督は誰ですか?  
 U: スピルバーグ  
 S: ではスティーヴン・スピルバーグさんが監督の映画を紹介しますね  
 U: お願いします。  
 S: 「ブリッジ・オブ・スパイ」という映画です  
 U: 知りませんが、どんな内容ですか。  
 S: 2015年に公開された作品です。スティーヴンスピルバーグ監督にコーエン兄弟の脚本、主演がトムハンクスと期待せずにはいられない組み合わせに、見事に期待に応える良作だと評判ですよ  
 U: もっと前の映画が好きです。  
*S<sub>before</sub>*: ベルリンの壁、ベルリンにかかる橋、チェックポイント、東ベルリン街中などなど当時の感じを丁寧に表現撮影しており見ごたえがありますよ  
*S<sub>after</sub>*: ベルリンの壁、ベルリンにかかる橋、チェックポイント、東ベルリン街中などなど当時の感じを丁寧に表現撮影しており見ごたえがありますよね

4. 映画の深い知識がある 映画の深い知識があると判断した場合、システムが推薦している映画をユーザが既に見たことがあると考え、視聴を促す発話を「是非もう一度見てみて下さい」といったように視聴経験があることを想定した発話に変更する。この場合では「どちらも不自然」のスコアはさほど高くないものの、「応答変更なし」のスコアが高くなっている。この原因としてはユーザの視聴経験の推定誤りが考えられる。例 (12) ではユーザは「アウトレイジ」シリーズは視聴経験があるが、推薦映画である「アウトレイジ最終章」については視聴経験はない。システムは知識ありと判定してしまったが、このように視聴経験がないユーザに「是非もう一度見てみて下さい」と発話するのはかなり不自然に感じられる。

- (12) S: お笑いタレントのビートたけしが30日放送のTBS「新・情報7DAY5ニュースキャスター」にリモート生出演と話題となっています
- U: そうなんですか。知らなかったです
- S: はい、ビートたけしさん、「アウトレイジ最終章」という映画に出演されています
- U: アウトレイジは見たことがあります。でも最終章はみてないかも
- S: 韓国フィクサーの立ち位置が今一つで、タケシも今までより控えめですが、西田敏行のズル賢い役者ぶりが目立ちますよね
- U: 韓国人がでてくるんですね
- S: 過去の出演者達がみんな歳をとって迫力不足だが、小心者のピエール瀧が今回は盛り上げてくれますよね
- U: ピエール瀧がアウトレイジというのはなんだか真実味があります
- S<sub>before</sub>*: 是非見てみてください
- S<sub>after</sub>*: 是非もう一度見てみてください

発話評価に用いた30サンプルをそれぞれのユーザの視聴経験のアンケート結果を照らし合わせると、30サンプル中23サンプルが実際に「視聴経験あり」で正しく視聴経験を推定できていた一方で、残りの7サンプルが「視聴経験なし」で推定が誤っていた。表12に、ユーザの実

視聴経験	応答変更あり	応答変更なし	どちらも自然	どちらも不自然
視聴経験あり (23)	<b>122</b>	70	6	32
視聴経験なし (7)	13	<b>41</b>	5	11
総計 (30)	<b>135</b>	111	11	43

表 12 視聴経験の有無別の自然さ。表内の数字はアノートの素点を、視聴経験の括弧内の数字はサンプル数を指す。

際の視聴経験の有無で自然さのスコアを比較した結果を示す。「視聴経験なし」は「応答変更なし」の方がスコアが高くなっており、発話の自然さに対するユーザの視聴経験の推定誤りの影響が大きいことが分かる。

## 6 おわりに

本研究では相手の発話の適切な意図解釈に向けて対話中の話者内部状態のモデル化に取り組んだ。収集した対話データの分析結果から知識、興味、対話意欲の3つの軸で話者内部状態をモデル化した。また、モデル化した3つの内部状態をクラウドソーシングによって対話データにアノテートすることで対話コーパスを構築した。構築した対話コーパスで学習した話者内部状態推定器は、推定対象の話者の発話と対話履歴から提案した内部状態を高い精度で推定できることを示した。

さらに、提案した話者内部状態に応じて応答を変更する対話システムの構築を行った。知識・興味・対話意欲それぞれの有無に応じて応答を適切に変更するルールを設計した。学習した話者内部状態推定器の推定結果を利用し、設計したルールに基づいて応答を適切に変更することで、より自然な応答を行えるシステムが構築できたことを対話単位での評価と発話単位での評価の両方で明らかにした。

ニューラルネットワーク技術が発展してきた近年、ニューラルモデルにはモデルの振る舞いや予測根拠を明らかにする「説明可能性」が求められつつある。今後、我々も対話システムの全体をニューラル化することを検討しているが、その際にも本研究で行ったユーザ内部状態のモデル化はシステムの振る舞いの根拠として意味を持ち、「説明可能性」に向けた研究の足がかりになると考えている。

## 謝 辞

この研究は国立情報学研究所 (NII) CRIS と LINE 株式会社とが推進する NII CRIS 共同研究の助成を受けて行った。

## 参考文献

Adiwardana, D. D. F., Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y., and Le, Q. V. (2020). “Towards a Human-like Open-Domain Chatbot.” *ArXiv*, **abs/2001.09977**.

