

# 音響情報を利用した音声対話システムにおける破綻検出

## Detection of Dialogue Breakdowns in Spoken Dialogue System using Acoustic Information

阿部元樹 梅井良太 狩野芳伸 綱川隆司 西田昌史 西村雅史

Motoki Abe, Ryota Togai, Yoshinobu Kano,  
Takashi Tsunakawa, Masafumi Nishida, Masafumi Nishimura

静岡大学総合科学技術研究科

Graduate School of Integrated Science and Technology, Shizuoka University

**Abstract:** Accurate detection of dialogue breakdown is important for recovering from mistakes that dialogue system made. As for spoken dialogue system, recognition errors are inevitable, and they tend to make breakdown detection difficult. In this study, first we evaluate how the speech recognition error affect the performance of the conventional breakdown detector using language information. Secondly, we construct a dialogue breakdown detector using acoustic information and evaluate how well it works. Finally, we compare their performances against spoken dialogue.

### 1. はじめに

2015年より shared task として、雑談対話における対話破綻検出性能を競う「対話破綻検出チャレンジ」が開催されている<sup>[1]</sup>。本タスクはテキストチャットを対象として行っており、言語情報に基づく破綻検出が可能である。対して音声対話の場合、音声認識の過程で認識誤りが発生することから正しい言語情報を得られるとは限らない。今回、自動対話システムで収集した雑談音声対話を対象として、誤認識がこれまでの言語ベースの破綻検出器に与える影響を調査するとともに、先に見出した対話破綻と発話の音響特徴量の関係を参考にして、新たに音響特徴量に基づく対話破綻検出器を構築し、その性能を既存の言語ベースの検出器と比較評価する。

### 2. 評価対象雑談音声対話データ

本研究で用いる雑談音声対話データとして、先行研究で収集された対話データを用いた<sup>[2]</sup>。このデータは自動対話制御を行う対話システムと被験者との間で雑談対話を行ったものである。本研究では男子大学生の被験者3名から合計44対話、1271発話を収録した。このデータを学習とテストの両方で用いる（交差検証）。

次にこれらに対し、3名のアノテータが、対話破綻ラベルの付与を行なった。付与するラベルは「対話破綻検出チャレンジ」と同様に、O:T:Xの3値とした（表1参照）。ただし、正解ラベルの決定について

は後述する言語情報に基づく対話破綻検出器が O:X の2値分類に特化したものであることを踏まえ、TをX側に含めた O:(T+X)として、多数決で決定した。なお、ラベル付与結果及びアノテータ間の一致度（weighted  $\kappa$  statistics）を表2に示す。

表1：ラベルの付与基準

label	基準
O	破綻ではない
T	破綻とは言い切れないが、違和感を感じる発話
X	明らかにおかしいと思う破綻した発話

表2：コーパスの O:(T+X)の内訳及び一致度

O	N=810 (0.637)
(T+X)	N=461 (0.363)
weighted $\kappa$ statistics	0.666

### 3. 言語ベース対話破綻検出器

ここでは言語情報に基づく対話破綻検出器として、先に提案した word2vec による単語間距離と関連ルールに基づく検出器を用いる<sup>[3]</sup>。

### 4. 音響ベース対話破綻検出器

今回、先に検討した対話破綻と音響特徴量の関係<sup>[4]</sup>を参考にして、音響情報ベースの識別器を構築した。識別器としては SVM(多項式カーネル)を用い、特徴量としては Interspeech 2009 emotion challenge の 384次元の特徴量<sup>[5]</sup>を用いた。なお、ここでは言語情

報は一切用いない。

## 5. 実験

本研究では以下の2つの実験を行った。

### 5.1. 言語ベース対話破綻検出器の性能評価

音声認識誤りの有無が性能にどの程度の影響を与えるかを検証するために、人手による書き起こし対話ログと Julius<sup>[6]</sup>による音声認識対話ログに対して、言語ベース対話破綻検出器を適用し、比較評価を行った。なお、今回のデータにおける Julius の文字誤り率(CER)は62%であった。

