外部知識に基づく発話生成に向けた日本語映画推薦対話データセットの構築

児玉 貴志 † 田中 リベカ † 黒橋 禎夫 †‡
† 京都大学 ‡ 科学技術振興機構 CREST
{kodama,tanaka,kuro}@nlp.ist.i.kyoto-u.ac.jp

1 はじめに

人と自然に話すことができる対話システムは人工知能の究極の目標の一つと言われている．対話はタスク指向型対話 [1] と非タスク指向型対話（雑談対話）[2] に大別される [3] が，その両者において，対話外の知識（外部知識）を必要とする場合がほとんどであり [4]，その理解・活用は重要視されている．

そうした背景もあり，外部知識を利用した発話生成の研究は近年注目を集めており，ベンチマークとなるデータセットも数多く提案されている [5, 6, 2, 4, 7]．しかしこれらのデータセットは英語または中国語での対話であり，外部知識に基づく発話生成を目的とした日本語対話データセットは存在しない．また，外部知識に基づいた，英語の対話データセットである Wizard of Wikipedia [2] を日本語に翻訳して活用を試みた研究もあるが，翻訳の誤りによるエラー伝搬が起きると考えられるのに加え，日本人の対話に出てこない話題があることが報告されており [8]，対話における言語間の壁は大きい．

そこで本研究では，外部知識に基づく発話生成を目的とした日本語映画推薦対話データセットを構築した．この対話データセットはクラウドワーカーによって作成され，約 2,500 対話からなる．図 1 に示すように，1 人の話者が映画の推薦者，もう 1 人の話者が被推薦者となり，推薦者が被推薦者に 1 つの映画を勧める設定である．推薦者だけが映画についての知識（＝外部知識）を参照することができ，推薦者はこの外部知識をできるだけ使用して発話をする．発話を行う際には，推薦者自ら，使用した知識を選んでアノテーションを行う．この手続きにより，対話中の推薦者側の発話すべてに，その発話の構成に使用された外部知識を紐付けることができる．また，外部知識である映画についての知識として，タイトル，製作年度などの構造化された知識に加え，あらすじやレビューといった構造化されていないテキストも用意することで，様々なタイプの外部知識を活用することを目的としたデータセットになっている．

本稿では，外部知識の収集方法とその知識に基づいた対話の収集方法を説明したのち，構築した映画推薦対話データセットを利用して訓練したベースラインモデルについて報告する．

2 映画推薦対話データセットの構築

対話のドメインには万人が興味を持ちやすく対話が円滑に進みやすいテーマであると考えられることから映画を選んだ．本節では映画推薦対話データセットの構築手法について説明する．

2.1 外部知識の収集

外部知識である映画情報は Wikipedia などのウェブテキストを中心に収集する．まず，過去の興行収入ランキング [1] を参考に 242 本の映画を選定した．この各映画に対して外部知識となる映画情報を収集する．

外部知識は図 1 にも示すように基本情報，レビュー，あらすじからなる．このうち，基本情報の大部分（タイトル，製作年度，監督名前，キャスト$_1$名前，キャスト$_2$名前）及びあらすじは各映画の Wikipedia 記事から取得する（監督は最大 1 名，キャストは最大 2 名まで）．また監督説明，キャスト$_1$説明，キャスト$_2$説明 はそれぞれの人物の Wikipedia 記事の第一段落から取得する．ジャンルについては Yahoo!映画 [2] のジャンル分けを使用する．レビューはクラウドソーシングで収集する（Yahoo!クラウドソーシング [3] を利用）．各ワーカは全 242 本からなる映画リストの中から自分が観たことがある映画を

1） http://www.eiren.org/toukei/index.html
2） https://movies.yahoo.co.jp/
3） https://crowdsourcing.yahoo.co.jp/

