

3.17 対話システムにおけるエラー分析

—雑談対話コーパスの収集と対話破綻検出—

東中 竜一郎 (NTTメディアインテリジェンス研究所)

船越 孝太郎 ((株) ホンダ・リサーチ・インスティテュート・ジャパン)

対話システム

対話システムとは言葉でやりとりを行うシステムのことであり¹⁾。一言で対話システムといってもさまざまである。たとえば、テキストのみでやりとりをするシステムもあれば音声を用いるシステムもある。ジェスチャや表情を用いて対話をするものもある。対話の目的も、タスク指向のもの（フライト予約といった特定の課題の遂行を目的とするもの）や、非タスク指向のもの（いわゆる雑談）もある。2010年ごろまではタスク指向型対話システムの研究が主に行われていたが、近年はPepperやJiboなどの会話機能付きのロボットや、Siriやしゃべってコンシェルのようなスマートフォンの音声アシスタント機能などが身近になるにつれ、会話すること自体を楽しむ雑談が注目されつつある。

対話システムにおける課題

対話システムは人工知能の究極の目標の1つとされており、技術的に難しいということはもちろんだが、そのほかの課題として、多くの研究者が連携し、システムを改善していくことが難しいという問題がある。なぜかというところ、対話システムの性能は、うまく対話できたかによって評価されるが、その「うまく」の定義はそれぞれのシステムの利用目的によって異なり、また、システム内部で利用される技術もシステムごとにまちまちであるため、システム内部に踏み込んだエラー分析をシステム・研究者横断で行うことが困難だからである。

対話破綻に基づくエラー分析

Project Next NLP 対話タスクの営みとして、対話システム自体をブラックボックスと考え、外から観

察可能な対話の表層（発話文字列）のみに着目してエラー分析を行うことにした。分析対象としては「対話破綻」に着目した。対話破綻とは、ユーザが対話を続けることができなくなるような状態を指し、対話システムにおいてよく見られる現象、かつ、対話システム研究者であれば誰しもが避けたい現象である²⁾。表層のみに着目し、また、どのようなシステムでも避けるべき現象である対話破綻に着目することで、多くの対話研究者が共同でエラー分析を行うことが可能となる。今回、我々の呼びかけに対し15の組織、合計32名の研究者が賛同し、普段研究しているシステムの垣根を越えた対話システムのエラー分析が実現した。

エラー分析を行うにあたり、対話データが必要となる。しかし、当初はこの取り組みに必要なデータがまだ整備されていなかった。そこで、まず人・システムの対話データを収集した。扱う対話の種類は近年注目されている雑談（テキストチャット）とし、対話データの収集には、NTTドコモが一般公開している雑談対話API^{☆1}を利用して実装された対話システムを用いた。そして、得られた対話データ（雑談対話コーパスと呼ぶ）に対し、対話破綻個所を複数のアノテータによりラベル付けした。最後に、そのラベル付けされた対話破綻個所を観察し、どういったシステムのエラーによってその対話破綻が引き起こされたかを研究者間で議論・整理し、エラーの分析および類型化を行った。

今回作成したエラーの類型を表-1に示す。まず大分類として4つの分類がある。発話、応答、文脈、環境である。それぞれ、発話単体として不適切、直

.....
^{☆1} https://www.nttdocomo.co.jp/service/developer/smart_phone/analysis/chat/

大分類	小分類	類型名	説明
発話	構文制約違反	構文的誤り	日本語の文として正しくない
	意味制約違反	意味的誤り	文としての意味が通らない
	不適切発話	解釈不能	著しく応答の体をなしていない
応答	量の公準違反	情報過不足	応答として内容の過不足がある
	質の公準違反	不理解	直前のユーザ発話を理解していない
	関係の公準違反	無関係	直前のユーザ発話の話題や発話意図と無関係
	様態の公準違反	意図不明	発話の意図がくみとれない
	誤解	誤解	直前のユーザ発話の内容を誤って理解している
文脈	量の公準違反	不要情報	冗長な発話
	質の公準違反	矛盾	発話内容や態度が急転換する発話
	関係の公準違反	無関係話題	話題が文脈から逸脱
	様態の公準違反	関連性不明	文脈のどの部分と関連しているのか不明
	話題展開への不追随	不追随	話題展開後も前の話題を続けている
環境	無根拠	共通基盤欠如	根拠のない主張
	矛盾	一般常識欠如	常識に反する主張
	非常識	社会性欠如	社会規範から外れ、相手を不快にする発話

