

発話の音響情報に基づく疑問表現判定手法

A method for determining question expression based on acoustic information of utterance

坂根 剛

Sakane Tsuyoshi

広島市立大学 言語音声メディア工学研究室

Language and Sound Media Laboratory, Hiroshima City University

概要 本研究では、音声対話システムにおいて話者からの質問や投げかけへの応答を改善するため、発話の音響情報に基づいて疑問表現発話を判定する手法を提案する。本報告では、発話の音響特徴量を使用して発話が平叙文か疑問文かを 2 クラス分類する機械学習器を提案する。また、疑問表現を判定するにあたって有効な音響特徴量についての検討も行った。

1 はじめに

近年、対話型音声返答システムは様々な分野で普及し、雑談のような非タスク指向型の対話についても研究が進められている。目良ら[1]は、従来の音声対話システムが“字面でしか判断をすることができない”といった問題点の解決法として、声の調子や表情などのノンバーバル情報から推定した話者感情を絵文字として音声認識結果に付与した入力文字列に対して返答発話文字列を出力する方法を提案している。しかし、音声認識では疑問表現を表す“?”という文字は表現できないため、図 1 に示すように話者からの質問や投げかけに対し、発話文字列に“?”がないことにより不適切な応答を返すといった問題が起こる。

そこで本研究では、音声対話システムが抑揚による疑問表現に対応できるようにするための、発話の音響情報から疑問表現を判定する手法を提案する。発話音声から音響特徴量を算出し、それらの音響特徴量を機械学習器に学習させることにより、疑問表現特有の声の上がり具合や声の高さなどから平叙文か疑問文かを 2 クラス分類する学習器を構築する。また、算出された音響特徴量について、疑問表現を判定するにあたって、有効な音響特徴量についての検討も行う。提案手法を[1]の音声対話システムに実装したシステムの流れを図 2 に示す。

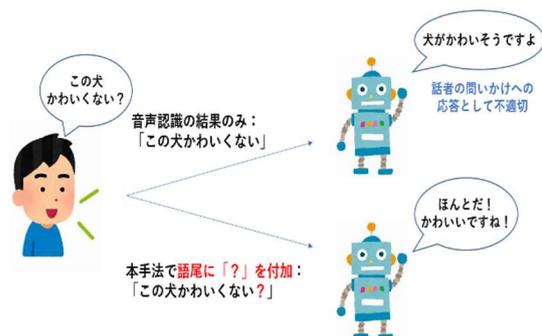


図 1：話者の問いかけに対する応答例

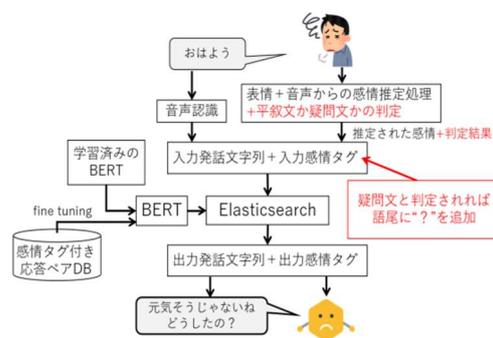


図 2：提案手法実装後の音声対話システムの流れ

2 音響情報に基づく疑問表現判定手法

本研究で構築する機械学習器は、発話音声データから音響特徴量を算出し、それらを使用した機械学習を行い、発話が平叙文か疑問文かを 2 クラス分類するものである。

2.1 発話音声データの収集

本研究では、学習用の発話音声データとして、「BTSJ 日本語自然会話コーパス」[2]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 名で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語尾の“?”の有無としている。

2.2 発話音声データからの音響特徴量

発話音声データから音響特徴量を算出するため、特徴量セットとしてopenSMILE[3]のeGeMAPSv02 feature set および emobase2010 feature set を使用した。それぞれ88種類、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も取得することができ、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を算出することができる。

2.3 機械学習器

本研究では機械学習器として決定木アルゴリズムに基づいた勾配ブースティングの手法を用いたLightGBM[4]を使用する。ブースティングは前の学習器の結果を次の学習データに反映をさせ学習を行っていく手法である。

3 評価実験

本実験では、それぞれの特徴量セットにて算出した音響特徴量を使用し、実験方法は層化10分割交叉検証にて2クラス分類の正解率の比較を行った。emobase2010 feature set で算出される静的特徴量は1,582種類であり、収集した学習データ数よりかなり多いために過学習を起こしてしまう恐れがある。そのため、LightGBMで特徴量の重要度を算出し、上位20種の静的特徴量を使用した実験も行った。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[5]を用いた。表1に特徴量重要度上位10種の特徴量を示す。大半は声の高さに関する基本周波数の静的特徴量であった。

機械学習実験の結果を表2に示す。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がわかる。また、特徴量重要度上位10種の大半は声の高さに関する基本周波数に関連する静的特徴量であり、疑問文では発話の末尾で声の高さが上がることから、疑

表1：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 iqrl-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：正解率の比較

eGeMAPSv02	emobase2010	上位20種
0.50	0.63	0.76

問表現を判定する際に有効な静的特徴量であるといえる。そのため、emobase2010 feature set で算出した静的特徴量を使用した方が、疑問表現をより正確に判定できたと考える。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、過学習を抑えられた、疑問表現を判定するにあたって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る。

4 おわりに

本研究では、発話の音響情報に基づいた疑問表現を判定する手法を提案し、発話の音響特徴量を使用した機械学習を行い、発話が平叙文か疑問文かを2クラス分類する機械学習器を構築した。

現在本研究で作成した機械学習器を[1]の音声対話システムに実装済であるが、対話実験を行った際に、応答ペアデータベースに適切な用例がなければ“?”の有無で応答が変化しなかった。そこで今後は応答を既存のデータから選択して決定するのではなく、モデルで応答をいくつか生成し、その中からフィルタリング等の応答選択手法を用いて応答を決定する手法に取り組む予定である。

5 参考文献

- [1] 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uzzy学会論文誌), 2022年8月 (予定).
- [2] 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.
- [3] F.Eyben , M.Wöllmer , and B.Schuller : “openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor, ” ACM Multimedia Conference – MM, pp.1459-1462, 2010.
- [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), Long Beach, CA, USA. pp.1-9, 2017.
- [5] Introduction to Boosted Trees,
<https://xgboost.readthedocs.io/en/latest/tutorials/model.html#learn-the-tree-structure> (2022/7/15 アクセス)