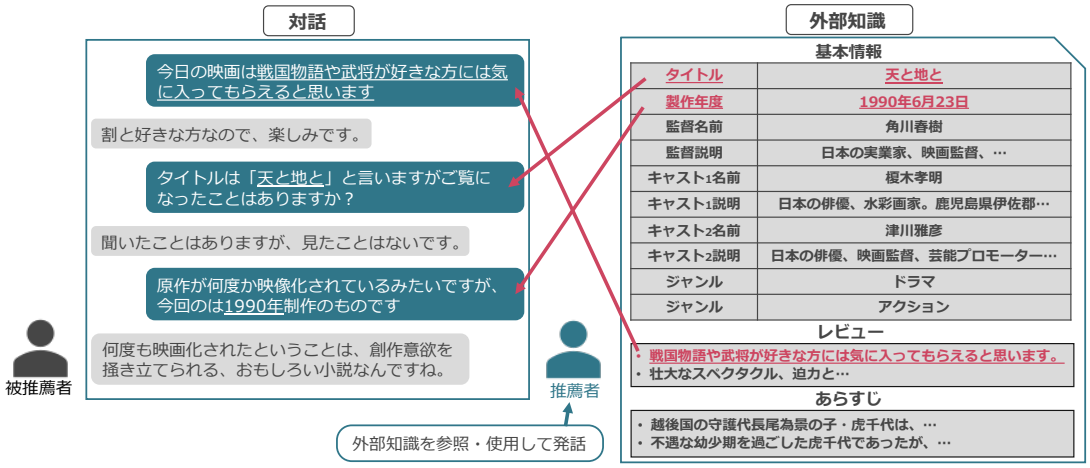

図 1　構築した映画推薦対話データセットの例．外部知識中の下線部は対話で使用されている知識であることを示す．

選び，その映画の「おすすめポイント」を 3 点に分けて文章で記入する．結果として 1 映画あたり平均 16.5 件のレビューを収集した．

推薦者に提示する知識が多くなりすぎないようにするため，あらすじは冒頭の 10 文のみ（10 文に満たない場合は全ての文）を文分割して提示する．レビューはワーカが記入したレビューを文分割等は行わずにそのまま使用する．収集したレビューの中から 15 文字以上 80 文字未満のレビューを各映画につき 5 つずつランダムに選び，その 5 つをその映画のレビューとして常に使用する．

2.2　クラウドソーシングによる対話収集

2.2.1　設定

2 人のワーカは役割が異なり，片方は映画の推薦者，もう片方は被推薦者として対話を行う．

推薦者　被推薦者が映画を見たくなるように映画を勧める．推薦する映画は映画リストの中から推薦者側が自由に決められる．この際，自分が勧めたい映画を選んでも，相手の好みを対話で聞き出しながらその好みにあった映画を選んでも構わない．推薦者は提示された知識をできる限り使用して映画を推薦し，発話を送信する際にはその発話の構成に使用した知識のチェックボックスにチェックを入れてから送信する（複数選択可）．また，挨拶や相槌など知識を使用しない発話の場合は別途用意してある「知識なし」の選択肢を選ぶことができる．

被推薦者　Wizard of Wikipedia [2] の指示と同様に「より楽しく，勧められる映画のことを知ろうと心がけて下さい」とだけ指示されている．

対話は以下の流れで行われる．

1. 推薦者・被推薦者のどちらから対話を始めてもよい．
2. 推薦者が推薦する映画を決める．映画が決まると，推薦者側の画面にはその映画についての知識が表示される．一方，被推薦者側の画面にはチャット画面のみが表示され，映画についての知識は表示されない．
3. 推薦者は提示された知識の中から選んだ知識に基づいた発話を行い，被推薦者は推薦者の発話に対して自由に返答する．
4. 対話は映画が決定されてから最低 20 ターン続ける．20 ターンを超えると対話を終了することができる．

2.2.2　対話収集システムの構築

対話収集システムは，児玉らの対話収集基盤 [9] をベースとして構築する．この対話収集基盤では指定した対話収集用 URL にワーカがアクセスするとペアのマッチングが行われ，ワーカがリアルタイムで対話するチャットルームの生成，およびそこでの対話の保存までを自動で行うことができる．本研究で使用する対話収集システムでは，ペアとなったワーカ 2 人にそれぞれ推薦者，被推薦者の役割を与え，推薦者のみ対話中に外部知識を参照できるような機構のみを追加で実装した．付録 A.1 に対話収集システムのインターフェースを示す．

2.2.3　収集結果

上述の設定，収集システムを利用して対話収集を行った．統計情報を表 1 に示す．全部で 2,565 対話

表 1　映画推薦対話データセットの統計情報．単語分割には Juman++ [10] を使用．

対話数	2,565
発話数	59,549
映画数	242
平均ターン数	23.2
1 発話あたりの平均単語数（推薦者）	23.9
1 発話あたりの平均単語数（被推薦者）	9.7
1 発話あたりの平均知識使用数	1.3
1 対話あたりの平均知識使用数	9.1

