

特定ドメインの対話システムにおける 発話矛盾検出のための前提文生成

堀口 勇輝[†]山本 賢太[†]武田 龍[†]駒谷 和範[†][†] 大阪大学産業科学研究所

1. はじめに

対話システムの担う役割の一つとして、動物園ガイドなどの情報案内が挙げられる。近年、大規模言語モデル(LLM)による発話文生成を利用した対話システムは流暢な発話が可能であり、利用者に違和感を与えることなく対話を行える。一方で、LLMに基づく発話文生成は一般的な知識に基づいているため、発話文の内容が特定ドメインに特有の情報と矛盾する場合がある。これを避けるため、発話文がドメイン知識と矛盾していないかを判定することが重要である。

先行研究における発話矛盾検出方法の多くは外部知識が関係データベースであることを前提としており[1]、発話文の内容と知識とを照合することが容易である。しかし、一般的にドメイン知識はデータベース化されておらず単に文の集合として表されており、先行研究における矛盾検出手法をそのまま適応することは困難である。

発話矛盾検出に類似するタスクとして含意関係認識がある[2]。これは二つの文に対して、前提文と呼ばれる一方の文の内容が真であると仮定したときに、もう一方の文の真偽を判定するタスクである。このタスクにおいてドメイン知識を前提文として与えることで、発話文がドメイン知識と矛盾しているか判定できる。しかし、含意関係認識は1文同士を対象としており、前提文として文の集合を与えることは想定されていない。

本研究では、図1のようなドメイン知識からの前提文生成と含意関係認識を組み合わせた発話矛盾検出モデルを提案する。このモデルでは、ドメイン知識として文集合を想定しており、データベース化されていないドメイン知識でも対応可能である。また、ドメイン知識から矛盾検出に必要な情報を1つの前提文に要約することにより、既存の含意関係認識モデルを利用して矛盾検出が可能となる。

本稿では提案モデル構築のための予備実験について述べる。この実験では、提案モデルにおける前提文生成モデルの学習における課題点を明らかにする。そのために、ベースラインモデルとして、前提文をドメイン知識である文集合から直接選択するモデルの精度と問題点について検証する。

2. 発話文矛盾検出とアプローチ

2.1 問題設定

本研究で提案する発話矛盾検出モデルは、情報案内タスクを行う対話システム内で、システム発話の候補を入力とし、事前に与えられたドメイン知識と照合して、発話内容がドメイン知識と矛盾するかどうかを判定する。矛盾の定義として、文集合の内容を真としたとき、発話文の内容が正しければ「無矛盾」、発話文の内容が誤り、もしくは正誤を判断できない場合は「矛盾」とする。このとき、ドメイン知識とは情報案内に必要な情報を記述

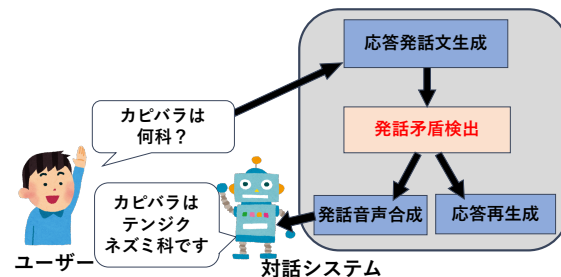


図1: 対話システムにおける発話矛盾検出の利用方法

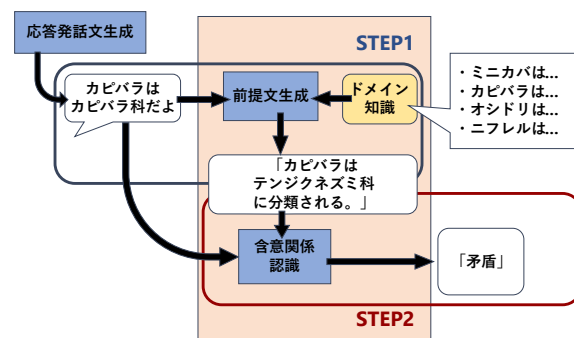


図2: 発話文矛盾検出モデルの概要

した文集合である。これは、対話システムが対話中に参照可能であり、複数のトピックに関して記述されている。特に、案内対象に特有の内容として一般常識とは異なる内容を含む場合がある。

