

会話によるニュース記事伝達のための 発話意図分類とデータベースの構築

Utterance Intention Classification and Database Construction for Conversational News Delivery System

横山 勝矢^{*1} 高津 弘明^{*1} 本田 裕^{*2} 藤江 真也^{*1*3} 林 良彦^{*1}
Katsuya Yokoyama Hiroaki Takatsu Hiroshi Honda Shinya Fujie Yoshihiko Hayashi

小林 哲則^{*1}
Tetsunori Kobayashi

^{*1}早稲田大学 ^{*2}本田技術研究所 ^{*3}千葉工業大学
Waseda University Honda R&D Co.,Ltd. Chiba Institute of Technology

In a human-human conversation, a listener frequently conveys various intentions explicitly/implicitly to the speaker with reflexive short responses. The speaker recognizes the intentions of these feedbacks and changes the utterance plans to make the communication more smooth and efficient. These functions are expected to be useful for human-system conversation also, but when human face the system they do not always give the same feedback as they did with human. We investigated the feedback phenomena of human users against our system which is designed to transfer a massive amount of information like news articles by spoken dialogue. First, user's intentions of feedbacks that can affect the behavior of the system were classified from the viewpoint of roles in the progress of conversation: presence/absence of a user's interest, comprehension state, release/keep turn, and so on. Then, using our news delivery conversation system, we gathered user's short responses, and labeled the intention through listening test by the subjects. We also attempted automatic identification of those intentions. As a result, in the current system, it is found that there are many feedbacks to convey doubt and automatic identification of them is possible stably, but there are not so many other feedbacks.

1. はじめに

実会話システム利用時における、システム発話に対するユーザのフィードバックの現象を分析するためのコーパス収集について検討する。

人同士の会話において、聞き手は、話し手の発話に対し、しばしば反射的な応答を行う。これらの応答は、発話の理解状態や発話内容への興味の程度など、聞き手の様々な状態を、話し手にフィードバックする。話し手は、これらのフィードバックをもとに、相手の理解が乏しければ情報を補足し、興味がなさそうなら話題を変えるなど、自らの発話計画を適宜修正し、聞き手の状態に合わせた発話を行うことになる。

こうした一連の処理を、人とシステムとの会話にも適用することで、従来の一問一答の枠組みを超えた、円滑で利便性の高い会話システムを実現できることが期待されている [Fujie 06, Kobayashi 13]。ここで、聞き手からのフィードバックは、必ずしも言語的に明示的な形でのみ表現されるとは限らず、抑揚や声色などに表れるパラ言語を伴った複雑な形で表現される [益永 14]。このため、こうした機能の実現には、自然なフィードバックを含む大規模な会話音声コーパスが必要となる。

会話における聞き手からのフィードバックを扱うためのコーパスの研究には、既に様々なものがある。牧野ら [牧野 17] は、一人一人の会話における話者交代の現象を分析対象として、話者交代の際に起こる韻律情報と視線情報を調査するためのコーパスを作成した。荒木ら [荒木 17] は、雑談における会話内容に対するユーザの興味の有無を分析対象として、MMDAgentを用い、WoZ (Wizard of Oz) 方式により対話データの収集を行った。井上ら [井上 17] は、エンゲージメント (興味や意

欲による対話継続欲求度) を分析対象として、自立型アンドロイド ERICA [Inoue 16] を用い、WoZ 方式により、会話時の相槌、笑い、うなずき、視線などのデータを収集した。

これら、一人一人の会話や、WoZ 方式のように、応答時に人が介入して収集したコーパスには、極めて自然なフィードバックが含まれ、研究対象としては興味深い。しかし、現状、実際の会話システムを使用したときのユーザの振る舞いとは乖離している可能性があり、システムへ組み込む機能を実現するための学習データとしては必ずしも適切ではない。

このような観点から、実際の自律システムとの対話を通じてデータ収録した研究もある。長澤ら [長澤 17] は、インタビューを行うシステムにおいて、ユーザの反応によって話題を掘り下げるか、話題を転換するかなどの対話戦略を決めるために、実際にロボットを用いた対話を行ってデータを収録し、ユーザの発話への積極性を 5 段階でアノテーションしている。

