

対話システムにおける個人特性を考慮した破綻度合い推定

山田竜彰¹ 坪倉和哉² 入部百合絵¹ 北岡教英³

¹愛知県立大学 情報科学部 ²愛知県立大学 大学院情報科学研究科

³豊橋技術科学大学 情報・知能工学系

{is211089, id231001}@cis.aichi-pu.ac.jp iribe@ist.aichi-pu.ac.jp

kitaoka@tut.jp

概要

ユーザが対話破綻に対して破綻と判断する度合い（以降、破綻度合いと呼ぶ）には個人差があることが報告されている。そのため、対話を円滑に進めるために必要な対話修復の戦略を考える際、破綻に対して修復すべきかどうか、あるいはどのように修復すべきかといった判断をするときに、破綻度合いの個人差は無視できない。そこで本研究では、そのような個人差が生じる要因が性格や年代、性別などの個人特性と関係していると考え、個人特性を考慮した破綻度合いの推定手法を提案する。加えて、破綻度合いはこれまでの対話履歴にも依存するため、対話履歴と個人特性に関する情報を併用して破綻度合いを推定する。システムの発話に対する破綻度合いとユーザの個人特性のデータを収集し、個人特性が破綻度合いに与える影響について調査を行った結果、勤勉性や神経症傾向などの7種類の個人特性によって破綻度合いに差が生じることが明らかとなった。また分析の結果に基づき破綻度合い推定器を構築した結果、8割以上の精度 (accuracy) を得ることができ、対話修復への活用が期待できる。

1 はじめに

現在の対話システムでは事実と異なる発言をする、急な話題転換を行う、同じ内容の質問を繰り返す、などの対話破綻が生じている。このような対話破綻が発生すると、ユーザが対話システムとの対話の継続が困難になることや、対話継続意欲が低下してしまうことが問題となるため、対話破綻を適切に検出する必要がある。対話破綻検出に関する先行研究として、「対話破綻検出チャレンジ」[2]が開催され、テキスト対話を対象に破綻検出が行われている。しかし、対話破綻に対するユーザの反応には個人差があることが報告されており[1]、破綻の感じ方に関する個人差を考慮することで、破綻対話に対する修復

精度の向上ができると考える。

本研究では、ユーザの破綻に対する個人差を考慮した対話破綻度合いの推定手法を提案する。まず破綻度合いがどのような個人特性によって異なるのかを調査し、その上で個人特性を考慮した破綻度合い推定器を構築する。これにより、ユーザに適応した対話破綻の検出が可能となる。

2 破綻度合いの収集と分析

2.1 対話破綻文の収集

対話破綻に対する度合いを示したラベルと個人特性を含む対話コーパスは存在しないため、破綻度合いのラベルと個人特性に関するデータを収集した。クラウドワーカーに対話破綻文を提示し、破綻度合いを回答してもらうことで、破綻度合いを収集する。次に、人と対話システムとの対話コーパスの中から、対話が破綻している対話破綻文と、その前の4発話を抽出した。破綻までの発話の流れをクラウドワーカーに示すためである。使用したコーパスは3種類あり、対話システム Hobbyist(HBY)[3]、ILYS-AOBA(ILA)[4]と人との対話データと、Capex 雑談対話コーパス[5]から対話破綻文を抽出した。HBY, ILAの対話データは対話システムライブコンペティションの予選データ[6]から収集したもので、HBY が 52 対話 (1 対話 30 発話)、ILA が 51 対話 (1 対話 30 発話) のデータである。Capex 雑談対話コーパスとは、株式会社 Capex が運営する雑談対話 AI アプリの対話ログに基づく約 15,000 対話 (1 対話 5 発話) のデータである。それぞれのデータについて、東中ら[7]の研究で示されている破綻類型 17 種類に非破綻を加えた、18 種類のアノテーションを行った。アノテーションは3人で実施した。破綻類型のうち、東中ら[7]の研究で示されている、発生確率の高い破綻類型である、誤情報、発話意図不明確、自己矛盾、繰り返しの4種類を破綻文として取り上げる。1 対

話はシステム発話, ユーザ発話, システム発話, ユーザ発話, システム発話 (破綻) の 5 発話からなり, 189 対話選定した.