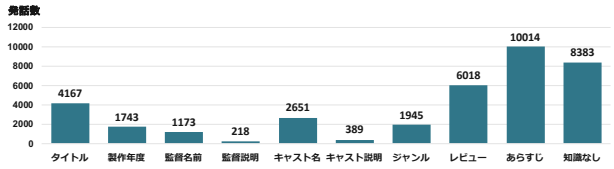

図 2　使用された外部知識の分布．キャスト$_1$名前 とキャスト$_2$名前 は「キャスト名」に，キャスト$_1$説明 とキャスト$_2$説明 は「キャスト説明」に集約して表示．

を収集した．推薦者側の 1 発話あたりの平均単語数は被推薦者側の 2 倍以上になっているが，これは推薦者側が映画を推薦するために，情報を提供するなどより多くのことを話す必要があるためと考えられる．また，1 発話につき知識を複数選択することも許容していたが，平均では 1.3 個となっており，各知識ごとに発話を分けて言及している傾向が見られた．1 対話あたりでは 9.1 個の異なる知識が使用されており，様々なタイプの外部知識が使用された対話が収集できていることが分かる．

図 2 に使用された外部知識のタイプ別の分布を示す．知識を使用していない発話は全体の約 2 割ほどに留まっており，多くの発話で何らかの外部知識が使用されている．また，レビューやあらすじといった構造化されていないテキストの使用頻度が高い傾向にあった．

さらに，今回収集した対話の一部（834 対話）については対話終了後に以下に示す 5 段階評価のアンケート（5 が最高評価，1 が最低評価）を実施した．

Q1: 映画が好きである
Q2: 対話を楽しめた
Q3: 推薦した（された）映画を知っているか
Q4: 上手く映画を推薦できた
Q5: 推薦された映画を見たくなった

Q1，Q2，Q4，Q5 の選択肢は [そう思う/ややそう思う/どちらとも言えない/ややそう思わない/そう思わない]，Q3 の選択肢は [観たことがあり，内容をよく覚えている/観たことがあり，内容を少し覚えている/観たことはないが，あらすじぐらいは知っている/観たことはないが，タイトルだけ知っている/全く知らない] である．推薦者は Q1，Q2，Q3，Q4 の 4 問に，被推薦者は Q1，Q2，Q3，Q5 の 4 問に回答する．

表 2　アンケート結果

	Q1	Q2	Q3	Q4	Q5
推薦者	4.38	3.97	4.01	3.90	-
被推薦者	4.27	3.78	2.63	-	3.78

アンケート結果を表 2 に示す．Q1 より多くのワーカが映画というテーマに対して高い興味を持っていることが，Q2 より最低 20 ターンという比較的長めの対話ながら，推薦者・被推薦者ともに対話を楽しんでいることが分かった．また，Q3 より推薦者は自分が観たことがある映画を推薦する映画として選ぶ傾向が見られた．最後に Q4，Q5 より，収集した対話が映画推薦の目的を十分に達成していることを確認できた．

3　実験

構築した映画推薦対話データセットでベースラインモデルを訓練した．本節ではベースラインモデルの外部知識の使用の有無による出力の違いを評価する．

3.1　モデル

ベースラインモデルには，Wizard of Wikipedia の論文でベースラインとして提案された生成ベースのモデルである，Generative End-to-End Transformer Memory Network [2] を用いる．このモデルは，対話履歴と外部知識の候補文を別々にエンコードし，対話履歴と各外部知識の候補のアテンションを計算して使用する外部知識を選択する．そして対話履歴と選んだ知識を連結してデコーダに渡すことで返答を生成する．このモデルの実装は ParlAI [4] で公開されているが，本研究では我々で再実装したものを用いる．

3.2　実験設定

データセットは学習データ：開発データ：テストデータ＝ 90%：5%：5%の割合で分割する．対話および外部知識は全て Juman++ [10] を使用して形態

4）https://github.com/facebookresearch/ParlAI

表 3　生成した応答に対する自動評価

	F1 (↑)	BLEU-1 (↑)	BLEU-2 (↑)	BLEU-3 (↑)	BLEU-4 (↑)
知識なし応答（予測知識）	6.7	5.5	0.0	0.0	0.0
知識あり応答（予測知識）	21.0	14.6	7.6	4.6	3.3
知識なし応答（正解知識）	5.5	4.8	0.0	0.0	0.0
知識あり応答（正解知識）	27.9	19.5	12.1	8.3	6.4

