

# 間接的課題導入を行う映画推薦対話システムの構築と 対話事例の収集・分析

児玉 貴志<sup>†</sup> 田中 リベカ<sup>‡</sup> 黒橋 禎夫<sup>†‡§</sup>

<sup>†</sup> 京都大学 <sup>‡</sup> 科学技術振興機構 CREST <sup>§</sup> 国立情報学研究所 CRIS

{kodama, tanaka, kuro}@nlp.ist.i.kyoto-u.ac.jp

## 1 はじめに

人間と機械の新たなインタフェースとして、知的な対話システムの実現が期待されている。Twitter などの大規模データを利用して、一回きりのレスポンスを適切に生成するオープンドメインの雑談研究なども盛んに行われている。しかし、たとえ雑談であっても、対話には一定の目的があり、複数回のターンテイキングによって情報がやりとりされる。人間と自然に対話できるシステムを実現するためには、このような複数発話の文脈を理解し、また対話参加者の心理状態、すなわち、興味・関心の度合いややりとりされている情報の理解度などを適切にモデル化する必要がある。このような対話研究を進めるためには、対話データの構築と分析が必要だが、人間同士の対話はレベルが高すぎて初期の対話システムの構築・改善のためには距離があり、一方、深層学習で生成される一回のレスポンスの連鎖を雑談対話とみなしても、現在のところその質は高くなく、文脈の分析に適しているとは言えない。

このような問題意識から、本研究では、適切な問題設定で、ある程度のターンテイキングが発生する対話システムを構築し、その対話システムと人との対話データを大量に収集・分析することで、文脈や人の心理状態をより深くとらえる対話研究をスタートさせたいというモチベーションで研究を行った。

具体的には、160 作品の映画データベースを構築し、これに基づく映画推薦システムを構築した。このシステムでは、(少し) ユーザの好みを聞いた上で、「魔法に興味はありますか?」「最近〇〇が話題ですが...」のように間接的に話題(推薦する映画)の導入を行う(図 1)。このような説得対話(persuasive dialogue)、すなわち、いきなり本題に入るのではなく、ビジネスマンが遠回しに商品売り込むような対話技術は将来の対話システムにとっても重要であろう。しかしそれ以上に本研究で重視したのは、推薦対話によってシステム側がイニシアティブをとれること、また、間接的課題導入によって必然的に一定回数以上のターンテイ

S1: 邦画と洋画どちらが好きですか?

U1: 洋画です

S2: そうですね。悪役って興味ありますか?

U2: わりと興味があります

S3: その悪役に関係のある映画があるんです

U3: どのような映画ですか?

S4: 「スーサイド・スクワッド」という名前の映画です

U4: 怖そうなタイトルですね

S5: 映画のほとんどの場面は夜ですが、登場人物がド派手なのでかなり明るい映画だと評判ですよ

U5: 目が悪いのでよく見えるか心配です

S6: 好き好きの別れやすい映画だな...、と言うこと  
だけは言えますよ 矛盾

U6: 見てみないとわからないですね

S7: 面白い作品なのでぜひ見て頂きたいです

図 1: 提案システムの対話例 (S はシステム発話、U はユーザ発話)

キングが生じる点である。これによって、現在の対話システムの技術レベルでも、人とシステムとの意味をなす対話ログが収集できる設定となる。実際、図 1 でのシステム発話 S6 は文脈上不適切であり、このような一定の文脈がある対話の中での不適切発話データの収集が可能となった。

本稿では、映画推薦システムのデザインと、クラウドソーシングで収集した本システムの対話ログの初期的分析を報告する。

## 2 関連研究

人と対話システムのテキスト対話の収集・分析は、タスク指向型・非タスク指向型対話システムの両方の文脈で行われてきており、エラーの類型も複数提案されている。対話原理の観点からエラーの類型化を行った古典的研究として、Dybkjaer et al. [1] の研究がある。人と航空券予約システムとの対話について、協調的な対話の原理を挙げた Grice の公準を参照しつつ、さらに人とシステムの対話に特有の観点を追加した分類を提案している。最近では東中ら [2] が雑談対話について