- 荒木雅弘, 富増紗也華, 中野幹生, 駒谷和範, 岡田将吾, 藤江真也, 杉山弘晃 (2017). マルチモーダル対話データの収集と興味判定アノテーションの分析. *SIG-SLUD*, **B5** (02), pp. 20–25. [M. Araki et al. (2017). Collection of Multimodal Dialog Data and Analysis of the Result of Annotation of Users’ Interests. *SIG-SLUD*, B5 (02), pp. 20–25.].
- Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A. C., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. (2012). “ISO 24617-2: A Semantically-based Standard for Dialogue Annotation.” In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC’12)*, pp. 430–437.
- 千葉祐弥, 能勢隆, 伊藤彰則 (2015). ユーザの対話意欲自動推定を目標とした対話データの分析と音声画像特徴量の検討. 情報処理学会研究報告. SLP, 音声言語情報処理, **2015** (10), pp. 1–6. [Y. Chiba et al. (2015). Yuza no Taiwa Iyoku Jido Suitei wo Mokuhyo toshita Taiwa Deta no Bunseki to Onsei Gazo Tokuchoryo no Kento. *IPSJ SIG Technical Report*, 2015 (10), pp. 1–6.].
- Chollet, M., Prendinger, H., and Scherer, S. (2016). “Native vs. Non-Native Language Fluency Implications on Multimodal Interaction for Interpersonal Skills Training.” In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 386–393.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186.
- Hangyo, M., Kawahara, D., and Kurohashi, S. (2012). “Building a Diverse Document Leads Corpus Annotated with Semantic Relations.” In *Proceedings of 26th Pacific Asia Conference on Language Information and Computing*, pp. 535–544.
- Inaba, M. and Takahashi, K. (2018). “Estimating User Interest from Open-Domain Dialogue.” In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 32–40.
- Inoue, K., Lala, D., Takanashi, K., and Kawahara, T. (2018). “Engagement Recognition in Spoken Dialogue via Neural Network by Aggregating Different Annotators’ Models.” In *Proceedings of Interspeech 2018*, pp. 616–620.
- Ishihara, T., Nitta, K., Nagasawa, F., and Okada, S. (2018). “Estimating Interviewee’s Willingness in Multimodal Human Robot Interview Interaction.” In *Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct*, pp. 1–6.
- Kawahara, D., Kurohashi, S., and Hasida, K. (2002). “Construction of a Japanese Relevance-tagged Corpus.” In *Proceedings of the 3rd International Conference on Language Resources Evaluation*, pp. 2008–2013.

- 駒谷和範, 岡田将吾, 西本遥人, 荒木雅弘, 中野幹生 (2018). 配布可能なマルチモーダル対話データの収集とアノテーション不一致傾向の分析. *SIG-SLUD*, **5** (02), pp. 45–50. [K. Komatani et al. (2018). Collecting Sharable Multimodal Dialogue Data and Analysis of Annotation Disagreement Tendency. *SIG-SLUD*, **5** (02), pp. 45–50.].
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology (second edition)*. Sage Publications.
- Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., and Dolan, B. (2016). “A Persona-Based Neural Conversation Model.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 994–1003.
- Loshchilov, I. and Hutter, F. (2019). “Decoupled Weight Decay Regularization.” In *International Conference on Learning Representations*.
- 松本紗規子, 荒木雅弘 (2018). 雑談対話におけるマルチモーダル情報を統合した興味判定手法. *SIG-SLUD*, **B5** (02), pp. 86–87. [S. Matsumoto and M. Araki (2018). Assessment of User’s Interests in Multimodal Dialog Integrating Multiple Modalities. *SIG-SLUD*, **B5** (02), pp. 86–87.].
- 宮崎千明, 東中竜一郎, 牧野俊朗, 松尾義博 (2017). 言語的・対話的特徴に着目したコールセンター対話における話者の知識量推定. *情報処理学会論文誌*, **58** (2), pp. 594–604. [C. Miyazaki et al. (2017). Estimating Callers’ Levels of Knowledge in Call Center Dialogues by Using Lexical and Dialogic Features. *IPSJ Journal*, **58** (2), pp. 594–604.].
- 中野幹生, 駒谷和範, 船越孝太郎, 中野有紀子, 奥村学 (2015). 対話システム. 自然言語処理シリーズ 7号. コロナ社. [M. Nakano et al. (2015). *Taiwa Shisutemu*. Shizen Gengoshori Series 7. Corona-sha.].
- 西本遥人, 武田龍, 駒谷和範 (2018). マルチモーダル対話における興味の有無の推定と追加コーパスを用いた性能評価. *SIG-SLUD*, **B5** (02), pp. 72–73. [H. Nishimoto et al. (2018). Estimating Interest Level in Multimodal Dialogues and Evaluating the Performance with Newly Collected Data. *SIG-SLUD*, **B5** (02), pp. 72–73.].
- 齋藤純, 坂口智洋, 柴田知秀, 河原大輔, 黒橋禎夫 (2018). 述語項構造に基づく言語情報の基本単位のデザインと可視化. 言語処理学会第24回年次大会発表論文集. [J. Saito et al. (2018). *Jutsugo Ko Kozo ni Motozuku Gengo Joho no Kihon Tan’i no Dezain to Kashika*. Proceedings of the 24th Annual Meeting of the Association for Natural Language Processing.].
- Saito, J., Murawaki, Y., and Kurohashi, S. (2019). “Minimally Supervised Learning of Affective Events Using Discourse Relations.” In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 5758–5765.

