Vol.2018-MUS-118 No.24
Vol.2018-SLP-120 No.24
2018/2/21

システム・ユーザ発話に着目した対話破綻検出

阿部元樹 [a)]　栂井良太　綱川隆司　西田昌史　西村雅史

概要：雑談対話システムにおいて，システムが対話制御に失敗すると不適切な応答をユーザに返す「対話破綻」という現象が発生することがある．これはユーザとの自然な対話の実現の阻害要因となるため，速やかな破綻検出とそのリカバリが必要である．従来研究では言語，音響情報それぞれを単独で用いて対話破綻検出が行われていたが，両者を併用することで更なる性能向上が期待される．本研究では，まずシステム発話から言語情報を，システム発話直後のユーザ発話からは言語情報並びに音響情報を抽出する．そして，それらを併用した対話破綻検出器を構築し，言語，音響情報単独の対話破綻検出器との性能比較評価を行うことで，単独時と比べて併用時にどれ程の性能向上が見込めるかを検証したので報告する．

キーワード：非タスク指向対話システム，音響情報，言語情報，対話破綻

Detection of Dialogue Breakdown Using Utterances Information

MOTOKI ABE [a)]　RYOTA TOGAI　TAKASHI TSUNAKAWA
MASAFUMI NISHIDA　MASAFUMI NISHIMURA

Abstract: In chat dialogue system, there may be phenomena called "dialogue breakdown" when the system fails in dialog control and responses an inappropriate return to the user. Dialogue breakdown is obstacle to realize of natural dialogue between system and user, so it is necessary for dialogue systems to detect dialogue breakdown quickly and recover it. Conventional research has tried to detect dialog breakdown using either linguistic or acoustic information, but further improvement in performance is expected by both. In this research, we developed a dialog breakdown detector that combines information extracted from system utterances (linguistic information) and user utterances (linguistic information and acoustic information). Then we conducted comparative evaluation with conventional dialog breakdown detectors.

Keywords: Non-task-Oriented Dialogue Systems, Acoustic Information, Linguistic Information, Dialogue Breakdown

1. はじめに

近年，雑談対話を行う非タスク指向型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，従来のタスク指向型対話システムとの相違点として，対話の目的が不明瞭であることや，話題が対話中に変化することが多いことから[1]，非タスク指向型対話システムにおける対話制御は難しい．また，音声対話の場合，ユーザ発話の認識の段階で認識誤りが発生することもある[2]．これらの問題点により，対話システムがユーザに対して不適切な応答をしてしまう「対話破綻」と呼ばれる現象が発生することがある．対話破綻はユーザに違和感を抱かせるものであり，自然な対話実現の阻害要因となる．

この対話破綻への対応は対話破綻の発生を抑制できるような対話制御[3]の他に，対話破綻を検出して事後的にリカバリ処理する方法が挙げられる．これは，対話破綻は発生するものという前提で，それを発生後にカバーすることでユーザの対話破綻への違和感を軽減することを目指したアプローチである．

2015 年より Shared Task「対話破綻検出チャレンジ」が開催されている[4]．同チャレンジでは，テキストチャットで収集された雑談対話データを用いて，言語情報による対話破綻検出に取り組んでいる[5][6]．一方，我々は音声対話中の破綻検出においてはユーザ発話に含まれる音響的特徴が有効ではないかと考え，音響情報を用いた対話破綻検出器を提案した．そして，既存の言語情報を用いた対話破綻検出器と性能比較実験を行い，両者の性質が異なることを確認した[7]．このことより，両者を併用することで対話破綻検出性能のさらなる向上が期待される．

本研究では，システム発話とユーザ発話からそれぞれ言語情報，システム発話から対話行為情報，ユーザ発話から音響情報を抽出する．そしてこれらの情報を併用することにより言語もしくは音響情報だけによる従来の破綻検出機との性能を比較した．

2. 雑談音声対話データ

本研究で用いるデータとして，自動対話制御を行う対話システムと被験者との間で行われた雑談音声対話データを

静岡大学大学院総合科学技術研究科
Graduate School of Integrated Science and Technology, Shizuoka University