素に分割し，BPE [11] を適用する [5]．モデルのハイパーパラメータは基本的に Dinan ら [2] の設定に従っているが，語彙数のみ 10,000 に変更した．

対話履歴はその対話中の全ての過去の発話を新しい方から順に最大 64 トークンまで入力する．外部知識の候補はその対話中で推薦している映画についての知識全てとし，「知識なし」も候補に含める．この候補数は各映画によって差があるが，平均 24.9 個であった．Generative End-to-End Transformer Memory Network では使用する知識を 1 つだけ選んでいる．これに合わせ，知識が複数アノテーションされている発話については以下の式で表されるトークン重なり率を計算し，値が最も高い知識 1 つのみを正解の知識として用いる．

$$\text{トークン重なり率} = \frac{\text{発話と当該知識の共通トークン数}}{\text{当該知識のトークン数}}$$

外部知識の各候補は，「知識ラベル <KNOWLEDGE> 知識内容」の形式でモデルに入力する（<KNOWLEDGE>は特殊トークン）．知識ラベルはその知識の種類を表したもので，監督名前・監督説明は「監督」，キャスト$_1$名前・キャスト$_1$説明・キャスト$_2$名前・キャスト$_2$説明 は「キャスト」とし，タイトル・製作年度・ジャンル・レビュー・あらすじについてはそのまま使用する（ただし，トークナイズは行う）．例えば，図 1 の例の製作年度の場合，「製作年度 <KNOWLEDGE> １９９０年６月２３日」となる．「知識なし」の場合は知識ラベル，知識内容の部分をともに特殊トークン<NO_KNOWLEDGE>に置き換えたものを入力とする．

3.3　結果

知識予測の正解率は 28.2%であり，ランダムベースラインである 4%を大きく上回る結果となった．

表 3 に生成した応答の自動評価の結果を示す．「知識なし応答」は生成時に使用された知識が「知識なし」だった応答，「知識あり応答」は「知識なし」以外の具体的な外部知識を使用して生成された応答である．また，「(予測知識)」はモデルが選択した知識を，「(正解知識)」は正解の知識を使用して生成した結果である．評価には unigram F1 (F1) [2], BLEU-1/2/3/4 [12] を使用した．予測知識と正解知識のどちらを使用した場合でも「知識あり応答」の方が「知識なし応答」の場合を全ての評価指標で大きく上回っており，具体的な外部知識を使用した発話では，より実際の発話と近い発話ができていることが分かる．実際「製作年度」を知識として使用した時には「製作年度は 2008 年です」といった発話が生成されるなど，使用した知識を発話に反映できているケースが多かった．一方で「監督名前」が「ピエール・コフィン」である知識を与えたときに「監督は、ジェームズ・キャメロンです。」というような知識とは異なる発話が生成されるケースも散見された．

一方，「知識なし応答」はその殆どが「こんにちは」などのつまらない応答になっていた．これは学習データ数の不足が主な原因として考えられる．

4　おわりに

本研究では，外部知識に基づく発話生成を目的とした日本語映画推薦対話データセットを構築した．本研究は我々の知る限り，外部知識に基づいた日本語対話データセットを構築した初めての研究であり，日本語による，外部知識に基づく発話生成の研究に取り組むための足がかりとなることが期待される．今後はデータセットの公開に向けて準備を進めるとともに，事前学習モデルを導入し，ベースラインモデルの精度向上に努める予定である．

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 共同研究の助成及び科学技術振興機構 CREST「知識と推論に基づいて言語で説明できる AI システム」の支援のもとで行われた．

5)　https://github.com/rsennrich/subword-nmt

参考文献

[1] Antoine Bordes, Y-Lan Boureau, and Jason Weston. Learning end-to-end goal-oriented dialog. *arXiv preprint arXiv:1605.07683*, 2016.

[2] Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. Wizard of Wikipedia: Knowledge-powered conversational agents. In *International Conference on Learning Representations*, 2019.

[3] 中野幹生, 駒谷和範, 船越孝太郎. 対話システム. 自然言語処理シリーズ. コロナ社, 2015.

[4] Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. Proactive human-machine conversation with explicit conversation goal. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3794–3804, 2019.

[5] Kangyan Zhou, Shrimai Prabhumoye, and Alan W Black. A dataset for document grounded conversations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 708–713, 2018.

[6] Nikita Moghe, Siddhartha Arora, Suman Banerjee, and Mitesh M. Khapra. Towards exploiting background knowledge for building conversation systems. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2322–2332, 2018.