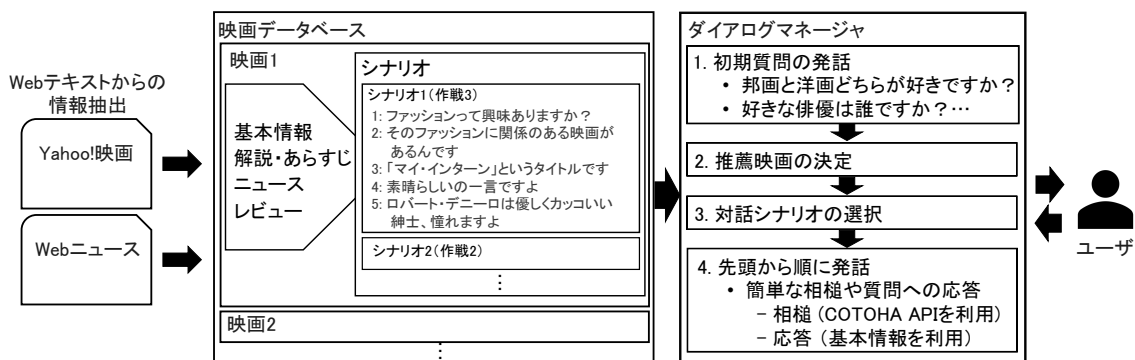


図 2: システム構成図

て Grice の公準を参照したエラー分類体系を提案し、対話の表層に注目した汎用的な分析を行っているが、その元となっている雑談対話は文脈の分析に十分な質だとは言えない。

推薦や交渉を含む説得対話を行うシステムも数多く提案されており、対話の設定も様々である。石川ら [3] は、依頼に対する受託度合いをユーザ発話から推定し、感情表現を用いて説得を行うシステムを提案している。また、He et al.[4] は、交渉戦略の決定と発話文生成の処理を切り分け、より適切な対話管理を実現した値段交渉システムを提案している。本研究では対話データの収集・分析に主眼を置くことから、間接的話題導入を行う対話システムを独自に構築した。

### 3 映画推薦対話システムの構成

システムの構成を図 2 に示す。提案システムでは、システム発話の内容はあらかじめ映画ごとに作成しておく。この発話集をシナリオと呼ぶ。対話管理はダイアログマネージャが担い、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。

#### 3.1 基本的な対話の戦略

本システムの基本的な対話戦略は、第一に、対話の始めに間接的に話題を導入することである。本研究では、間接的話題導入の方法として以下の 3 つのパターンを用意した。

パターン 1 : 最近のニュースから対話を始める

パターン 2 : 映画のテーマから対話を始める

パターン 3 : 映画情報 (人名) から対話を始める

第二に、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をすることである。

第三に、必要最低限の相槌や応答をする以外、ユーザ発話に関わらずシナリオ通りに発話を行うことである。この素朴な設定も、システムがイニシアティブを

とる推薦対話では十分機能し、現在の技術レベルでも意味のある対話が可能となる。

#### 3.2 映画データベース

各映画の基本情報、解説・あらすじ、ニュース情報、レビュー、シナリオは映画データベースに格納する。現在、Yahoo!映画のユーザレビュー数が多い映画上位 160 作品を対象としている。

基本情報、解説・あらすじ、レビューは、Yahoo!映画から 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り、解説・あらすじは数文からなる。レビューは、Yahoo!映画のユーザが映画を見た感想や評価などを書き込んだものであり、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 dmenu ニュースの人気記事ランキングから Web スクレイピングで取得した。社会、経済などの 9 つのドメインについて、1 日あたり 20 件ずつ、直近 5 日分の記事がランキング形式で掲載されている。このうちエンタメとスポーツのニュース記事 5 日分 (計 200 件) を取得した。このニュース情報は間接的話題導入パターン 1 のシナリオの作成に用いる。

#### 3.3 対話シナリオの作成

シナリオは対話戦略に沿って映画ごとに 1 つ以上作成した。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。後半部分では、推薦ポイントを述べる発話を 2 回行い、最後に視聴を促した。

シナリオの前半部分に用いる間接的話題導入の発話は、3.1 節で述べた 3 つの導入パターンに基づき、映画データベース中の情報を用いて生成する。

パターン 1 では、最近のニュースの話題から対話を始め、そのニュースの登場人物が出演する映画を推薦する。映画データベースのニュース記事から最初の一

文を抽出し、キャストの名前が含まれていたなら、その文に「～というニュースが話題になっていますね」などをつなげる。

パターン2では、映画のテーマとなる普通名詞から対話を始める。その映画を代表する名詞を、映画データベースの解説・あらすじから tf-idf 法を用いて選ぶ。まず、解説・あらすじの文章から全ての名詞を抽出したあとキャスト・スタッフの名前を取り除く。それぞれの映画につき tf-idf のスコアがもっとも高い名詞をその映画のテーマとする。そのテーマを用いて「～って興味ありますか?」といった文で対話を始める。例えば、テーマが「魔法」ならば「魔法って興味ありますか?」と導入する。