[a)] abe.motoki.17@shizuoka.ac.jp

Vol.2018-MUS-118 No.24
Vol.2018-SLP-120 No.24
2018/2/21

用意した[8]．この対話システムでは，システム発話から抽出される話題転換や質問といったシステム状態と，ユーザ発話から抽出される平均発話音量・発話長・有音率を用いて対話制御を行う．そして，得られた特徴量を踏まえて次の対話状態に遷移を行い，あらかじめ対話状態ごとに用意された発話文を発話する．各対話状態における発話例を表 1 に示す．今回，男子大学生 3 名との雑談対話を 44 セッション（1 セッション当たり 25～29 ターン）行い，合計 1240 発話を収集した．なお，1 セッション当たりのターン数が異なる理由として，対話中にシステムがエラーを起こしたセッションがあり，その場合はエラーが起こるまでの対話データを 1 セッション分として用いたためである．収集データの内訳を表 2 に示す．

次に，収集した雑談音声対話データに対して人手で書き起こしたテキストログと大語彙音声認識エンジン Julius のディクテーションキット（GMM ベース）を用いた音声認識で得られたテキストログを用意する[9]．なお，この対話音声データにおける Julius の文字誤り率（CER）は 62%であった．次に，書き起こしのテキストログを用いて 3 名のアノテータが対話破綻ラベルの付与を行い，コーパスを構築した．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 値分類とした（表 3 参照）．ただし，今回は問題を 2 値分類として設定したため，T を X 側に含む O : (T＋X) とし，3 名のアノテータの多数決でラベルを決定した．なお，アノテータ間のラベル付与の一致度（weighted kappa statistics）は 0.666 であった．コーパスの内訳を表 4 に示す．

表 1　システム状態と各状態における発話例
Table 1　System states and examples of utterance in each state

システム状態	発話例
話題転換	話題を変えましょう．最近ハマっている趣味は何ですか？
質問	その趣味はいつごろ好きになりましたか？
傾聴	うんうん，それで？
自己開示	私の趣味は自作パソコンです
共感	やっぱりそうですよね

表 2　収集データ内訳
Table 2　Details of collected data

被験者	セッション数	発話数
A	14	406
B	15	431
C	15	403
合計	44	1240

表 3　ラベルの付与基準
Table 3　Standard for giving labels

ラベル	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 4　コーパス内訳
Table 4　Details of corpus

ラベル	発話数（割合）
O	788（0.635）
(T＋X)	452（0.365）
合計	1240

3. 提案手法

提案手法はシステム発話から言語情報並びに対話行為情報を，ユーザ発話から言語情報並びに音響情報を抽出し，これらを統合して 1 つの学習モデルを構築するものである．以下，各抽出特徴量について述べる．

なお，これらの特徴量の値は，算出した値を直接用いるのではなく，話者ごとに標準化変量に正規化した値を用いる．

3.1 言語特徴量

3.1.1 対話行為情報

対話行為情報とは，システムが行う社会的発話の分類情報である．先行研究において，雑談対話に代表されるような，対人間関係の擁立・維持を主目的とする対話は「社会的対話」と呼ばれ，その中でも共感・自己開示に関する発話は，ユーザに親近感を抱かせるために重要であると明らかにされている[10]．我々は先に，共感や自己開示の発話行為をシステムが行った際，対話破綻が発生しやすいことを発見した[11]．これは雑談対話ならではの重要な対話行為であるが，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唆している．この結果から，対話行為情報も対話破綻検出の際にこうした情報も対話破綻検出の際に 1 つの有効な素性になるのではと考え，特徴量として採用した．

Vol.2018-MUS-118 No.24
Vol.2018-SLP-120 No.24
2018/2/21

今回用いた雑談音声対話データは，既にシステム状態ごとに発話テンプレートが用意されていたが，これは実験者の主観に基づいて分類されたこと，先行研究の分類と若干異なることから，本研究では改めてアノテータ 3 名による多数決で評価を行った．なお，アノテータ間の一致度（weighted kappa statistics）は 0.866 であった．また，３名の間で一致しなかったものが今回 4 発話あり，それらについては unknown state とした．今回用いた対話行為分類及び発話例，分類結果を表 5 に示す．