本研究においても、長澤らの研究と同様に、実システムの利用時におけるユーザの応答を調査対象とすることを試みる。現在我々が作成しているニュース伝達のための会話システム [高津 18a] を対象として、ユーザが実際にシステムとの会話時にどのようなフィードバックを行い、システムはどのようにそれを利用するかを検討する。システムからの提示情報に対するユーザの興味の有無、発話交代に関する働きかけ (発話権を取りたいか、継続したいかなど)、システム発話の理解状態など、ユーザがフィードバックの発話に込めた、システムの振る舞いに影響を与えうる情報を広くとりあげ、これらの機能を整理・分類するとともに、これらが、現状のシステムとの会話におけるユーザの応答にどの程度含まれるかを調査する。

以下本稿では、2. でユーザが会話時に行うフィードバックの分類について説明し、3. でデータセットの収集方法やアノテーションについてとその結果について述べる。

2. 発話意図分類

システム会話におけるユーザのフィードバック現象を分析するコーパスの収集に先駆け、フィードバックがシステムの振る舞いに影響を与える情報を広くとりあげ、これらの機能を整理・分類した。

フィードバックの役割は、システムにある種の要求を伝えることにあがるが、これらは、ユーザの要求を直接明示的に伝えるものと、ユーザ状態等を伝達することで間接的、あるいは暗示的に伝えるものがある。本稿ではこれらをまとめて発話意図(intention)と呼ぶ。

円滑なシステムを実現するための機能として、人の発話とシステムの発話が同時に起きるべきではないと考え(発話の衝突回避)、また、利便性の高いシステムを実現するための機能として、システムが伝える情報の量という観点から、その増減を行うフィードバックがあることを考えた。これらのフィードバックを分類し、以下の6つの発話意図を提案する。

- ・待機要求
- ・質問
- ・補足要求
- ・反復要求
- ・無関心
- ・既知

「待機要求」はユーザがある発話の後少し間を空けても話し続けることを示し、システムの発話は待ってほしいという意図である。従来の音声対話システムではVAD(Voice Activity Detection)により発話の区切りが検出されたら次の発話を開始するというものであった。そのためユーザはまだ発話を継続するつもりにも関わらず、システムが発話を開始し発話の衝突が起こっていた。この「待機要求」を認識することでシステムは発話を待機し、衝突を陽に回避することができると考えられる。

「質問」「補足要求」「反復要求」は伝達する情報を増やすための意図である。まとまった量の情報を伝達するシステムにおいてユーザの理解状態に応じて情報量の調整を行う必要があるが、中でもユーザの疑問解消や十分な情報の伝達は不可欠である。そのため、これらの意図を認識することで伝達する情報量を増やす調整を行うことができると期待した。「質問」はユーザが疑問をもち陽に質問を行ったもので、対話行為で分類される「質問」と同様のものである。「補足要求」は「質問」のように陽に質問は行っていないが、システム発話に疑問点を感じているもしくはシステム発話に理解が追いついていないため補足説明がほしいという意図である。「反復要求」はシステム発話を聞き取れなかった場合や受け取れなかった場合にもう一度同じ発話を繰り返してほしいという意図である。

「無関心」「既知」は、伝達する情報を減らすための意図である。興味のないトピックや既に詳細を知っている情報を長々と伝えるシステムは利便性のあるものではない。そういった情報をユーザの反応に応じて省くことで利便性があり、継続して利用したくなるようなシステムになると考えられる。「無関心」は対話を行うトピックに対してユーザが興味を持っていないという意図である。ユーザの興味の無い情報を伝達することは無駄だと考え、この意図を認識した場合は他のトピックへ遷移することが考えられる。「既知」はシステムが伝える情報がすでにユーザが有する情報であることを示す意図である。既知情報を含むトピックに関しては詳細な情報を省き伝達することで、伝達効率を上げることができると考えられる。