パターン3では、映画データベースの基本情報に含まれている人物名などの固有名詞から対話を始め、その固有名詞に関連する映画を推薦する。今回は、各映画の主要キャスト2名と監督を対象とした。その人物名を用いて「～さんってご存知ですか?」といった発話から対話を始める。

シナリオの後半部分に用いる映画の推薦ポイントを述べる発話には、Yahoo!映画のユーザーレビューの文のうち、映画を褒めている文を用いた。それぞれの映画について、レビューの各文に褒めている度合いのスコアをつけて上位100文を抽出し、シナリオにはランダムに2文を使用した。具体的には、まず、映画を褒めるときに使われる名詞(「傑作」など)・形容詞(「面白い」など)計50語を選んだ。その後、単語ベクトルの平均とその50語の単語ベクトルの平均のコサイン類似度をスコアとした。なお、単語ベクトルの計算には学習済み word2vec を用いた。

### 3.4 ダイアログマネージャ

ダイアログマネージャでは、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。

まず推薦映画を決定する。推薦映画をランダムに決定する方式と、ユーザーの好みを尋ねてから、好みに応じた映画を推薦する方式を用意し、これらの方式のうちどちらかを無作為に選択する。この質問を初期質問と呼び、以下からランダムに1つを発話する。

- 好きな俳優は誰ですか?
- 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、例えばユーザー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。そして、「～さんいいですね」「その～さんが出演している映画を紹介しますね」といった形で話題導入を行う。ユーザーが「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した後、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。

ユーザー発話に関わらずシナリオ通りの発話を行うため、システム発話とユーザー発話にズレが生じる可能性がある。このズレを吸収するため、ユーザーの発話の種類に応じて、システム発話の先頭に相槌を挟んだ。ユーザー発話の種類判定には NTT コミュニケーションズが提供する COTOHA API を用いた。COTOHA API では文のタイプ(平叙文、疑問文、命令)と発話行為(挨拶、情報提供など14種類)の判定が可能である。今回は、ユーザー発話が[文タイプ: 疑問文, 発話行為: 情報獲得以外]なら「ちょっとわかりませんが、」を、[文タイプ: 平叙文, 発話行為: フィードバック/相槌]なら「そうですか...、」を挿入した。なお、推薦映画の主演・監督・キャストの名前を問う簡単な質問をユーザーがした場合には、「ちょっとわかりませんが」を挿入する代わりに映画情報データベースを参照して応答を挿入し、内容を問う質問があった場合には「内容については詳しく言えないのですが、」を挿入した。

## 4 対話例の収集と分析

### 4.1 クラウドソーシングでの対話収集

本システムを用いてクラウドソーシングで対話例を収集した。ワーカーには事前に、システムとの対話であること、そのシステムが160作品の中から対話的に映画を薦め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た。最後はシステム発話で対話を終了する。また、対話後には以下に示す5段階評価選択式のアンケートを3問実施した。

- (1) 潜在的興味: 映画は好きですか?  
{5:好き, 4:どちらかと言えば好き, 3:どちらとも言えない, 2:どちらかと言えば嫌い, 1:嫌い}
- (2) 説得度: 推薦された映画を見たくありませんか?  
{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}
- (3) 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}

S1: 邦画と洋画どちらが好きですか？  
 U1: 邦画  
 S2: そうですね。宮崎吾朗さんをご存知ですか？  
 U2: いいえ知りません  
 S3: そうですね...、その宮崎吾朗さんが監督の映画を紹介したいんですが  
 U3: お願いします  
 S4: 「ゲド戦記」という名前の映画です  
 U4: 聞いたことがあります  
 S5: 二度、三度と鑑賞して、初めて深い感動を味わえるらしいですよ  
 U5: そうなんですか。いちど見てみたいです。  
 S6: 素直に感動しますよ  
 U6: 見てみたいです  
 S7: 面白い作品なのでぜひ見て頂きたいです

