抑揚による疑問表現を考慮した音声対話システムの提案

坂根剛 目良和也 黒澤義明 竹澤寿幸
広島市立大学大学院　情報科学研究科
sakane@ls.info.hiroshima-cu.ac.jp
{mera, kurosawa, takezawa}@hiroshima-cu.ac.jp

概要

従来の音声対話システムでは音声認識結果をそのまま入力としているため，疑問文と平叙文の区別がつかず，問いかけに対して不適切な応答をすることがあった．そこで本研究では，ノンバーバル情報による疑問表現を考慮できるように，発話の音響的情報に基づいて疑問表現発話を判定し，GPT モデルを用いて疑問符の有無を考慮した応答を生成する手法を組み込んだ音声対話システムを提案する．疑問表現判定実験における正解率は 0.76，また応答発話テキスト生成実験では疑問符を付加した入力に対しては疑問表現を考慮した応答候補が生成されていることが確認された．

1 はじめに

近年，音声対話システムは様々な分野で普及しており，雑談のような非タスク指向型対話についても研究が進められている[1, 2]．しかし音声認識では疑問表現を表す疑問符を表現できないため，従来の音声対話システムでは図 1 に示すように話者からの問いかけを平叙文と誤認識してしまい不適切な応答を返すといった問題が起こる．このような発話テキストに依らない疑問表現は日常対話においてたびたび用いられている．目良ら[3]の音声対話システム実験では, 実験参加者 7 人による延べ 209 発話中 57 件が疑問発話であり, うち 20 件が発話文字列だけでは区別できない疑問表現であった．

そこで本研究では，従来の音声対話システムでは扱えない“抑揚による疑問表現”を考慮して応答を生成する音声対話システムを提案する．また提案システムの要素技術である“発話音声の音響的特徴から疑問表現を判定する手法”と“疑問符の有無を考慮した GPT モデルでの応答生成手法”を実装する．

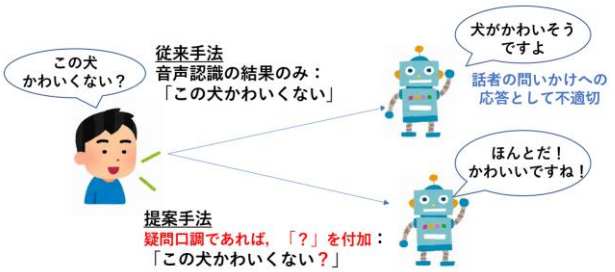

図 1：抑揚による疑問表現への対応

2 関連研究

杉山ら[1]は，雑談を通じて名所についての知識，行った経験，具体的な印象の 3 要素を順に引き出す対話を行うシステムを提案している．このシステムでは，主にテキストベースで対話を進めていき，アジェンダに基づく対話制御の考えに沿って自然な対話の流れを実現していく．

また，藤原ら[2]は，系列変換応答生成モジュールと知識ベース応答生成モジュールによって生成された応答候補からフィルタリングによって最終的な応答を決定する手法を提案している．この手法では知識ベースによりある程度の質問に対応可能であるが，テキストベースのフィルタリング手法を用いているため発話文字列のみでは区別することができない疑問表現に対応することは困難である．

これらの手法ではいずれも話者の口調や表情などのノンバーバル情報を考慮していない．一方，目良ら[3]はユーザのノンバーバル情報から話者感情を推定し，推定した感情を考慮して応答できる音声対話システムを提案している．そこで本研究では，目良ら[3]の手法をベースとして，発話の音響特徴量から話者感情の代わりに疑問表現か否かを判定することで，ノンバーバルな疑問表現を考慮した音声対話システムを実現する．

3 抑揚による疑問表現を考慮した音声対話システム

本論文では，目良ら[3]の“発話の音響的特徴から推定した**話者感情を絵文字として**発話文字列の末尾に付与することによって話者感情を考慮した応答を生成する手法”をベースとして，“発話の音響的特徴から判定した**疑問表現を疑問符として**発話文字列の末尾に付与することによって疑問表現を考慮した応答を生成する手法”を提案する．

3.1 提案手法のシステム構成

図 2 に提案するシステムの概要を示す．まず，話者の発話音声に対して音声認識を行うと同時に，疑問表現判定部で入力発話音声の音響的特徴から疑問表現か否かを判定し，疑問文と判定された場合は音声認識から得られた入力発話文字列の末尾に疑問符を付加する．その入力発話テキストに対して，応答発話生成部ではファインチューニングを行った GPT モデルによって応答発話テキストを生成する．疑問表現判定処理については 3.2 節，応答発話テキスト生成手法については 3.3 節で詳しく説明する．

3.2 音響情報に基づく疑問表現判定

本手法では，発話音声データから算出した静的音響特徴量を LightGBM[4]に学習させることで，入力発話音声が平叙文か疑問文か 2 クラス分類する機械学習分類器を構築する．