[7] Hao Zhou, Chujie Zheng, Kaili Huang, Minlie Huang, and Xiaoyan Zhu. KdConv: A Chinese multi-domain dialogue dataset towards multi-turn knowledge-driven conversation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7098–7108, 2020.

[8] 杉山弘晃, 成松宏美, 水上雅博, 有本庸浩, 千葉祐弥, 目黒豊美, 中嶋秀治. Transformer encoder-decoder モデルによる趣味雑談システムの構築. *SIG-SLUD*, No. 02, pp. 104–109, 2020.

[9] 児玉貴志, Frederic Bergeron, 新隼人, 田中リベカ, 坂田亘, 黒橋禎夫. クラウドソーシングで利用可能な日本語対話収集基盤. 言語処理学会第 27 回年次大会, 2021.

[10] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Design and structure of the Juman++ morphological analyzer toolkit. *Journal of Natural Language Processing*, Vol. 27, No. 1, pp. 89–132, 2020.

[11] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1715–1725, 2016.

[12] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.

A　付録

A.1　対話収集システムのインターフェース

図 3, 4 にそれぞれ映画決定前後の推薦者側のチャットルームのスクリーンショットを示す．映画決定前は画面左部に映画を決定するための簡単な検索機能が表示されており，推薦者はこの検索機能を利用して映画を決定する．映画決定後は画面左部に決定した映画についての知識が表示される．発話を送信するときには各知識の先頭にあるチェックボックスにチェックを入れてから送信する．なお，被推薦者側の画面には推薦者側の画面左部は表示されず，画面右部のチャット画面のみ表示される．

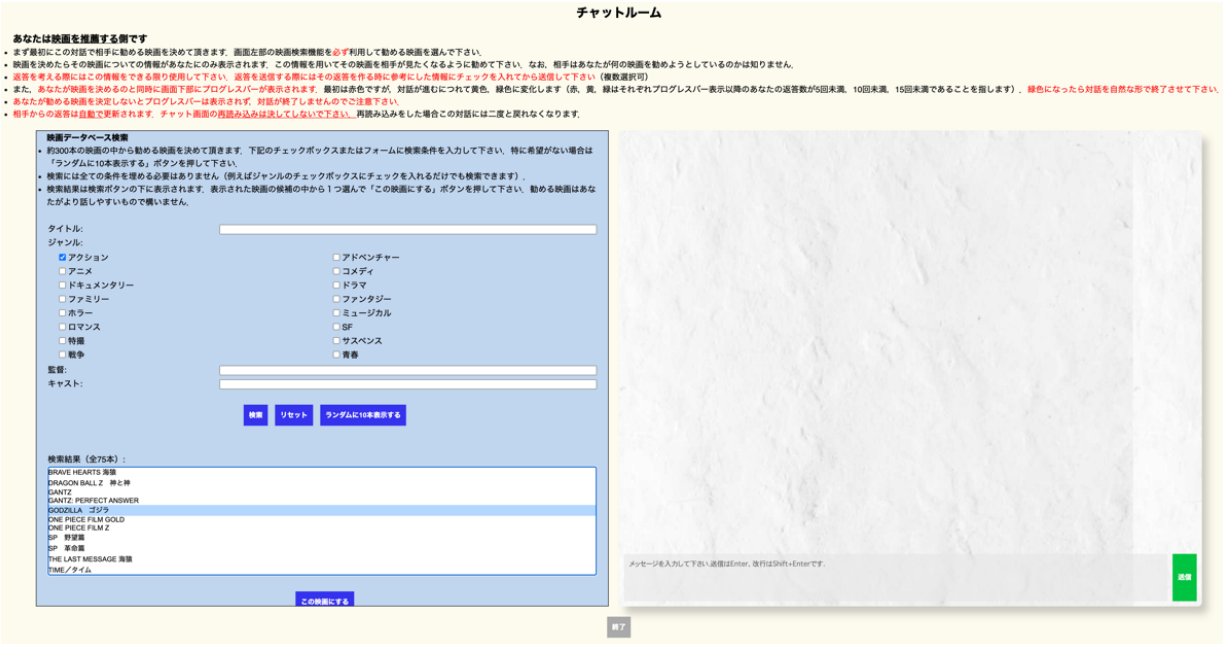

図 3　映画決定前のチャットルームのスクリーンショット（推薦者側）

図 4　映画決定後のチャットルームのスクリーンショット（推薦者側）