

コンピュータとの自然な会話を実現する雑談対話技術

人とコンピュータが自然に会話できる対話システムをめざし、NTTメディアインテリジェンス研究所の技術協力のもと、雑談対話システムを開発した。本対話システムの特長は、対話における話題と文脈を正しく認識し、大規模データから返答を作成・選択することにより、人間どうしの会話に近い応答を行えることである。本システムは、スマート家電への適用や家庭用ロボットの対話機能として利用するなどの応用が期待される。

サービス&ソリューション開発部 おおにし かなこ よしむら たけし
大西 可奈子 吉村 健

1. まえがき

近年、しゃべってコンシェルのような音声認識対応エージェントが普及してきた。しゃべってコンシェルは、「メールを送る」や「電話をかける」などのような何かしらの作業目的をもった発話から、「富士山の高さは？」や「世界一高い山は？」などのような質問、さらに、「好きだよ」や「こんにちは」などのような雑談に対しても返答が可能な音声エージェントである。ユーザーにとって作業目的の発話や情報要求目的の発話が行えることは非常に便利であるが、しゃべってコンシェルは便利さを享受するためだけに使用されているわけではなく、日々多くの雑談も投げかけられている。ここから、ユーザーの雑談欲求が非常に高いことがうかがえる。しかし、しゃべって

コンシェルは、想定内の発話に対しては的確な返答が返ってくるものの、バリエーションがまだ十分ではない。ユーザーの雑談欲求を満たすことができれば、雑談対話そのものを人気コンテンツとしてユーザーに提供したり、ロボットやゲーム機、車載器等の新領域デバイスの通信モジュール搭載を拡大するなど、さまざまなビジネス応用が考えられる。

このようなユーザーの雑談欲求にさらに応えるためドコモでは、NTTメディアインテリジェンス研究所の成果技術に基づき、雑談対話システムを開発した。この雑談対話システムは、コンピュータと人間が自然に会話できることを目的にしたもので、大規模データから作成した発話データを利用し豊富な返答が可能なシステムである。また本システムは、ユーザーと単発の発話を繰り返すのでは

なく、複数回のやりとりを可能としている。本稿では雑談対話システムの概要と対話技術について解説する。

2. 雑談対話システム概要

開発した雑談対話システムの概要を述べる。ここで“雑談対話”とは、例えば「長野への行き方を教えて」のような特定の情報（この例では道案内）を要求する発話ではなく、「長野へ行きたいなあ」のような明確な目的のない発話によって構成される対話のことを指す。

図1は開発した雑談対話システムを用いたアプリの例である。左側がシステムの発話、右側がユーザーの発話となっている。例えば、「これから長野に行きます」というユーザーの発言に対して、システムは「長野羨ましい。」と返答している。このよ

うな雑談に対して柔軟で的確な返答を行う一般的な手法は知られておらず、本システムは対話システムとして、学術的にも非常に大きな意味をもっている。

2.1 システム特長

本システムの特長は、ユーザの発話とその発話に対する返答データとの単純なマッチングによって返答を行うシステムではなく、入力されたユーザの発話内容を解析し、発話の意図や文脈を理解したうえで、大規模データに基づいて返答を行うことである。文脈を理解できることから、人間どうしの会話のように文脈をもった自然な対話を行うことができる。

2.2 システム動作

図2に示すシステム概要図を基にして、システムの動作を簡単に説明する。本システムは大きく分けて6つのパートから成る。この6つのパートとは、①発話理解部、②対話制御部、③発話タイプ分類部、④システム発話生成部、⑤システム発話選択部、⑥出力前変換部である。

ユーザから「長野に行きたい」という発話を受け取ったときの処理を説明する。

①発話理解部

発話理解部では、入力発話を解析することにより、現在ユーザが話している話題を「長野」とであると認識する。

②対話制御部

対話制御部では、ユーザ発話の対話行為を推定し、これまで

の対話の履歴を基に、次にどのような返答を行うべきかを、“ポジティブな発言をする”、“ネガティブな発言をする”などという大まかな枠組みの中で推定する。例えば「長野に行きたい」に対して、共感する発言をする、とシステムが推定した場合、「いいですね」と返答するかもしれない。

③発話タイプ分類部

発話タイプ分類部ではユーザの発話がどのようなものを大雑把に判定するため、ユーザの発話を“雑談対話”“質問応答システム向けの質問”“システム自身への質問”の3タイプに分類する。例えば、「長野に行きたい」は“雑談対話”，「富士山の高さは？」は“質問応答シ



