各対話エラー破綻類型の特徴を考慮した対話破綻検出システムの提案

初田 玲音

2021 年 11 月 8 日　中間発表

1　研究背景

現在研究が進められている雑談対話システム技術の 1 つとして, 対話破綻検出技術がある．対話破綻検出とは, システムの対話破綻させる発話を検出する技術であり, その競技会である対話破綻検出チャレンジ (Dialogue Breakdown Detection Challenge: DBDC) も 2015 年から開催されている [1].

対話破綻検出では,「破綻ではない (o), 破綻 (x), 違和感あり (t)」の評価のみを行うが, これを発展させた研究として対話破綻エラー類型分類がある [2]．対話破綻エラー類型分類では, 2 つの軸で整理された 8 グループに, 17 種類に分類された破綻エラー類型に基づき, 発話がどの破綻エラー類型で破綻したかをマルチラベルに分類する．対話破綻エラー類型分類技術は, 任意の対話システムに対し, どのような破綻をする傾向にあるか等, 対話システムの評価指標として用いることが出来るため, 重要な技術だと言える．しかし, 破綻エラー類型のマルチラベル分類に関する研究は私の知るところ存在しない.

そこで, 本研究では, 各破綻エラー類型が属するグループの特性を考慮した特徴を用意することで, 各グループの推定精度を改善し, それによって全体のグループ推定精度を向上させる手法を提案する.

2　関連研究

堀井ら [3] は, Project Next NLP の日本語対話タスクグループによる雑談対話の破綻原因類型化案に基づき, その類型毎に破綻識別器を作成し, それらを組み合わせる手法を提案した．破綻の原因を類型化した点は同様だが, 出力については「破綻ではない (o), 破綻 (x), 違和感あり (t)」の 3 種のみに留まり, どの破綻類型で破綻するかは考慮していない.

社会性に関する関連研究として, Khatri ら [4] は, 攻撃的な内容を含む発話を検出する為に, 大規模データをブートストラップする 2 段階の半教師付きアプローチを提案した．ブートストラップモデルによって, 対話システムにおける, 攻撃的な発話を高い精度で検出した結果から, ブートストラップモデルは様々なドメインに対して一般化を行うことを示した.

3　研究構想

対話破綻エラー類型は, 全部で 17 種類存在し, 2 つの軸によって整理され, 8 つのグループに分けられる．本研究では, 破綻した発話を, それぞれの破綻エラー類型が属する 8 グループにマルチラベル分類するタスクを対象とする．　本節では, 各グループのベースラインと提案手法について述べる.

4　今後の課題

- 各グループの特徴, 特性を調べる
- どのような特徴が有用か決定する方法を検討する
- 外部コーパスが必要なグループには, コーパスを用意する
- 実験方法, 評価方法について検討する

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. 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-75-B502, pp. 27-32(2015)

[2] Ryuichiro Higashinaka, Masahiro Araki, Hiroshi Tsukahara and Masahiro Mizukami：Integrated taxonomy of errors in chat-oriented dialogue systems, SIGDIAL 2021.

[3] 堀井朋, 森秀晃, 林卓也, 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

[4] Khatri C, Hedayatnia B, Goel R, Venkatesh A, Gabriel R, Mandal A. Detecting Offensive Content in Open-domain Conversations using Two

Stage Semi-supervision. In: Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS); 2018. p. 1–9.