評価方法としては交差検証を用いて実施した。評価指標はO, (T+X)に対してPrecision, Recall, F-Measure, 全体のAccuracyである。なお、Juliusについてはディクテーションキットを用いた。

表3, 表4に結果を示す。これらより、音声認識誤りによって対話破綻検出性能が劣化することが分かる。また、本検出器の傾向として(T+X)側に判定をすることが多い。

表3: 書き起こし対話ログを用いた場合の性能評価結果

label	Precision	Recall	F-Measure	Accuracy
O	0.826	0.156	0.262	0.430
(T+X)	0.384	0.950	0.547	

表4: 音声認識対話ログを用いた場合の性能評価結果

label	Precision	Recall	F-Measure	Accuracy
O	0.820	0.113	0.198	0.413
(T+X)	0.378	0.961	0.543	

### 5.2. 音響ベース対話破綻検出器の性能評価

次に、音響ベース対話破綻検出器が雑談音声対話における対話破綻検出にどの程度有効かについて評価を行った。なお、評価方法と評価指標については前述した言語ベース対話破綻検出器の性能評価時と同様である。

表5に結果を示す。表4と表5を比較すると、Accuracyでは音響ベースの検出器の性能が言語ベースのものを上回っていることが分かる。なお、検出の傾向としては、(T+X)側の判定が多い言語ベース対話破綻検出器に対して音響ベース対話破綻検出器はO側への判定が多いことが分かる。また、(T+X)側についても適合率は上回っており、再現率は低いものの比較的正しく判定できていることが分かる。これらより、言語ベース対話破綻検出器と音響ベース対話破綻検出器の判定傾向に大きな差異があることが

確認できる。

表5: 音響情報を用いた対話破綻検出器の性能評価結果

label	Precision	Recall	F-Measure	Accuracy
O	0.661	0.681	0.671	0.575
(T+X)	0.414	0.392	0.402	

## 6. おわりに

本研究では、雑談音声対話システムにおいて、既存の言語ベースの対話破綻検出器が音声認識誤りによって受ける影響の大きさを確認するとともに、新たに音響ベースの対話破綻検出器を提案し、その性能を比較した。結果として、言語情報を一切用いない音響ベースの対話破綻検出器が、言語ベースのものと比較して高い性能を出すことを確認することができた。今後の課題として、言語情報と音響情報を併用した検出器の構築に向けての検討が挙げられる。

## 謝辞

言語情報を用いた対話破綻検出器の評価について、ご協力を頂きました柴淳氏に感謝します。本研究の一部はJSPS 科研費 16K01543 の助成を受けたものである。

## 参考文献

- [1] 東中竜一郎, 船越孝太郎, 小林優佳, 稲葉通将: 対話破綻検出チャレンジ, 言語・音声理解と対話処理研究会, vol. 75, pp.27-32, (2015)
- [2] 梅井良太, 中島悠, 伊東伸泰, 西田昌史, 西村雅史: 非言語音響情報を利用した聞き役対話システムに関する検討, 情報処理学会第78回全国大会, 6Q-03, (2016)
- [3] 柴淳, 狩野芳伸: 単語の意味の距離から検出する対話破綻, 言語・音声理解と対話処理研究会, vol.78, pp.72-74, (2016)
- [4] 阿部元樹, 梅井良太, 綱川隆司, 西田昌史, 西村雅史: 非言語音響情報を利用した対話破綻検出に関する検討, 情報処理学会第79回全国大会, 6M-04, pp.225-226, (2017).
- [5] Bjorn Schuller, Stefan Steidl, Anton Batliner: The INTERSPEECH 2009 Emotion Challenge, INTERSPEECH, pp.312-315, (2009)
- [6] 河原達也, 李伸晃: 連続音声認識ソフトウェア Julius, 人工知能学会誌, vol.20, No.1, pp.41-59, (2005)