図 1: 収録環境

3. データセット構築

2. で提案した発話意図に関するラベルを付与したデータセットの構築について述べる。発話意図のあらわれ方は、人同士で対話した場合と、システムと対話した場合とは異なることが考えられる。本研究ではシステムと対話した際の現象をとらえるため、実際に[高津 18a]のシステムを用い、24人の被験者が対話を行った際の音声データを収集した。さらに、収集したデータを10人のアノテータにより2. で提案した発話意図に関するアノテーションを行った。

3.1 音声収録方法

人がシステムと対話した際の現象をとらえるため、実際に[高津 18a]のシステムを用い、被験者が対話を行った際の音声データを収集した。現在2103対話分の収録が完了している。音声データ収集には24人(男:12人, 女:12人)の被験者がシステムとの対話を行い、その際の音声データを随時収録した。

3.2 アノテーション

アノテーションは、図2に示すアノテーションツールを作成し、10人のアノテータで行った。このツールで用いるラベルは2. で提案した発話意図を示すことができるかどうかと、このツール自体の使いやすさを考慮し設計した。具体的には、ラベルを現象、行為、内部状態の3つのレイヤに分け内部状態には5つの軸を設定した。

現象では、「発話の完了」を示す終了、「発話が途切れていたたり、遮られている」ことを示す継続、「音声のない区間」を示す無音・雑音、「システム対話と関係のない音声」を示す無効、の4ラベルを付与する。

行為では、発話行為として分類される相槌、ターンをとる、返答、質問、情報開示、リアクションの6ラベルについて当てはまるものを選択する。

内部状態では「システム発話を受け取れているかどうか」「システム発話の内容に疑問を感じているかどうか」「対話内容に興味を抱いているかどうか」「システム発話の内容に好意的かどうか」「システム発話の内容が予期しえたかどうか」の5つの軸についてそれぞれ3段階で評価を行う。



図 2: 作成したアノテーションツール

表 1: 各ラベルに対応する発話意図とそれに対して想定されるシステムアクション

ラベル	発話意図	システムアクション
質問	質問	質問応答
継続	待機要求	傾聴
ターンをとる		
意外	補足要求	補足説明
相槌		
or	& 不受理	繰り返す
リアクション	& 無関心	他トピックへ移行
当然	既知	詳細説明省略

アノテーションのルールは、1. 現象のレイヤで「終了」のラベルを付けた場合には行為、内部状態のレイヤに関しても必ずラベルを付与する。2. 「継続」のラベルを付けた場合には分かれれば行為、内部状態のレイヤに関してラベルを付与する。3. 「無音・雑音」「無効」のラベルを付けた場合には行為、内部状態のレイヤにはラベルを付与しない。とした。

各レイヤでの選択は独立に行われ、内部状態のレイヤでは軸ごとに独立に評価が行われる。各ラベルに対応するユーザー意図とそれに対して想定されるシステムアクションを表 1 に示す。

対象音声の分割は WebRTC vad (aggressiveness=2) *1 を用い、即時的な反応を対象とするため 1.5 秒以下の音声を対象とした。また [森 05] において前の文脈が対象発話の評価に影響を及ぼすことが示されている事を考慮し、評価の際の材料として「対象音声の 3 秒前から再生」「対象音声の 5 秒前から再生」を作成した。さらに 5 秒前の音声からのみでは評価できなかった場合を考慮して、「対話全体」を聞けるようにし、対象発話が全体の中で何秒から始まるかの記述も行った。

他にも、アノテータ間の意見の相違をなくすためアノテーションの試用期間を設け、不一致が多かったものに対する意見のすり合わせ等も行った。

アノテーションを行う際には、アノテータ同士の一致度を求めるため、200 発話分に対しては全員がラベルを付与し、そ

表 2: 各ラベルの数

ラベル名		数	ラベル名		数
現象	終了	9240	内部状態	受理	8511
	継続	271		不受理	396
	無音・雑音	930		納得	3426
	無効	215		懷疑	2337
行為	相槌	3671		興味	2287
	ターンをとる	88		無関心	1156
	返答	233		当然	301
	質問	2161		意外	1594
	情報開示	366		好感	844
	リアクション	4950		嫌悪	542

