6Q-09

言語情報と音響情報を併用した対話破綻分類器の構築

阿部元樹　栂井良太　綱川隆司　西田昌史　西村雅史
静岡大学大学院総合科学技術研究科

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．

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ったアプローチ[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」といった研究課題が提唱され，多くの関連研究が行われている[2][3].

我々は先に雑談音声対話を対象とし，音響ベース対話破綻検出器を提案した．そして，既存の言語ベース対話破綻検出器との性能比較評価を行い，両者の性質が異なることを確認した[4]．このことより，両者を併用することで破綻検出性能の更なる向上が期待される．

本研究では，言語ベース特徴量・音響ベース特徴量を雑談音声対話から抽出し，両者を併用することで言語ベース・音響ベース単体と比べてどの程度対話破綻分類性能が向上するかを検証した．

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[5]．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，1240 発話を収録した．そして，人手による書き起こしログを併せて作成した．なお，学習と評価用として，3 : 2 にデータを分割する．

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 値とした（表 1 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 側に含めた O:(T+X)として，多数決で決定した．なお，学習・評価データに対してのラベル付与結果及び全体のデータセットにおけるアノテータ間の一致度（weighted κ statistics）を表 2 に示す．

表 1 ：ラベルの付与基準

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 2：学習・評価用データの O:(T+X)の内訳

データセット	学習用（割合）	評価用（割合）
O	447(0.619)	341(0.658)
(T+X)	275(0.381)	177(0.342)
weighted κ statistics	0.666	

3. 言語ベース特徴量

言語ベース特徴量として最初に，ユーザ発話とシステム発話から抽出される名詞ペア群を用いて算出される発話間距離に着目した先行研究の手法を参考にした[6]．まず書き起こしテキストログに含まれるユーザ発話及びシステム発話から形態素解析器 Janome（使用辞書：IPA 辞書）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[7]）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）．

これに加えて，対話破綻・非破綻時に特に出現する名詞 14 個の出現頻度（14 次元）と，システム発話の 1 発話から抽出される名詞数（1 次元）を用いる．

以上，16 次元を言語ベース特徴量とする．

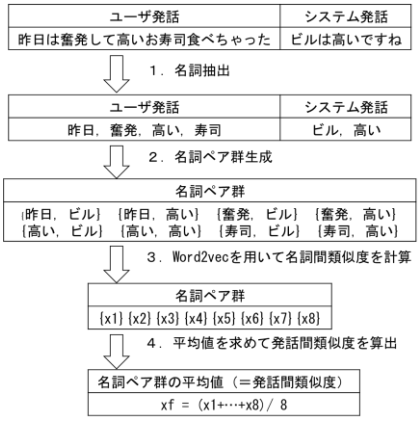

図 1：発話間類似度算出フロー

Construction of Dialogue Breakdown Classifier Using Linguistic and Acoustic Information
Motoki Abe, Ryota Togai, Takashi Tsunakawa, Masafumi Nishida, Masafumi Nishimura
Graduate School of Integrated Science and Technology, Shizuoka University

4. 音響ベース特徴量

システムが対話破綻を起こした直後のユーザ発話には，破綻に対しての困惑やとまどいといったものが音響情報として表出し，通常時と異なるゆらぎが出ることが確認されている[8]．音響ベース特徴量として，先行研究[3]で用いた特徴量を採用する．音響解析ツールキットのOpenSMILE を用いて，システム発話直後のユーザ発話に着目し，そのユーザ発話から INTERSPEECH2009 Emotion Challenge の 384 次元音響特徴量を抽出した[9].

次に，この 384 次元音響特徴量に対して学習データを用いて次元削減を行う．この理由として，今回用意したデータセットのサイズを考慮すると，384 次元は過大な次元数であり十分な学習をさせることが困難であるためである．そして，次元削減手法として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）を用いた．これによって選定された 21 次元音響特徴量を使用する．

なお，今回選定された音響特徴量を分析してみると RMS energy や F0，Voice Probability が主として重要度が高いことを確認できた．

5. 評価実験

言語ベース特徴量と音響ベース特徴量を併用することによって，対話破綻分類性能が向上するかを検証する．そのために，言語＋音響対話破綻分類器（16+21 次元）を構築し，言語単体（16 次元），音響単体（21 次元）の対話破綻分類器との性能比較評価を行った．評価指標は ROC 曲線を用いた．

今回，識別器は SVM（多項式カーネル）を用いた．学習用データについては破綻・非破綻の偏りを補正し，1:1 にした状態で学習を行った．評価データについては補正を行わずに偏りがある状態で評価を行った．

図 2 に結果を示す．図 2 を見てみると，言語対話破綻分類器が音響対話破綻分類器，言語＋音響対話破綻分類器より優れており，併用した効果が十分に出ていないことが確認できる．今回言語＋音響対話破綻分類器の性能が低かった原因として，ROC 曲線や AUC を鑑みるに，音響ベース特徴量が言語ベース特徴量に対して破綻識別性能が劣っていることが挙げられ，組み合わせ時に相乗効果が生まれなかったと考えられる．ただし，今回言語ベース特徴量は書き起こしテキストログを使用したことから，音声認識誤りを含んだものではない．音声認識誤りが生じた場合，言語ベース特徴量の有効性は低下することが予想され，その際に音響ベース特徴量が良好に動作するかの検証は今後行う必要がある．また，音響ベース特徴量についても，他の有効な音響ベース特徴量の検討を行う必要がある．

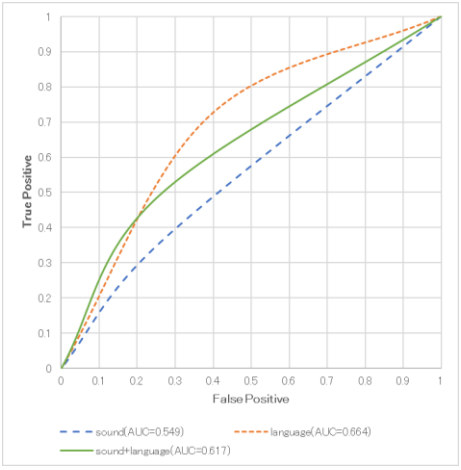

図 2：対話破綻分類実験結果（ROC 曲線）

6. おわりに

本研究では，言語ベース特徴量と音響ベース特徴量を併用することによる対話破綻分類器の性能改善の有効性の検証を行った．今後の課題として，音声認識誤りが生じた場合の対話破綻分類性能の検証と，音響ベース特徴量の更なる検討，言語＋音響の組み合わせ手法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] 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] 稲葉通将，高橋 健一，“RNN エンコーダによる文脈を考慮した対話破綻検出”，言語・音声理解と対話処理研究会，Vol.78，pp.98-101，(2016).
[4] 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).
[5] 栂井良太，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] 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).
[7] 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).
[8] 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).
[9] Bjorn Schuller, Stefan Steidl, Anton Batliner: The INTERSPEECH 2009 Emotion Challenge, INTERSPEECH, pp.312-315, (2009).