2.2 アプローチ

提案する発話矛盾検出モデルは、図2のように、発話文の矛盾検出を二段階の手順で行う。まずSTEP1では、発話文とドメイン知識の文集合から、含意関係認識に用いる前提文を生成する。発話矛盾を含意関係認識によって検出するために必要な内容を、発話文と文集合を参照することで文集合から抜き出し、前提文として出力する。次にSTEP2では、前提文と発話文で含意関係認識を行い、発話矛盾の有無を判定する。含意関係認識モデルに入力し、出力が「含意する(entailment)」であったときは「無矛盾」、「矛盾する(contradiction)」「関係がない(neutral)」であったときに「矛盾」と読み替え、「無矛盾」または「矛盾」のラベルを出力する。

3. 実験

予備実験として、ベースラインとなる抽出ベースのモデルの精度を検証する。この実験では図2のSTEP1における前提文生成の精度が発話矛盾検出精度に与える影響について分析する。

Premise Sentence Generation for Contradiction Detection of Utterances in Domain-Specific Dialogue Systems: Yuki Horiguchi, Kenta Yamamoto, Ryu Takeda, and Kazunori Komatani (Osaka Univ.)

3.1 実験条件

本実験では、対話システムを水族館ガイドとして利用することを想定する。水族館の利用者から質問を受けて動物や施設に関する情報に回答することを想定したデータセットを作成する。

3.1.1 使用データ

大阪府吹田市の水族館施設であるニフレル[‡]で使用される対話システムを想定した発話文データおよびドメイン知識データセットを作成した。発話文データは、ChatGPT API の gpt-3.5-turbo-0613 モデルを用いて、ユーザによる質問を想定した質問文に対する回答を作成することで136文の発話文を収集した。想定質問は、ニフレルで鑑賞可能な34種類の動物に関して、代表的な特徴、展示場所、分類、生息地の4つを問うものである。この際、モデルの入力にドメイン知識は用いていない。ドメイン知識データは、Wikipedia およびニフレルの公式サイトからニフレルおよびニフレルに存在する動物の知識を記述した文を収集することで、1149文からなる文集合のデータセット構築した。

また、発話文データとドメイン知識データを用いて、各発話文に「矛盾」もしくは「無矛盾」のラベルを付与するアノテーションをおこなった。アノテーションの際はドメイン知識の文集合全体を参照し、発話文内容が文集合に含まれていれば「無矛盾」、含まれていなければ「矛盾」のラベルを付与した。また、アノテーションの判断基準とした文集合内の文を最大5文まで選び、それらを参照文として記録した。

3.1.2 実験手法

前提文抽出手法として、キーワード検索、全文抽出、人手抽出の3種類の手法を用いた場合の矛盾検出精度を比較した。はじめに、各方法を用いて前提文集合を抽出した。キーワード検索は、発話文に含まれるキーワードをもとに前提文を抽出する手法である。ここでは、まず発話文をMeCabで形態素解析し、形態素のうち名詞・動詞・形容詞・副詞の品詞タグが付いた内容語をキーワードとする。次に、ドメイン知識の文集合内の各文が何種類のキーワードを含むかを計測した。そして、キーワードを最も多く含んだ上位 k 文を前提文として抽出した。 $k=1, 2, \dots, 5$ の場合で実験をおこなった。全文抽出は、ドメイン知識の全文を前提文として抽出する。人手抽出は、3.1.1節で述べた参照文のうち1文、もしくは全ての参照文(1~5文)を前提文として抽出する。

抽出した前提文と発話文を既存の含意関係認識モデル[§]に入力し含意関係認識を行った。前提文を複数文を抽出した場合、すべての文に対して「矛盾」と判定した場合に「矛盾」とし、それ以外は「無矛盾」とした。