表-1 エラーの類型

前発話の応答として不適切、直前発話の応答としては問題ないが対話全体に照らせば不適切、対話文脈より広い範囲を見たときに不適切という分類である。そして、大分類の下に小分類として全部で16のエラーがある³⁾。表における類型名は我々が分かりやすさのため考案したエラーの名前である。実際の対話においてエラーを分類した例を以下に挙げる。ここで、Uは人間、Sがシステムの発話である。

【類型名：無関係】

U: それでは趣味はなんですか?

S: 最後に旅行されたのはいつですか?

【類型名：矛盾】

S: 買い物は1人が楽です

U: 確かに気が楽ですね

S: 買い物は長いです

U: まあ見るのも楽しいし

S: 買い物は一緒に楽しいですね

もちろん、今回考案した類型は完全というわけではない。しかしながら、雑談を行うシステムにおいて、このようなエラーがあるということを対話研究者が共有できた意義は大きい。また、直近の営みとして、対話破綻箇所について、どのようなエラーで引き起こされたと考えられるかをコメントとして記述してもらい、これらのコメントを自動分類する手法によるエラー分析も行っている⁴⁾。この分析から、理解しづらい発話、ユーザ質問の無視、発話意図が不明といったエラーが

特に対話破綻に関係することも分かってきている。

対話破綻はなくなるか

Project Next NLP 対話タスクの営みによって対処すべきエラーが明確化され、システムの改善が進むことを期待している。また、我々は雑談対話コーパスを公開し、そのデータを用いた対話破綻検出の評価型ワークショップ^{☆2}をオーガナイズしている⁵⁾。対話

システムが自身の発話の前にその発話が対話破綻につながりそうと判断できれば、対話破綻は未然に防げるようになる可能性がある。ただ、人間同士でも話が続けにくいときがあるように、対話破綻を完全になくすることはできないだろう。対話破綻からうまくリカバリしていく技術も併せて研究していく必要がある。

参考文献

- 1) 奥村 学 監修, 中野幹生, 船越孝太郎, 中野有紀子, 駒谷和範 著: 対話システム, コロナ社 (2015).
- 2) Martinovsky, B. and Traum, D.: The Error is the Clue: Breakdown in Human-Machine Interaction, In Proc. ISCA Workshop on Error Handling in Spoken Dialogue Systems, pp.11-16 (2003).
- 3) 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: Project Next NLP 対話タスク: 雑談対話データの収集と対話破綻アノテーションおよびその類型化, 言語処理学会第21回年次大会ワークショップ (2015).
- 4) Higashinaka, R., Mizukami, M., Funakoshi, K., Araki, M., Tsukahara, H. and Kobayashi, Y.: Fatal or Not? Finding Errors that Lead to Dialogue Breakdowns in Chat-oriented Dialogue Systems, In Proc. EMNLP, pp.2243-2248 (2015).
- 5) 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6回対話システムシンポジウム (2015). (2015年10月1日受付)

☆2 <https://sites.google.com/site/dialoguebreakdownndetection/>

東中竜一郎 (正会員) higashinaka.ryuichiro@lab.ntt.co.jp
慶應義塾大学政策・メディア研究科修士。博士 (学術)。NTTメディアインテリジェンス研究所にて勤務。入社以来、質問応答システム・対話システムの研究に従事。しゃべってコンシエルの質問応答機能の研究開発に携わる。

船越孝太郎 (正会員) funakoshi@jp.honda-ri.com
東京工業大学大学院情報理工学専攻修士。博士 (工学)。2006年より (株) ホンダ・リサーチ・インスティテュート・ジャパンにて勤務。対話システムに関する研究に従事。