

個人差と対話行為を考慮した対話破綻検出に関する検討

A Study on Dialogue Breakdown Detection Considering Individual Variations and Interactive Behavior

阿部 元樹[†] 梶井 良太[†] 綱川 隆司[†] 西田 昌史[†] 西村 雅史[†]
 Motoki Abe Ryota Togai Takashi Tsunakawa Masafumi Nishida Masafumi Nishimura

1. はじめに

近年、非タスク指向型の雑談対話システムが注目を集めており、当該分野の研究が積極的に行われている。しかし、ドメインが限定されない雑談対話は対話制御が困難である。加えて音声対話の場合は音声認識誤りが生じることから、結果としてシステムがユーザに対して不適切な返答をする「対話破綻」という状況が発生することが多い。

対話破綻への対処法については、まず対話破綻が起こらないように対話制御を高度化するというアプローチ^[1]がある一方、事後的ではあるがシステムが、対話破綻が発生したことを自動検知するというアプローチがあり、もし自動検出が出来れば素早いリカバリが可能になると考えられる。これまでにテキストチャットを対象とした「対話破綻検出チャレンジ」といった研究課題^[2]が提唱され、多くの関連研究が行われている^[3]。

我々は先に雑談音声対話を対象とし、非言語音響情報を用いた対話破綻検出方法を提案した^[4]。今回、更に対話破綻発生時のシステム側の対話行為によって対話破綻発生比率やその際の音響的变化が異なること、対話破綻時に生じると考えられるユーザ音声の音響的变化は話者ごとに異なるということ、以上 2 つの仮説を立て、先行研究^[5]で収集された雑談対話音声データを用い、対話破綻時と非対話破綻時とで対話行為と話者間での音響的差異を分析した。

2. 雑談音声対話データ

2.1 対話コーパス

本研究で用いる雑談音声対話データとして、先行研究^[5]で収集された対話コーパスを用いた。このコーパスは発話区間中の平均発話音量、発話長、有音率をパラメータとして用いた対話システムで収集されたもので、本研究では男子大学生の被験者 3 名に対して被験者ごとに 6 対話 (1 対話 29 ターン) 行ったものを用いた。

2.2 ラベル付与

このコーパスは音声ファイルのみのため、アノテータ 1 名に書き起こしと対話行為ラベル、対話破綻ラベルの付与を行ってもらった。対話行為ラベルと対話破綻ラベルについてはそれぞれ先行研究^{[2][6]}で用いられているものを用いた (表 1, 2)。なお、対話行為ラベルのうち非共感ラベルについては、本コーパスでそれに該当する発話がなかったため、ここでは省略する。

表 1: 対話行為ラベルとその発話例

| ラベル | 発話例 |
|------|---|
| 対話管理 | こんにちは、今日は色々教えてください、最近ハマっている趣味はありますか、教えてください |
| 質問 | あなたは近くのスーパーでコロッケを買いますか? |
| 共感 | 私もそう思います |
| 自己開示 | 私は豆腐が好きなんですよ |
| 相槌 | うんうん、それで? |

表 2: 対話破綻ラベルとその基準

| ラベル | 基準 |
|-----|---|
| ○ | 破綻ではない (当該システム発話の後対話を問題なく継続できる) |
| △ | 破綻とは言い切れないが、違和感を感じる発話 (当該システム発話の後対話をスムーズに継続することが困難) |
| × | 明らかにおかしいと思う破綻した発話 (当該システム発話の後対話を継続することが困難) |

3. 対話破綻分析

3.1 音響的素性

音響的素性としては、表 3 に示す 6 個を用いた。これらは予備検討や先行研究で用いられているものを参考に選定した。なお、発話長、無音区間長、有音区間長、有音率については Audacity^[7]を用いて、平均ピッチ、ピッチレンジについては Praat^[8]を用いて算出した。

表 3: 音響的素性とその概要

| 音響的素性 | 概要 |
|--------|------------------------------|
| 発話長 | システム発話終了時からユーザ発話終了時までの時間長[s] |
| 無音区間長 | 発話長と有音区間長の差[s] |
| 有音区間長 | 300ms 以上有音が続いた区間の合算[s] |
| 有音率 | 発話長における有音区間長の割合 |
| 平均ピッチ | 発話区間の平均 F0[Hz] |
| ピッチレンジ | 発話区間の最大 F0 と最小 F0 の差[Hz] |

3.2 分析結果

2.2 節で行った対話破綻ラベルと対話行為ラベル付与の結果、表 4 の結果が得られた。また、各被験者の対話破綻時と非対話破綻時での 6 個の音響的素性の差異について平均と分散を分析し、図 1 の結果が得られた。表 4 より、3 名の被験者に共通してシステムが「共感」「自己開示」の対話行為の発話をした時、対話破綻が起こる比率が高いことが分かった。これらの対話行為は雑談対話ならではの重要な対話行為ではあるが、それ故に対話破綻が発生しやすいという課題を抱えていることを示している。加えて、図 1 より、被験者ごとに変化する音響的素性とその変化度合いは異なることが分かった。