3.2 実験結果と考察

各前提文抽出手法を用いた際の発話文矛盾検出の結果を表1に示す。全文抽出モデルでは、実験で用いたモデルの中で再現率が最も高いが、適合率は最も低くなった。ドメイン知識の文集合から抽出する文を増やすほど、「無矛盾」な発話を「無矛盾」と正しく判定するために必要な文が前提文集合に含まれる可能性が高くなり、再現率が増加する。その一方で、前提文集合が大きくなるほど正解が「矛盾」であるものを「無矛盾」と誤判定する可能性が高くなり、適合率が減少する。適切な前提文を用

表 1: 発話文矛盾検出精度

モデル	正解率	適合率	再現率	F1	
全文抽出	1文抽出	0.77	0.52	0.42	0.47
	2文抽出	0.77	0.53	0.55	0.54
	3文抽出	0.76	0.50	0.58	0.54
	4文抽出	0.76	0.50	0.61	0.55
	5文抽出	0.75	0.49	0.61	0.54
全文抽出	0.49	0.30	1.0	0.46	
人手抽出(1文)	0.77	0.55	0.36	0.44	
人手抽出(全参照文)	0.80	0.60	0.57	0.57	

表 2: キーワード検索モデルで前提文抽出が誤った例

発話文	前提文
カピバラは、世界最大のネズミ科の動物であり、社会的で水生の生活を好むことが特徴です。	特徴。北アメリカ最大の齧歯目であり、世界でもヨーロッパビーバーと共に南アメリカのカピバラに次いで大きい種である。
カピバラ (Hydrochoerus hydrochaeris, 和名: オニテンジクネズミ (鬼天竺鼠)) は、テンジクネズミ科カピバラ属に分類される齧歯類。	カピバラ (Hydrochoerus hydrochaeris, 和名: オニテンジクネズミ (鬼天竺鼠)) は、テンジクネズミ科カピバラ属に分類される齧歯類。

いるだけでなく、不要な前提文を用いることのないよう前提文の数を調整する必要がある。

キーワード検索モデル同士で比較すると、1文抽出より複数文を抽出するモデルのほうがF1スコアが高くなった。矛盾を判定するために必要な情報は1文ではなく複数文に分散していると考えられる。また、キーワード検索と人手抽出を比較すると、1文抽出の場合はキーワード検索のほうがF1スコアが高くなった。これは人手では主語が欠けた文でも前後の文脈から内容を推察し前提文として抽出できるが、その場合の含意関係認識精度は下がるからだと推測できる。

キーワード検索による前提文抽出が誤った例を表2に示す。キーワード検索により抽出した前提文はアメリカビーバーについて記述した文であり、カピバラについての発話文に対する前提文として相応しくない。アメリカビーバーという主語が欠落しているため、人手では前後の文脈から情報を補完できるため抽出を避けられるが、単純なキーワード検索による抽出では誤りが起きる。このような誤りがキーワード検索の前提文抽出精度を下げる要因となっている。以上の分析結果から、前提文生成においては、複数文にわたる情報の統合と主語の補完の2点が重要な要素となる。

4. おわりに

本研究では、ドメイン知識の文集合から前提文を生成して含意関係認識を行うことで、発話文矛盾を検出する手法を提案した。また、今回は予備実験として、ドメイン知識の文集合から前提文を抽出する手法で発話文矛盾検出を行い、前提文を複数抽出することで精度が向上すること、主語などの情報が欠落した文が誤った前提文抽出の原因となることを確認した。今後は複数の文を用いて情報を補完した前提文を生成するモデルの学習に取り組み、発話矛盾検出システムの精度向上を目指す。

参考文献

- [1] Libo Qin, et al. Don't be contradicted with anything! CI-ToD: Towards benchmarking consistency for task-oriented dialogue system. In *Proc. EMNLP*, 2021.
- [2] 栗原健太郎ほか. JGLUE: 日本語言語理解ベンチマーク. *自然言語処理*, Vol. 30, No. 1, pp. 63–87, 2023.

[‡]<https://www.nifrel.jp/>

[§]<https://huggingface.co/llm-book/bert-base-japanese-v3-jnli>