- Smith, E. M., Williamson, M., Shuster, K., Weston, J., and Boureau, Y.-L. (2020). “Can You Put it All Together: Evaluating Conversational Agents’ Ability to Blend Skills.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2021–2030.
- Song, Z., Zheng, X., Liu, L., Xu, M., and Huang, X. (2019). “Generating Responses with a Specific Emotion in Dialog.” In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3685–3695.
- Stevens, S. S. (1946). “On the Theory of Scales of Measurement.” *Science*, **103** (2684), pp. 677–680.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M. (2000). “Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech.” *Computational Linguistics*, **26** (3), pp. 339–374.
- 徳久良子, 乾健太郎, 松本裕治 (2009). Web から獲得した感情生起要因コーパスに基づく感情推定. 情報処理学会論文誌, **50** (4), pp. 1365–1374. [R. Tokuhisa et al. (2009). Web kara Kakutoku shita Kanjo Seiki Yoin Kopasu ni Motozoku Kanjo Suitei. IPSJ Journal, 50 (4), pp. 1365–1374.].
- Tolmachev, A., Kawahara, D., and Kurohashi, S. (2020). “Design and Structure of The Juman++ Morphological Analyzer.” *Journal of Natural Language Processing*, **27** (1), pp. 89–132.
- 富増紗也華, 荒木雅弘 (2016). 雑談対話におけるマルチモーダル情報からの興味の有無の判定. 人工知能学会全国大会論文集, **JSAI2016**, 204OS23a1. [S. Tomimasu and M. Araki (2016). Multimodal Classification of Users’ Interests on Topics in Chat-oriented Dialog. Proceedings of the 30th Annual Conference of the Japanese Society for Artificial Intelligence, 204OS23a1.].
- 植田暢大, 河原大輔, 黒橋禎夫 (2020). BERT と Refinement ネットワークによる統合的照応・共参照解析. 言語処理学会第 26 回年次大会発表論文集. [N. Ueda et al. (2020). BERT to Refinement Nettowaku niyuru Togoteki Shoou · Kyosansho Kaiseki. Proceedings of the 26th Annual Meeting of the Association for Natural Language Processing.].
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). “Attention is All You Need.” In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 5998–6008.
- Yoshino, K. and Kawahara, T. (2015). “Conversational System for Information Navigation Based on POMDP with User Focus Tracking.” *Computer Speech & Language*, **34** (1), pp. 275–291.
- Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., and Weston, J. (2018). “Personalizing Dialogue Agents: I Have a Dog, Do You Have Pets Too?” In *Proceedings of the 56th Annual*

*Meeting of the Association for Computational Linguistics*, pp. 2204–2213.

Zhou, X. and Wang, W. Y. (2018). “MojiTalk: Generating Emotional Responses at Scale.” In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 1128–1137.