表 5　対話行為分類と各分類における発話例及び分類数
Table 5　Dialogue act classification, examples of utterance in each class and number of classes

対話行為分類	発話例	分類数
対話管理	ところで最近買って後悔したものはありますか？	213
質問	それは何故ですか？	351
自己開示	私の趣味は旅行なんですよ	418
相槌	うんうん，それで？	103
共感	それはいいですね	151
unknown class		4
合計		1240

3.1.2 発話間類似度

ユーザ発話とシステム発話の発話間類似度を，対話破綻検出の特徴量とする先行手法がある[12]．これは発話同士の類似度が低い時，システムがユーザの発話に対して，文脈に沿わない発話をした可能性が高いという仮説のもと提案された手法で，対話破綻検出においても有効であることが確認されている．そのため，この発話間類似度を今回の言語ベース破綻検出器の特徴量として採用した．

まず，テキストログに含まれる 1 ターンのユーザ発話及びシステム発話のペアから形態素解析器 Janome（使用辞書：IPA 辞書）[13]を用いて名詞抽出を行う．次に，ユーザ発話，システム発話それぞれから抽出した名詞を組み合わせて名詞ペア群を作る．そして，名詞ペアごとにペアとなっている名詞間の類似度を Word2vec（使用モデル：日本語 Wikipedia エンティティベクトル[14]）を用いて算出する．最後に，名詞ペアごとに算出された名詞間類似度の平均値を算出し，これを発話間類似度とする（図 1 参照）．

なお, ペアのユーザ発話とシステム発話のうち一方から抽出された名詞数が 0 の場合，ペアが生成できないため該当セッションの発話間類似度の平均値をその値とした. また，Wikipedia エンティティベクトルで抽出された名詞が辞書に入っておらず，名詞間類似度が算出できない場合は当該ペアを無視して処理した．

図 1　発話間類似度算出フロー
Fig. 1　Calculation flow of utterance similarity

3.1.3 語集サイズ

対話破綻が起きたとき，システム発話やユーザ発話は短くなる傾向がある．具体例を挙げると，システム発話では対話制御に失敗した場合，無意味に相槌を行ったり，ユーザ発話では対話破綻を起こしたシステムに対して呆れや侮蔑の感情から素っ気ない発話をしたり（例：ユーザ「あぁ，そうですか」「もういいよ」）といったものである．こうした場合，該当発話から抽出される名詞数は少なくなることが予想される．このことから，システム発話・ユーザ発話から抽出される名詞数が対話破綻検出に有効な素性であることが考えられるため，今回特徴量として採用する．

3.2 音響特徴量

今回用いる音響情報特徴量は，音響解析ツールキット OpenSMILE[15]を用いて抽出できる INTERSPEECH 2009 Emotion Challenge の 384 次元音響特徴量を用いた[16]．ただし，今回用意したデータセットのサイズを考慮すると 384 次元という特徴量数は過剰であり，学習が困難であることが考えられる．そこで，特徴量選択手法の 1 つであるステップワイズ変数選択法（変数選択基準：赤池情報基準量）を用いて，特徴量選択を実施した．その結果，以下表 6 に示す 9 個の音響特徴量を用いることとする．

Vol.2018-MUS-118 No.24
Vol.2018-SLP-120 No.24
2018/2/21

表 6　使用音響特徴量
Table 6　Selected acoustic features

特徴量	概要
pcm_fftMag_mfcc_sma[3]_min	MFCC3 次元目の最小値
pcm_fftMag_mfcc_sma[5]_kurtosis	MFCC5 次元目の尖度
pcm_fftMag_mfcc_sma[6]_max	MFCC6 次元目の最大値
pcm_fftMag_mfcc_sma[6]_amean	MFCC6 次元目の平均
pcm_fftMag_mfcc_sma[7]_linregc2	MFCC7 次元目のオフセット
pcm_fftMag_mfcc_sma[12]_linregc2	MFCC12 次元目のオフセット
voiceProb_sma_max	声である確率の最大値
voiceProb_sma_range	声である確率のレンジ
pcm_fftMag_mfcc_sma_de[1]_stddev	MFCC1 次元目の時間微分値の標準偏差