2.2 破綻度合いの収集

126 人のクラウドワーカーに対話破綻文に対する破綻度合いの回答を依頼した. 破綻度合いは, 1: 破綻していない, 2: どちらかといえば破綻していない, 3: どちらともいえない, 4: どちらかといえば破綻している, 5: 破綻している, の 5 段階である. 既存研究[2]では, 破綻, 非破綻, どちらでもない, の 3 段階でシステム発話が破綻か否かをアノテーションしているが, 本研究では 5 段階に設定することで, 個人特性と破綻度合いの関連を詳細に分析する.

2.3 個人特性の収集

本研究では, 個人特性による破綻度合いの違いを分析するため, クラウドワーカーの個人特性を調査するアンケートも実施した. 個人特性には, 既存研究[2]で個人特性が破綻に対する反応に影響を及ぼしていることが示されている TIPI-J[8] に加え, 特性シャイネス尺度[9], KiSS-18[10], ENDCORE (簡易版) [11], 性別, 年齢の 15 種の個人特性を用いることで, 多種多様な個人特性が破綻度合いに与える影響について探る.

2.4 破綻度合いの分析

破綻度合いが個人特性によって異なるのかを調査するため, マン=ホイットニーの U 検定 (Mann-Whitney U test) を行った. まず, クラウドワーカー毎に破綻度合いの平均スコアを算出した. 次に, 個人特性毎に平均値を算出し, 各クラウドワーカーの個人特性がその基準より高いか低いかで, 上位群と下位群に分類した. その後, 上位群のクラウドワーカーと下位群に属するクラウドワーカーの破綻度合いをもとにそれぞれの平均値を算出し, 上位群/下位群間で U 検定を行った. このような検定を 15 種の個人特性と破綻類型毎に実施した. 紙幅の関係上, 5%水準 (両側検定) で有意差が認められた破綻類型と個人特性の組み合わせのみを表 1 に示す. 対話破綻の類型が誤情報である破綻発話に対する破綻度合いと神経症傾向, 繰り返しの破綻発話に対する破綻度合いと開放性についての箱ひげ図を図 1, 2 に示す.

表 1 有意差が認められた破綻度合い

破綻類型	個人特性	p 値	下位群 の破綻 度合い の平均	上位群 の破綻 度合い の平均
誤情報	勤勉性	0.038	4.31	4.09
	神経症傾向	0.039	4.09	4.29
	開放性	0.013	4.29	4.04
	他者受容	0.011	4.33	4.09
	関係調整	0.003	4.30	4.01
繰り返し	外向性	0.030	4.26	4.05
	開放性	0.039	4.24	4.06
	自己主張	0.029	4.27	4.08
	他者受容	0.027	4.28	4.09
	関係調整	0.044	4.23	4.05