## 付録

### A アノテータ毎のアノテーションの信頼性

本節では話者内部状態アノテーション（4.3 節）におけるアノテータ毎のアノテーションの信頼性について述べる。アノテーションはクラウドソーシングによって行い、458 名のアノテータが参加した。アノテータ毎のアノテーションの信頼性は他のアノテータとのアノテーションの一致率  $m$  で測る。 $m$  は以下の計算式で計算される。

$$m = \frac{\text{他のアノテータとアノテーションが一致した数}}{\text{当該アノテータによるアノテーション数}}$$

全データ (=Full) での平均一致率  $\bar{m}$  は 0.51 であった。一致率がチャンスレート ( $m = 0.33$ ) を下回るアノテータをアノテーションの信頼性が低いアノテータとみなすとする、今回の話者内部状態のアノテーションに参加した 458 名中、該当するアノテータは 41 名であった。割合に直すと全体の約 9% であり、アノテーションの信頼性が低いアノテータはごく少数であったと言える。

実際、この 41 名を除外した場合の Krippendorff’s  $\alpha$  (Krippendorff 2004) によって計算されるアノテータ間の一致率は、知識が 0.45、興味が 0.45、対話意欲が 0.39 と全データにおけるアノテータ間の一致率（知識: 0.41, 興味: 0.40, 対話意欲: 0.35）（表 6 参照）と大きな差はなかった。

### B 映画推薦対話システムの改良

評価時の対話収集時に映画推薦対話システムの一部のモジュールを改良した。この節で述べる内容については評価時（5.2 節）にのみ適用し、コーパス構築時には適用していない。

#### 映画データベースの拡張

まず、最新の映画についても推薦できるよう映画データベースに格納する映画情報を 213 作品から 331 作品へと増やした。

#### 不適切な推薦ポイント文のフィルタリング

次に、推薦ポイント文抽出の改善を行った。まず、これまでの word2vec での手法によって抽

出された推薦ポイント文について分析した。その結果「面白かったのでまた来週観に行きます」といったような人間の発話としては適切だが、対話システムの立場として不適切な文があることが分かった。そこで以下の条件を満たすレビュー文を推薦ポイント文の候補から除外した。

**条件 1** : レビュー文中のいずれかの述語の主格が著者 (=レビューを書いたユーザ) である

**条件 2** : 条件 1 に該当する述語が意思性を持つ

**条件 3** : 条件 1 に該当する述語が対話システムの立場として使用するのが不適切である

条件 1 ではレビュー文を既存の解析器を用いて述語の主格を特定することで判定する。解析器には 植田 他 (2020) の *Base+coref+noun+bridge* モデルを京都大学ウェブ文書リードコーパス (KWDLIC) (Hangyo et al. 2012) と京都大学テキストコーパス (Kawahara et al. 2002) で学習したモデルを用いた。KWDLIC のテストデータに対して、異なる初期値で実験を 3 回行った外界ゼロ照応解析の平均 F1 スコアは 0.745 であった。条件 2 については述語の品詞、意味、接尾辞 (モダリティ・態)、項にもとづき識別する<sup>8</sup>。条件 3 は独自に選んだ全 53 単語からなる不適切述語リストに含まれているものとする。不適切述語リストは条件 1, 2 に該当する述語全 2,901 種類の中から出現回数 50 回以上のもの 103 個を取り出し、対話システムの立場として使用するのが不適切であるかどうかという指標で著者らで選定した。

#### レビュー文の極性判定手法の変更

レビュー文が映画を褒めているかどうかの極性判定の手法についても変更を行った。Saito et al. (2019) はイベント (齋藤 他 2018) 単位で極性を判定する手法を提案している。本研究ではこの極性判定器を利用し、レビュー文をイベント単位に分割し、文中の最も後ろのイベントの極性をそのレビュー文全体の極性とみなして極性判定を行う。極性のスコアは 1 から -1 までの範囲で出力される (1 が最もポジティブ、-1 が最もネガティブ)。この極性スコアと各文の文字数の積をスコアとし、推薦ポイント文の候補の中からスコア上位 50 文を抽出する。これを各映画について行い、各シナリオでは 50 文の中からランダムに 2 文を用いる。

#### 略歴

**児玉 貴志** : 2019 年京都大学工学部電気電子工学科卒業。2020 年京都大学大学院情報学研究科博士前期課程修了。同年より同大学院情報学研究科博士後期課程に進学、現在に至る。修士 (情報学)。自然言語処理の研究に従事。言語処理学会会員。

**田中リベカ** : 2013 年お茶の水女子大学理学部情報科学科卒業。2015 年同大学院博士前期課程修了。2018 年同博士後期課程単位修得退学。同年より京都大学大学院情報学研究科特定研究員。

<sup>8</sup> <https://github.com/ku-nlp/ishi>

**黒橋 禎夫**：1994年京都大学大学院工学研究科電気工学第二専攻博士課程修了。博士（工学）。2006年4月より京都大学大学院情報学研究科教授。自然言語処理，知識情報処理の研究に従事。言語処理学会10周年記念論文賞，同20周年記念論文賞，第8回船井情報科学振興賞，2009 IBM Faculty Award 等を受賞。2014年より日本学術会議連携会員。

(2020年7月1日 受付)

(2020年10月1日 再受付)

(2020年11月7日 採録)