4. 評価実験

今回提案した言語情報と音響情報を併用した言語＋音響ベース対話破綻検出器が，言語単体及び音響単体の対話破綻検出器に対してどの程度良好に動作するかを検証した．

4.1 実験概要

実験は以下の条件で実施し，評価を行った．

識別器は SVM（多項式カーネル）を使用した．学習と評価は 10 分割交差検証で実施した．

データセットについては偏りが大きいため，O：（T＋X）＝1：1 になるように補正を行った．

評価指標は O，（T＋X）それぞれの F-Measure，全体の Accuracy，Receiver Operating Characteristic curve（ROC curve）と Area under curve（AUC）を用いた．

4.2 言語＋音響ベース対話破綻検出器の性能評価

4.2.1 評価手法

今回以下の 3 つの対話破綻検出器を構築し，書き起こしテキストログと音声認識テキストログの 2 つに対して評価を行った．

- 音響ベース対話破綻検出器（Acoustic）
  特徴量：INTERSPEECH 2009 Emotion Challenge 384 次元音響特徴量より 9 次元
- 言語ベース対話破綻検出器（Linguistic）
  特徴量：対話行為（6 次元），発話間類似度，（システム発話・ユーザ発話の）語彙サイズ…合計 9 次元
- 言語＋音響ベース対話破綻検出器（L＋A）
  特徴量：前者 2 つの特徴量を併用…合計 18 次元

4.2.2 実験結果

実験結果を表 7，表 8，図 2 に示す．これらより，書き起こしテキストログと音声認識テキストログ共通で，言語情報と音響情報を併用した場合の方がそれぞれ単体で用いる場合よりも性能が向上していることが確認できる．単体で評価した場合，言語ベース対話破綻分類器の方が音響ベース対話破綻分類器に対して破綻検出性能が優れていることが分かる．しかし，併用時には言語単体の際よりも性能が向上する他，音声認識誤りが発生しても音響情報の抽出に影響はないことから，音響ベース対話破綻分類器は比較的安定に動作するといったことが利点として挙げられ，言語ベース対話破綻検出器に補助的に音響情報を用いることで効果が得られる可能性が示された．

また，音声認識誤りを含んだデータでも破綻検出性能の低下は軽微であることが確認された．これについて詳細を分析したところ，音声認識誤りの影響を受けない対話行為情報や語彙サイズ（システム発話）の貢献度が高いことが確認された．特に対話行為情報の貢献度が大きく，これらの特徴量によって性能低下度合が抑えられたことが分かる．しかし，他の言語特徴量を用いた対話破綻検出器においては性能低下度合が異なることが予想されるため，追加検証が必要である．

表 7　書き起こしのテキストログでの対話破綻検出実験結果
Table 7　Experimental results of dialogue breakdown detection (Transcription)

検出器	F-Measure (O)	F-Measure (T+X)	Accuracy	AUC
Acoustic	0.579	0.516	0.549	0.575
Language	0.582	**0.606**	0.594	0.613
L+A	**0.613**	0.591	**0.602**	**0.652**

表 8　音声認識のテキストログでの対話破綻検出実験結果
Table 8　Experimental results of dialogue breakdown detection (Voice recognition)

検出器	F-Measure (O)	F-Measure (T+X)	Accuracy	AUC
Acoustic	0.579	0.516	0.549	0.575
Language	0.578	**0.604**	0.592	0.605
L+A	**0.615**	0.594	**0.605**	**0.640**