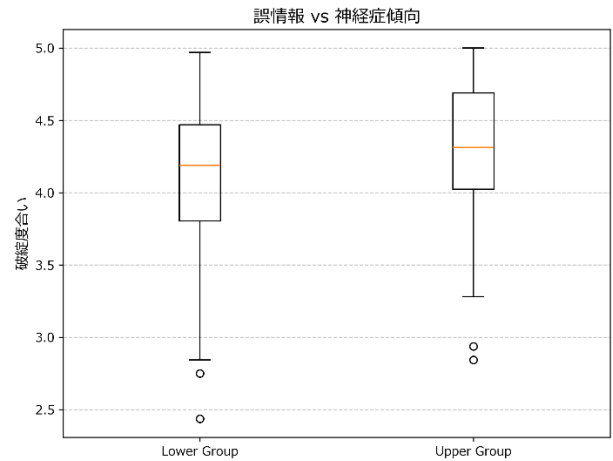


図 1 誤情報, 神経症傾向における箱ひげ図

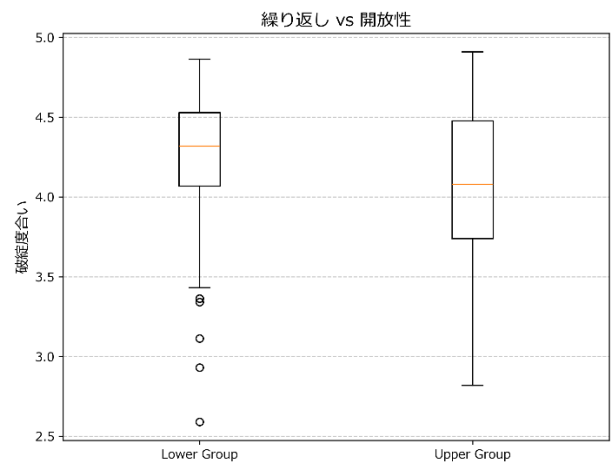


図 2 繰り返し, 開放性における箱ひげ図

破綻類型が誤情報である対話に関しては、神経症傾向のスコアが高い人ほど破綻度合いが高い結果であった。これは、神経症傾向が高い人はシステムの誤った情報に対して真偽を敏感に捉えて、破綻と感じやすいからだと考えられる。

破綻類型が誤情報、繰り返しである対話に関しては開放性、他者受容、関係調整のスコアが高い人ほど破綻度合いが高い。このような特性を持つ人は相手の意見を受け入れる傾向にあるためだと考えられる。

以上より、同じ対話破綻に対しても、個人特性によって破綻度に対する感じ方に有意な差が生じていることが明らかとなった。

3 提案する破綻度合いの推定手法

前章の結果から、個人特性の違いが破綻度合いに影響を及ぼしていることを確認したため、本章では、個人特性を考慮した破綻度合いの推定手法を2種類提案する。手法について以下に詳述する。

3.1 破綻度合いの推定手法 1

推定手法1の概要図を図3に示す。2.1で述べた4つの発話文と1つの破綻文の計5つの発話文を事前学習済みモデルBERT(Bidirectional Encoder Representations from Transformers)に投入し、得られた768次元の文ベクトル表現と、個人特性を表す15次元の特徴量(上位群, 下位群)を連結して783次元の特徴量ベクトルを生成した。この特徴量を入力として、5分割交差検証を行った。識別器にはSVM(Support Vector Machine)(RBFカーネル)とXGBoosterを使用した。破綻度合いを5段階で収集したため、出力は5クラスの分類である。加えて、各段階に対する解釈がクラウドワーカーによって異なる可能性もあるため、各段階の範囲を広げることにも試み、破綻度合い1と2、4と5を結合した3クラス分類も実施した。

3.2 破綻度合いの推定手法 2

推定手法2の概要図を図4に示す。推定手法1では、5発話からなる特徴量ベクトルが768次元であったのに対し、個人特性は15次元であったため、発話に関する特徴と個人特性に関する特徴の次元数に大きな偏りが生じていた。推定手法2では、この問題を解決する。具体的には、推定手法1ではクラウ