表 3: 各発話意図に対応するラベルとそのデータ数

ラベル		発話意図	データ数		
質問		質問	2161		
継続		待機要求	312		
ターンをとる					
意外		補足要求	2055		
相槌	& 懷疑				
or	& 不受理			反復要求	192
リアクション	& 無関心			無関心	1145
当然		既知	301		

の一致度を求めた。一致度の計算にはフライスのカッパ係数 (Fleiss' Kappa) を用い、「現象」「行為」ではそれぞれのラベルを選択したか、「内部状態」では各軸毎にどのラベルを選択したかで一致度を求めた。

3.3 結果と考察

データセットの構築により得られた結果について述べる。現在、10420 発話に対するラベル付与が済んでおり、集まったデータについて各ラベルの数は表 2 の通りである。この数とは、アノテータが重複している場合は一人でも付与しているラベルがあればカウントしたものである。

この中から表 1 で示した発話意図の分類に当てはまるラベルの数は表 3 の通りである。また、アノテータ全員がラベルを付与した 200 発話に対する一致度 (Fleiss' Kappa) は図 3 のようになった。

提案した音声対話によりまとまった情報を伝えるシステムにおいて理解すべきユーザーの発話意図において、最も表出度が高かったものは「質問」が 2161 発話、次いで「補足要求」が 2055 発話、「無関心」が 1145 発話であり、他の意図に関しては表出度合いが低かった。また、一致度についてみると現象や行為のラベルについては κ 係数が約 0.8 とほぼ完全な一致に近い値であったが、内部状態のラベルについてはどれも 0.4 を下回り、アノテータの評価一致度は低かった。

内部状態のラベル数は、「システム発話を受け取れているかどうか」の軸に関するラベルの付いた発話数が 8907 と「終了」のラベルが付いた発話のほぼすべてにラベルがついていたが、「システム発話の内容に疑問を感じているかどうか」に関するラベルの付いた発話数は 5763、「対話内容に興味を抱いている

*1 <https://github.com/wiseman/py-webrtcvad>

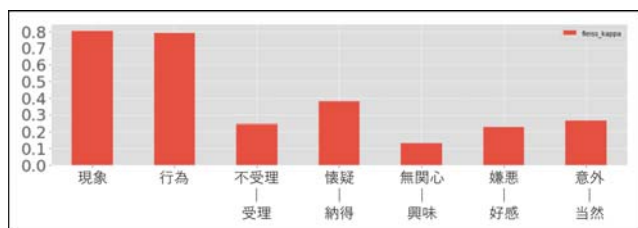


図 3: アノテータ同士のラベル一致度 (Fleiss' Kappa)

かどうか」に関するラベルが付いた発話数は 3443, 「システム発話の内容に好意的かどうか」に関するラベルが付いた発話数が 1386, 「システム発話の内容が予期しえたかどうか」に関するラベルが付いた発話数が 1895 と「システム発話を受け取れているかどうか」の軸以外は内部状態のラベルの付いた発話が少なくなることがわかる。このことから、現状のシステムとの対話において話者はそれほど内部状態の表出を行っていないことが分かる。

使用したシステムのニュースを伝えるという特性上, 「質問」「補足説明」といった発話意図の表出が増えることは容易に想像でき, それは表出度合にも表れている。中でも「質問」に関してはラベルの一致度も高いため, システム実運用においても有用な意図であり, データセットとしての質も高く学習などに用いることができると考えられる。しかし, 「補足要求」についてはその構成するラベルの一致度が低いためシステム実運用の際に判断のブレが強く出てくることが考えられる。「無関心」については, 話者の内部状態表出が少ないことから, 棒読みに聞こえることが多く, 第三者からすると無関心に聞こえることが多かったのではないかと考えられる。また, ラベルの一致度については「補足説明」と同様のことが言える。他の意図に関しては表出度合が低く, 構成するラベルの一致度も低い。