学習用音声データとしては，「BTSJ 日本語自然会話コーパス」[5]の対話音声データを使用する．本コーパスには友人同士や初対面同士などの状況での 2 名間の対話が収録されており，対話時間は約 10 分から 30 分である．また，それぞれの対話音声データに対する書き起こし文も用意されている．この対話音声データを 1 発話単位で切り取ることにより，平叙文，疑問文の発話音声データをそれぞれ 200 件ずつ収集した．なお，平叙文，疑問文とする基準は各対話音声データに対する書き起こし文末尾の疑問符の有無としている．

発話音声データから算出する音響特徴量セットとしては，openSMILE の eGeMAPSv02 feature set および emobase2010 feature set を使用する．両 feature set では，それぞれ 88 種類および 1,582 種類の静的特徴量が算出される．eGeMAPSv02 feature set はほとんどの特徴量が算術平均と変動係数のみとなっているのに対して，emobase2010 feature set は各波形の一次導関数を取得することができ，また各波形から線形近似直線の傾きや四分位，標準偏差等の静的特徴量も算出することができる．

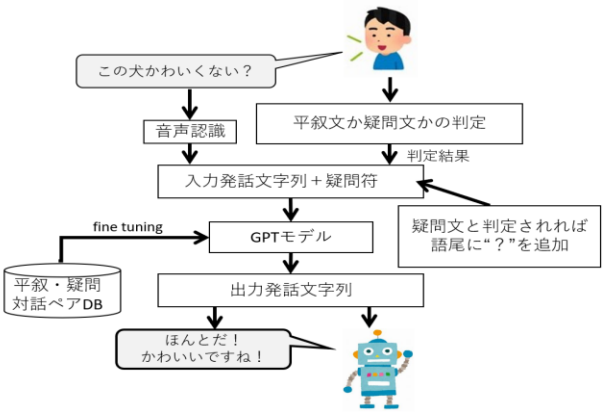

図 2：提案手法のシステム構成

3.3 応答発話テキスト生成手法

本節では，疑問表現発話の末尾に疑問符を付与した入力発話テキストに対して，疑問表現であることを考慮した応答を生成するシステムについて説明する．応答発話テキストを生成するシステムは，rinna 社が公開している GPT-2[6]をベースとして，日本語日常会話コーパス[7]によってファインチューニングを行うことで構築する．日本語日常会話コーパスは，性別・年齢を考慮して選別された協力者 40 名により，日常で発生する会話を協力者自身によって記録してもらうことで収集したコーパスである．本コーパスには収集した対話音声に対応する転記テキストが収録されており，本手法では転記テキストを全てファインチューニング用データとして使用する．

本コーパスの転記テキストには笑いが生じている箇所や音の詰まり等の非言語的情報がタグで表記されているが，本手法ではテキストデータを扱うため，学習の際は読点を除く非言語的情報のタグを削除している．

このようにして前処理を行ったデータから，末尾に疑問符が存在する転記テキストとその直後の転記テキストをペアとする疑問符付きファインチューニング用データを作成した．平叙文にも対応できるように，入力文の末尾が疑問符ではなく読点であるペアも同様に作成している．これにより作成したファインチューニング用データの件数は，平叙文が 523,984 件，疑問文が 53,489 件となった．

4 評価実験

本実験では，発話の音響特徴量を用いた疑問表現の有無の判定処理および，疑問符の有無による応答候補の変化について評価実験を行う．

4.1 疑問表現推定実験

本実験では, 3.2 節で説明した音響特徴量セットから算出される発話の音響特徴量を用いて，疑問表現の有無での 2 クラス分類の正解率の評価を行う．なお emobase2010 feature set は算出される音響特徴量が 1,582 種類と学習データ数よりかなり多いため過学習を起こしてしまう恐れがある．そこで特徴量の重要度上位 20 種の静的特徴量に限定した学習実験も行った．重要度の算出は，ある特徴量が機械学習モデル全体において目的関数の改善に貢献した度合いを示す gain[8]を用いた．表 1 に特徴量重要度上位 20 種を示す．

機械学習実験の結果を表 2 に示す．emobase2010 feature set での特徴量重要度上位 20 種の静的特徴量のみ使用した結果が正解率 0.76 と最も高い結果となった．これは emobase2010 feature set には各波形の一次導関数が存在するため，各波形の時系列変化の傾向を学習できたものと考えられる．また，特徴量重要度上位 10 種の大半は声の高さに関係する基本周波数に関連する静的特徴量であったことから，声の高さの変化は疑問表現を判定する際に有効であるといえる．また，1,582 種類の静的特徴量を特徴量重要度上位 20 種に絞り込むことにより, 関連性が低い特徴量を除外することができたため，正解率が大幅に向上したと考えられる．

4.2 疑問表現に対する応答生成実験

次に，提案手法によって生成された応答発話テキストが入力発話の疑問表現に合った応答となっているかについての実験を行う．発話文字列のみでは疑問表現か否か区別できない文字列を実際の対話事例から収集し，疑問符がある場合とない場合の両方のテキストに対して応答発話テキストを生成した．

本手法によって生成された応答発話テキストの例を表 3 に示す．この結果より，発話文字列のみでは疑問表現として区別できない文字列でも疑問符を付加することで適切な応答発話テキストを生成できることを確認できた．