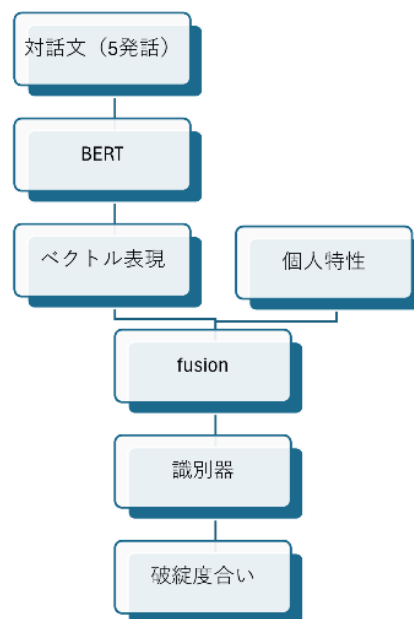


図 3 推定手法1の概要図

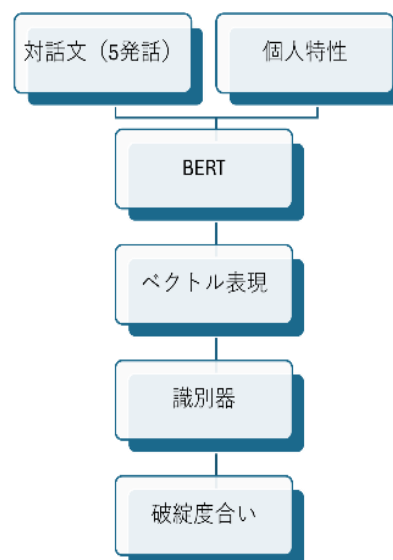


図 4 推定手法2の概要図

ドワーカーから収集した性格特性の数値を上位群と下位群に分けて2値を識別器に入力したが、個人特性を数値表現にするのではなく、「勤勉性のスコアは高いです」のように個人特性の種類と高低がわかるような説明文を生成する手法を提案する。先述した、4つの発話文と1つの破綻文からなる5発話に個人特性を説明する文章を連結したものをBERTと投入し、768次元の特徴量ベクトルを生成した。この特徴量を用いて推定手法1と同じように2種(3クラスと5クラス)の識別を実施した。評価方法は5分割交差検証である。

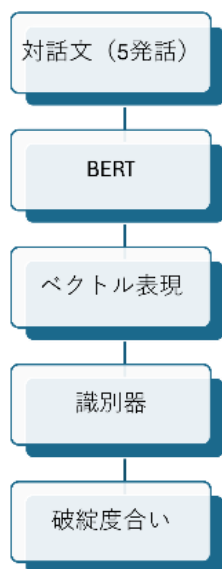


図 5 比較手法の概要図

3.3 提案手法の評価

提案手法 1 と 2 の破綻度合いの推定性能を評価した。結果を表 2 と表 3 に示す。なお、本研究では、個人特性を考慮した破綻度合いの推定手法を提案しているため、個人特性によって精度が向上しているかを確認するため、提案手法 1 の個人特性を結合する部分を省略し、個人特性を与えていない推定器も評価し、提案手法との比較を行った。比較手法の概要を図 5 に示す。

表 2 より 3 クラス分類において、すべての手法で 8 割を超える高い精度 (accuracy) で破綻度合いを推定できていることがわかる。F1 Score を比較すると提案手法 1 の XGBoost が最も高く 0.798 となっている。また、表 3 より、5 クラス分類においては、すべての手法に対して 6 割を超える精度 (accuracy) を得たが、3 クラス分類と比較すると、精度は低下している。3.1 で述べたように、5 段階の 1 と 2、4 と 5 の線引きがクラウドワーカーにより解釈が異なる可能性もあり、本データにおいては 3 段階にすることで、破綻度合いによる差異を適切に学習できることが分かる。表 3 の 5 クラス分類については、F1 Score を比較すると提案手法 1 の XGBoost が最も高く 0.598 となっている。5 クラス分類、3 クラス分類において提案手法 1 (XGBoost) のスコア (F1 Score) が最も高く、比較手法のスコアを上回っているため、

表 23 クラス分類における評価結果

3 クラス分類	Acc.	Pre.	Rec.	F1
提案手法 1(SVM)	0.821	0.776	0.821	0.795
提案手法 1 (XGBoost)	0.818	0.790	0.818	0.798
提案手法 2(SVM)	0.821	0.776	0.821	0.795
提案手法 2 (XGBoost)	0.817	0.783	0.817	0.793
比較手法(SVM)	0.821	0.776	0.821	0.795
比較手法(XGBoost)	0.821	0.776	0.821	0.795