このように話者のフィードバックが少ない対話では, 従来感情認識や意図認識として行われてきた研究の成果が生かせず, 効率的な対話の実現ができない。また, 雑談対話のようなタスクにおいても, システムの対話戦略をユーザの意図に沿わずに決定することになり, 継続的に使われるシステムの実現は難しい。そのため, ユーザが内部状態を表出するためには現状のシステムで改善するべきものは何か, どのような機能を加えるべきなのか, そもそもシステムという時点で内部状態は表出しなのか, といった検討が必要である。

また, 今回作成したコーパスの中でもデータ量が多く, アノテータ間の一致度も高かった発話意図「質問」については各発話が「質問」かどうかの識別実験も行った [高津 18b]。この結果, 識別率約 95 % と非常に高い識別性能を得ることができた。

4. おわりに

実会話システム利用時における, システム発話に対するユーザのフィードバックの現象を分析するためのコーパス収集について検討を行った。現在我々が作成しているニュース伝達のための会話システム [高津 18a] を用いて, ユーザが実際にシステムとの会話時にどのようなフィードバックを行い, システムはどのようにそれを利用しうるかを調査した結果, 現状のシステムでは発話意図「質問」においては出現頻度は高く, アノテータ間の一致度も高いことが分かった。一方で, その他の発話意図についてフィードバックは陽に得られないことが分かった。今後の課題として話者のフィードバックを陽に得られるよ

うなシステムの作成があげられる。

参考文献

- [Fujie 06] Shinya Fujie, Riho Miyake, Tetsunori Kobayashi: Spoken Dialogue System Using Recognition of User's Feedback for Rhythmic Dialogue, in Proceedings of Speech Prosody (2006)
- [Kobayashi 13] Tetsunori Kobayashi, Shinya Fujie: Conversational Robots: An Approach to conversation protocol issues that utilizes the paralinguistic information available in a robot-human setting, Acoustical Science and Technology, Vol. 34, pp64-72 (2013)
- [益永 14] 益永祐吾, 川端豪: 発話意図判定における文型・韻律・音色の相補的效果, 情報処理学会研究報告音声言語情報処理 (SLP), Vol.2010-SLP-80, No.14 (2014)
- [牧野 17] 牧野孝史, 三浦寛也, 竹川佳成, 平田圭二: 複数人対話における話者交替現象分析のためのコーパスの作成, 人工知能学会言語・音声理解と対話処理研究会 (SIG-SLUD), pp.3-8 (2017)
- [荒木 17] 荒木雅弘, 富増紗也華, 中野幹生, 駒谷和範, 岡田将吾, 藤江真也, 杉山弘晃: マルチモーダル対話データの収集と興味判定アノテーションの分析, 人工知能学会言語・音声理解と対話処理研究会 (SIG-SLUD), pp.20-25 (2017)
- [井上 17] 井上昂治, Divesh Lala Pierrick Milhorat, 高梨克也, 河原達也: 潜在キャラクタモデルによるリアルタイム対話エンゲージメント推定, 人工知能学会言語・音声理解と対話処理研究会 (SIG-SLUD), pp.78-83 (2017)
- [Inoue 16] K. Inoue, P. Milhorat, T. Zhao, T. Kawahara: Talking with ERICA, an autonomous android, in Proceedings of SIGDIAL Conference, pp.212-215 (2016)
- [長澤 17] 長澤史記, 石原卓弥, 岡田将吾, 新田克己: ユーザーの態度推定に基づき適応的なインタビューを行うロボット対話システムの開発, 人工知能学会言語・音声理解と対話処理研究会, pp.72-77 (2017)
- [高津 18a] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 意図性の異なる多様な情報行動を可能とする音声対話システム, 人工知能学会論文誌, Vol.33, No.1 (2018)
- [森 05] 森大毅, 相澤宏, 粕谷英樹: 対話音声のパラ言語情報ラベリングの安定性, 日本音響学会誌 61 巻 12 号, pp.690-697 (2005)
- [高津 18b] 高津弘明, 横山勝矢, 本田裕, 藤江真也, 林良彦, 小林哲則: 会話によるニュース記事伝達のための発話意図理解, 第 32 回人工知能学会全国大会論文集 (2018)