表 1：重要度上位 20 種の静的特徴量

順位	特徴量
1	F0final_sma_de_linregc1
2	F0finEnv_sma_kurtosis
3	F0finEnv_sma_maxPos
4	shimmerLocal_sma_de_linregc2
5	pcm_fftMag_mfcc_sma_de[13]_kurtosis
6	pcm_fftMag_mfcc_sma[9]_linregc1
7	F0final_sma_linregc2
8	pcm_fftMag_mfcc_sma_de[3]_minPos
9	F0finEnv_sma_iqr1-2
10	shimmerLocal_sma_linregc1
11	lspFreq_sma_de[1]_amean
12	F0final_sma_linregc1
13	lspFreq_sma[7]_minPos
14	lspFreq_sma[4]_linregerrA
15	logMelFreqBand_sma[1]_kurtosis
16	pcm_fftMag_mfcc_sma_de[2]_quartile3
17	lspFreq_sma_de[1]_quartile3
18	logMelFreqBand_sma_de[1]_amean
19	logMelFreqBand_sma_de[2]_amean
20	pcm_fftMag_mfcc_sma_de[10]_kurtosis

表 2：抑揚による疑問表現推定手法の正解率の比較

eGeMAPSv02	emobase2010	重要度上位 20 種
0.50	0.63	0.76

表 3：提案手法によって生成された応答発話

(a) 「この犬かわいくない（？）」への応答発話

疑問符	生成された応答発話候補
有り	・そうだよね ・犬好き？ ・うんかわいい
無し	・いい子なんだけどね ・かわいくない ・ちっちゃいころはよさげだけどね

(b) 「元気（？）」への応答発話

疑問符	生成された応答発話候補
有り	・うん。 ・元気です元気です。 ・元気なのよ。
無し	・それはよかった。 ・すげえ元気だ。 ・ね。

(c) 「体調は大丈夫（？）」への応答発話

疑問符	生成された応答発話候補
有り	・大丈夫。 ・風邪引いてない ・大丈夫だよ。
無し	・よかった。飲み物を飲んでます ・お。そうね。 ・病院行って。

5 おわりに

本研究では，従来の音声対話システムでは扱えない“入力発話音声の抑揚による疑問表現”を考慮して応答を生成する音声対話システムを提案し，システムの要素技術である“発話音声の音響特徴量から疑問表現を判定する手法”と“疑問符の有無を考慮した GPT モデルでの応答生成手法”を実装した．要素技術に対する評価実験の結果，疑問表現判定実験における正解率は 0.76 であった．また，「抑揚による疑問表現を疑問符によって表現した入力発話テキストに対しては，疑問表現を考慮した応答発話テキストを生成することができた．

今後は，抑揚による疑問表現検出性能の向上と，得られた応答候補から適切な応答を選択するフィルタリング機能を開発することで，提案手法に基づく音声対話システム全体を構築する予定である．

謝辞

本研究の一部は国立研究開発法人科学技術振興機構（JST）の「COI プログラム令和 4 年度加速支援（グラント番号 JPMJCA2208）」の支援によって行われた．

参考文献

[1] 杉山弘晃，成松宏美，水上雅博，有本庸浩：自然な流れに沿って対話を進めるアジェンダベース雑談対話システム，人工知能学会研究会資料 SIG-SLUD-B902-12 , pp58-61, 2019.

[2] 藤原吏生，岸波洋介，今野颯人，佐藤志貴，佐藤汰亮，宮脇峻平，加藤拓真，鈴木潤，乾健太郎：ILYS aoba bot：大規模ニューラル応答生成モデルとルールベースを統合した雑談対話システム，人工知能学会研究会資料 SIG-SLUD-C002-25, pp110-115, 2020.

[3] 目良和也，黒澤義明，竹澤寿幸：ユーザの非言語的な感情表出を考慮した音声対話手法，知能と情報（日本知能情報ファジィ学会論文誌）, Vol. 34, No. 3, pp.555-567, 2022.

[4] G.Ke, Q.Meng, T.Finley, T.Wang, W.Chen, W.Ma, Q.Ye , and T.Liu. LightGBM : A Highly Efficient Gradient Boosting Decision Tree, In 31st Conference on Neural Information Processing Systems (NIPS 2017)，pp.1-9, 2017.

[5] 宇佐美まゆみ監修，『BTSJ 日本語自然会話コーパス(トランスクリプト・音声)2021 年 3 月版』，国立国語研究所，機関拠点型基幹研究プロジェクト「日本語学習者のコミュニケーションの多角的解明」, 2021.

[6] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever: Language Models are Unsupervised Multitask Learners, Technical Report OpenAI, 2019.

[7] 小磯花絵・天谷晴香・石本祐一・居關友里子・臼田秦如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉・渡邊友香「日本語日常会話コーパスの設計と特徴」（言語処理学会第 28 回年次大会発表論文集）

[8] xgboost developers, Introduction to Boosted Trees, https://xgboost.readthedocs.io/en/latest/tutorials/model.html#learn-the-tree-structure (2022 年 1 月 12 日アクセス)