表 35 クラス分類における評価結果

5 クラス分類	Acc.	Pre.	Rec.	F1
提案手法 1(SVM)	0.646	0.540	0.646	0.564
提案手法 1 (XGBoost)	0.645	0.580	0.645	0.598
提案手法 2(SVM)	0.646	0.543	0.646	0.564
提案手法 2 (XGBoost)	0.636	0.544	0.636	0.566
比較手法(SVM)	0.646	0.540	0.646	0.564
比較手法(XGBoost)	0.646	0.546	0.646	0.566

個人特性を特徴量として組み込むことにより、推定精度が向上することが確認された。

提案手法 1 と提案手法 2 を比較すると、すべての評価項目において提案手法 1 の評価が上回っていることがわかる。個人特性の高低の情報を単純に結合する提案手法 1 に対し、提案手法 2 では対話文と個人特性を結合した文章により特徴量を得るため、個人特性を考慮した埋め込みベクトルを取得できることを期待したが、個人特性の情報が対話文の意味と混在して曖昧になり、対話文の主たる意味構造に影響を与え、推定精度が相対的に低下したと推測される。

4 おわりに

本研究では破綻度合いにおける個人差を考慮することで、対話破綻検出精度の向上を目指した。個人特性が破綻度合いに与える影響について調査を行った結果、個人特性によって破綻度合いが変化していることがわかった。分析の結果に基づいた破綻度合い推定器を開発した結果、個人特性を考慮した提案手法が最も高い精度 (F1 Score) で破綻度合いを推定した。今後は、破綻度合い推定を対話システムに適用するための検討を行う。

謝辞

本研究は JSPS 科研費 JP22K19793, JP23H00493 の助成を受けたものです。また、本研究は国立研究開発法人産業技術総合研究所と行っている「高齢者を対象とした音声言語処理に関する研究開発」の一部として実施されたものです

参考文献

1. Kazuya Tsubokura, Yurie Iribe & Norihide Kitaoka: Analysis of the relationship between user response to dialog breakdown and personality traits, *Advanced Robotics*, Vol.38, No.4, pp.246-255, 2024.
2. 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, 第 75 回言語・音声理解と対話処理研究会 (第 6 回対話システムシンポジウム), 人工知能学会研究会資料 SIG-SLUD-B502-07, pp.27-32(2015).
3. 杉山弘晃, 成松宏美, 水上雅博, 千葉祐弥, 目黒豊美, 中嶋秀治: Transformer encoder-decoder モデルによる趣味雑談システムの構築.SIG-SLUD, B5(02), pp.104-109(2020).
4. 藤原更生, 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 .SIG-SLUD,B5(02),pp.110-115(2020).
5. 叶内晨, 赤崎智, 堀江伸太郎, 小亀俊太郎: Capex 雑談対話コーパスの構築とその分析, 言語処理学会第 28 回年次大会 発表論文集, pp.1090-1094(2022).
6. 東中竜一郎, 船越孝太郎, 高橋哲郎, 稲葉通将, 角森唯子, 宇佐美まゆみ, 川端良子, 水上雅博, 小室允人: 対話システムライブコンペティション 3, 第 90 回人工知能学会 言語・音声理解と対話処理研究会(第 11 回対話システムシンポジウム), 人工知能学会研究会資料, pp.96-103(2020).
7. 東中 竜一郎, 荒木 雅弘, 塚原裕史, 水上 雅博: 雑談対話システムにおける対話破綻を生じさせる発話の類型化, 自然言語処理, 29 巻, 2 号, pp. 443-466, (2022).
8. R. Lewis Goldberg: An alternative “Description of Personality”: The Big-Five Factor Structure, *Journal of Personality and Social Psychology*, pp.1216-1229(1990).

9. 堀 弘道, 山本真理子: 心理測定尺度集 I—人間の内面を探索する<自己・個人の内過程>. サイエンス社, 東京, p226, (2001)
10. 菊池 章夫, 思いやりを科学する一向社会的行動の心理とスキル, 川島書店, 212p (1988)
11. 藤本 学: スキルとしてのコミュニケーション 大坊郁夫 (編) 幸福を目指す対人社会心理学—対人コミュニケーションと対人関係の科学—, ナカニシヤ出版, p.193-210, (2012)