図 3: 収集した対話例

このうち (2) の説得度は、推薦というシステムの目標が達成されたかどうかに加えて、対話から読み取れる興味・関心状態と照合する目的で調査した。さらに、推薦という目標の達成しやすさは、今回の話題である映画が好きか否かにも依存すると考えられるため、(1) の潜在的興味でワーカーの傾向を調査した。

## 4.2 考察

収集した対話例を図 3 に示す。全部で 80 対話収集し、その平均ターン数は約 6 ターンであった。特に多く出現したユーザ発話は「はい」と「そうですね」で、出現回数はそれぞれ 35 と 24 であった。一方、ユーザ発話全体の 56.2% が出現回数 1 の発話だったことから、ある程度のバリエーションが得られたと考えられる。

対話後のアンケートの結果を表 1 に示す。元々映画が好きな参加者が多数だったことも手伝ってか、57 の対話で説得度が 4 以上であった。また、60 の対話で自然さについて 4 以上の評価を得た。説得度が 2 以下の対話では、ユーザ発話に単調な発話やシステムの発話に対しての否定的な発話が多く見受けられた。

対話中の個々のシステム発話について、エラーの調査を行った。エラーの分類にあたっては、東中ら [2] の類型化にしたいが、発話・応答・文脈の 3 つのレベルを区別した。その上で、発話を最も下位のレベル、文脈を最も上位のレベルと位置づけ、下位のレベルでのエラーを優先した。今回の調査で出現したエラーの分類ラベルと分布を表 2 に示す。東中ら [2] が提示したエラータイプのうち、不適切発話や話題転換への不追随などを除く一部のみが観察された。たとえば、冒頭の図 1 の対話の S6 は推薦の意図から外れる発話であり、文脈レベルの「矛盾」に該当する。これは推薦ポイン

評価	潜在的興味	説得度	自然さ
5	37 (46.3%)	21 (26.3%)	29 (36.3%)
4	34 (42.5%)	36 (45.0%)	31 (38.8%)
3	10 (12.5%)	14 (17.5%)	12 (15.0%)
2	2 (2.5%)	9 (11.3%)	9 (11.3%)
1	2 (2.5%)	5 (6.3%)	4 (5.0%)

表 1: アンケート評価と回答数 (括弧内は割合)

レベル	分類ラベル	発話数
発話	構文的誤り	10
	意味的誤り	9
応答	不理解	21
	無関係	35
文脈	不要情報	1
	矛盾	26
	関連性不明	25

表 2: エラー分類ラベルとシステム発話数の分布

トの抽出が上手くいかなかったことが原因であり、今後システムを改良する必要がある。

さらに、破綻とまでは言えないが情報の流れに違和感がある例は少なくなかった。たとえば、図 3 の対話の自然さの評価は 5 であったが、より注意深く見ると、S5 と S6 をこの順に発話しているのは不自然である。システム発話への評価の主観性の高さは対話システム評価の問題の 1 つであり [5]、精査が必要である。

## 5 おわりに

本研究では、映画推薦対話システムをデザインし、クラウドソーシングでの対話収集とその分析を行った。間接的話題導入を行う推薦対話という設定によって、一定の文脈を維持した対話データが得られた。今後、情報源の取り扱いに注意した上で大規模データの収集・公開を進めていく予定である。

## 参考文献

- [1] Laila Dybkjær, Niels Ole Bernsen, and Hans Dybkjær. GRICE INCORPORATED: Cooperativity in spoken dialogue. In *COLING1996*, Vol. 1, pp. 328–333, 1996.
- [2] 東中竜一郎, 船越孝太郎, 荒木雅弘, 塚原裕史, 小林優佳, 水上雅博. テキストチャットを用いた雑談対話コーパスの構築と対話破綻の分析. *自然言語処理*, Vol. 23, No. 1, pp. 59–86, 2016.
- [3] 石川葉子, 水上雅博, 吉野幸一郎, Sakti Sakriani, 鈴木優, 中村哲. 感情表現を用いた説得対話システム. *人工知能学会論文誌*, Vol. 33, No. 1, pp. DSH-B.1-9, 2018.
- [4] He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. Decoupling strategy and generation in negotiation dialogues. In *EMNLP2018*, pp. 2333–2343, 2018.
- [5] 東中竜一郎. 雑談対話システムの評価とその問題点 (特集「人工知能・機械学習技術の他分野への応用」および一般). *人工知能基本問題研究会*, Vol. 101, pp. 21–23, 2016.