[†] 静岡大学 総合科学技術研究科

表 4: 各被験者 (A, B, C) の対話破綻ラベル、対話行為ラベル付与結果

| A | ○ | △ | × | B | ○ | △ | × | C | ○ | △ | × |
|------|----|---|---|------|----|---|----|------|----|---|---|
| 対話管理 | 28 | 1 | 0 | 対話管理 | 25 | 0 | 0 | 対話管理 | 31 | 5 | 5 |
| 質問 | 41 | 4 | 2 | 質問 | 44 | 1 | 3 | 質問 | 35 | 2 | 4 |
| 共感 | 21 | 3 | 8 | 共感 | 23 | 4 | 9 | 共感 | 23 | 4 | 6 |
| 自己開示 | 38 | 2 | 3 | 自己開示 | 28 | 4 | 0 | 自己開示 | 26 | 2 | 4 |
| 相槌 | 11 | 6 | 6 | 相槌 | 16 | 7 | 10 | 相槌 | 16 | 6 | 5 |

| 被験者 | ○ : (△ + ×) |
|-----|-------------|
| A | 139 : 35 |
| B | 136 : 38 |
| C | 131 : 43 |

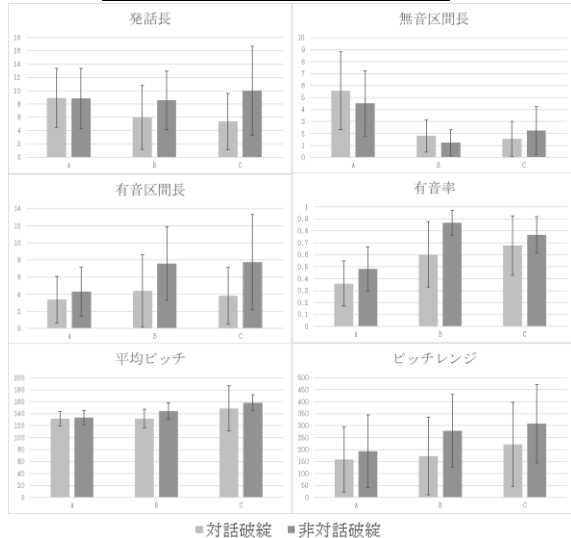


図 1: 各被験者 (A, B, C) の対話破綻時と非対話破綻時での音響的素性の平均・標準誤差

4. 対話破綻検出実験

4.1 各被験者の音響的素性

3.2 節の分析結果より、各被験者で対話破綻検出に有効な音響的素性が異なることが確認されているため、対話破綻検出器の音響的素性も被験者ごとに対話破綻検出に有効なものを選定する必要がある。音響的素性の選定は greedy stepwise (forwards) を用い、表 5 に示す音響的素性を選定した。

表 5: 各被験者の音響的素性選定結果

| 被験者 | 音響的素性 |
|-----|---------------------------------------|
| A | 無音区間長, 有音区間長, 有音率 |
| B | 発話長, 無音区間長, 有音区間長, 有音率, 平均ピッチ, ピッチレンジ |
| C | 発話長, 有音区間長, 平均ピッチ, ピッチレンジ |

4.2 対話破綻検出実験の結果

4.1 節で選定した音響的素性を用いて対話破綻の自動識別を試みる。識別器としてはナイーブベイズを用いた、正解は表 2 の○の場合を非対話破綻、△と×を対話破綻とした。Leave-one-out 法による交差検証で性能評価を行った。また、データの偏り補正を行い、対話破綻 : 非対話破綻の比率を 1 : 1 にした。評価指標としては対話破綻, 非対話

破綻それぞれの適合率・再現率・F 値, 検出器の識別精度, AUC である。以下表 6 に結果を示す。

表 6 より、被験者 B, C はある程度の検出は出来ている一方、被験者 A についてはモデルが上手く構築できなかったことが分かる。原因として、図 1 にあるように被験者 A は対話破綻時と非対話破綻時の音響的变化が少ないことが挙げられる。

表 6: 対話破綻検出実験の結果

| 被験者 | 対話破綻時 | | | 精度 | AUC |
|-----|--------|-------|-------|-------|-------|
| | 適合率 | 再現率 | F 値 | | |
| A | 0.565 | 0.626 | 0.594 | 0.572 | 0.594 |
| | 非対話破綻時 | | | | |
| B | 0.781 | 0.603 | 0.680 | 0.717 | 0.767 |
| | 非対話破綻時 | | | | |
| C | 0.633 | 0.710 | 0.669 | 0.649 | 0.725 |
| | 非対話破綻時 | | | | |

5. おわりに

本研究により、雑談音声対話における対話破綻検出において、システムの対話行為に応じて対話破綻発生比率が変化すること、対話破綻時の音響的素性の変化は話者ごとに異なることが確認された。今後の方針として、より有用な音響的素性を見出し、言語情報との組合せによる破綻検出器の構築を行いたい。

謝辞

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] 大西 可奈子, 吉村 健, “コンピュータとの自然な会話を実現する雑談対話技術”, NTT DoCoMo テクニカル・ジャーナル, Vol.21, No.4, pp.17-21, (2014).
- [2] 東中 竜一郎, 船越 孝太郎, 荒木 雅弘, 塚原 裕史, 小林 優佳, 水上 雅博, “テキストチャットを用いた雑談対話コーパスの構築と対話破綻の分析”, 自然言語処理, Vol.23, No.1, pp.59-86, (2016).
- [3] 杉山 弘晃, “発話生成における誤りパターンの分析に基づく対話破綻検出”, 言語・音声理解と対話処理研究会, No.78, pp.81-84, (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] 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).
- [6] 東中 竜一郎, 堂坂 浩二, 磯崎 秀樹, “対話システムにおける共感と自己開示の効果”, 言語処理学会第 15 回年次大会発表論文集, pp.446-449, (2009).
- [7] Audacity (<http://forest.watch.impress.co.jp/library/software/audacity/>)
- [8] Praat (<http://www.fon.hum.uva.nl/praat/>)