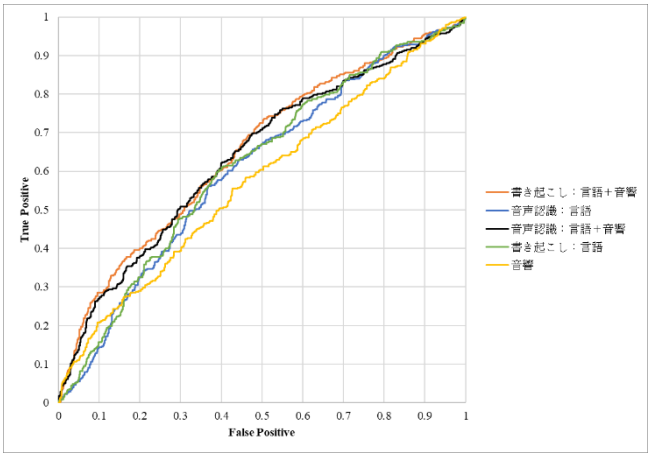

図 2　ROC 曲線

Fig. 2　Receiver Operating Characteristic curve

5. おわりに

本研究では，雑談音声対話での対話破綻検出において，システム発話・ユーザ発話から抽出される言語情報とユーザ発話から抽出される音響情報を併用することで対話破綻検出性能の改善を試みた．そして，言語情報と音響情報の併用，さらには対話行為情報などの利用によって破綻検出性能を改善できる可能性を示した．

今後の課題として，ドコモ雑談対話 API[17]など公開されている対話制御手法を用いて収集した雑談音声対話データでの評価及びデータセットの拡充，他の言語情報を用いた対話破綻検出器での検証，そして音響特徴量の追加検討が挙げられる．

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] 徳久良子，寺嶌立太，“非課題遂行対話における発話の特徴とその分析”，人工知能学会論文誌，vol.22，No.4，pp.425-435，2007.
[2] 石川開，隅田英一郎，”テキストデータを使った音声認識誤りの訂正”，自然言語処理，vol.7，No.4，pp.205-227，2000.
[3] 稲葉通将，高橋健一，”対話破綻検出による対話システムの応答性能の向上”，言語・音声理解と対話処理研究会，vol.81，pp.110-115，2017.
[4] 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.
[5] 小林颯介，海野裕也，福田昌昭，”再帰型ニューラルネットワークを用いた対話破綻検出と言語モデルのマルチタスク学習”，言語・音声理解と対話処理研究会，vol.75, pp.41-46, 2015.
[6] 堀井朋，森秀晃，林卓矢，荒木雅弘，”破綻類型情報に基づく雑談対話破綻検出”，言語・音声理解と対話処理研究会，vol.78, pp.75-80, 2016.
[7] 阿部元樹，栂井良太，狩野芳伸，綱川隆司，西田昌史，西村雅史，”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”，言語・音声理解と対話処理研究会，vol.81, pp.102-103, 2017.
[8] 栂井良太，中島悠，伊藤伸泰，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.
[9] 河原達也，李伸晃，”連続音声認識ソフトウェア Julius”，人工知能学会論文誌，vol.20，No.1，pp.41-59，2005.
[10] 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.
[11] 阿部元樹，栂井良太，綱川隆司，西田昌史，西村雅史，”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”，第 16 回情報科学技術フォーラム，E-002，2017.
[12] 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.
[13] Janome v0.3 documentation(ja) http://mocobeta.github.io/janome/ （取得日時：2018/01/25）
[14] 鈴木正敏，松田耕史，関根聡，岡崎直観，乾健太郎，” Wikipedia 記事に対する拡張固有表現ラベルの多重付与”，言語処理学会第 22 回年次大会発表論文集, pp.797-800, 2016.
[15] openSMILE by audEERINGTM http://audeering.com/technology/opensmile/ （取得日時：2018/01/25）
[16] Bjorn Schuller, Stefan Steidl, Anton Batliner, ” The INTERSPEECH 2009 Emotion Challenge”, INTERSPEECH, pp.312-315, 2009.
[17] docomo Develoler support：雑談対話 https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_name=dialogue&p_name=api_reference （取得日時：2018/01/25）