図1 雑談対話システムを用いたアプリの例 (左:システム, 右:ユーザ)

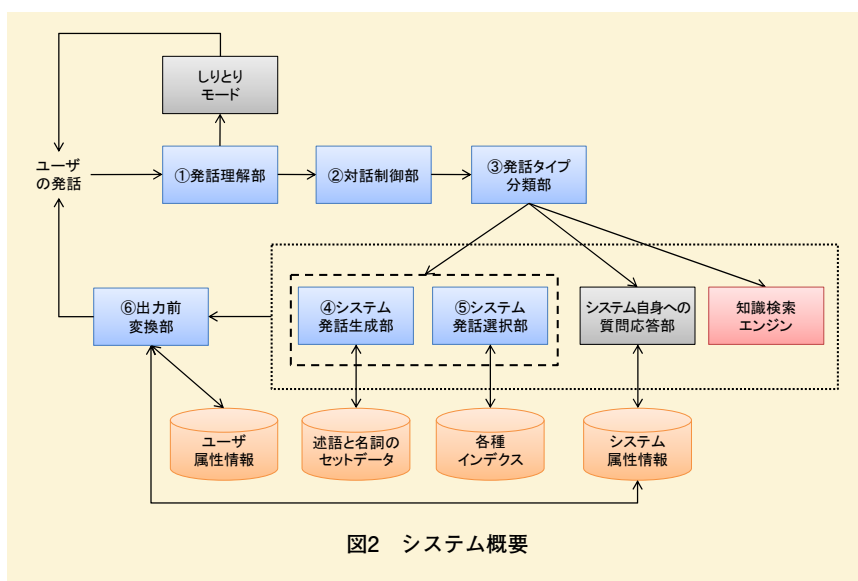


図2 システム概要

システム向けの質問”，「あなたの名前は」は“システム自身への質問”と分類される。発話が“質問応答システム向けの質問”である場合は，外部の知識検索エンジン[1] [2]を用いて返答する。発話が“システム自身への質問”である場合は，システム自身の属性情報を参照し「20歳です」などと答える。ユーザの発話が“雑談対話”である場合は，後述するシステム発話生成部へ進み，返答を行う。

④システム発話生成部

システム発話生成部では，システムがあらかじめ保持している知識を利用し，返答を生成する。例えば“長野”に対しては，「修学旅行で長野に行く」「温泉が多い」「空気が美味しい」のような知識を保有しており，それらを利用して「温泉が多いですよ」などの発話を生成する。

⑤システム発話選択部

発話は生成するだけでなく，大規模発話データから選択することによっても可能である。本対話システムでは人手で作成したものやインターネットから取得したものなど合わせて大規模な発話データベースを保持しており，その中から現在の返答として最も適切と考えられるものを選択し出力する。

⑥出力前変換部

より自然な対話を行うため，出力前変換部でシステム返答の語尾変換を行う。例えば「明日

はいい天気になるのか」という文に対して，システムに女性というキャラクタを付与したい場合は「明日はいい天気になるかしら」と変換することができる。これにより，システムに対して，一貫した個性を付与できる。

また，本対話システムはゲームなど，雑談対話以外のモードへ入ることもできる。これはユーザ発話の内容を解析し，自動でモードを切り替える。現在はモードとして，しりとりゲームモードがある。

3. 対話技術

本対話システムにおいて，最も重要な対話技術である，焦点認識，対話制御，発話生成について図3に示す例を用いて解説する。

3.1 焦点認識

対話を行うために重要なことは，現在ユーザが何について話しているのかを理解することである。話題（ここでは焦点と呼ぶ）は対話中

継続し，またあるタイミングで遷移する。この継続と遷移を認識するために，本対話システムでは主に次の2つの解析を行っている。

(1)焦点の抽出

1つは焦点の抽出である。例で，ユーザは「そこは良かったですか」という発言まではオランダの話をしているが，次に「はい，とても」と答えたシステムを遮り，続いてカメラの話を始めている。ここでシステムは焦点が“オランダ”から“カメラ”に移ったことを認識できなければならない。この問題を解決するために，本対話システムでは発話から焦点としてふさわしい単語を推定し，抽出する。焦点の抽出は機械学習*1によって実現されている。

(2)照応解析

もう1つは，照応解析*2である。例では，システムの「オランダでチーズを食べたよ」という発言に対して，ユーザが「美味しいよね」と答えている。ここでシステムはユーザの直近の発話だけを解析したのでは，何が美味しかったのか認識できない。

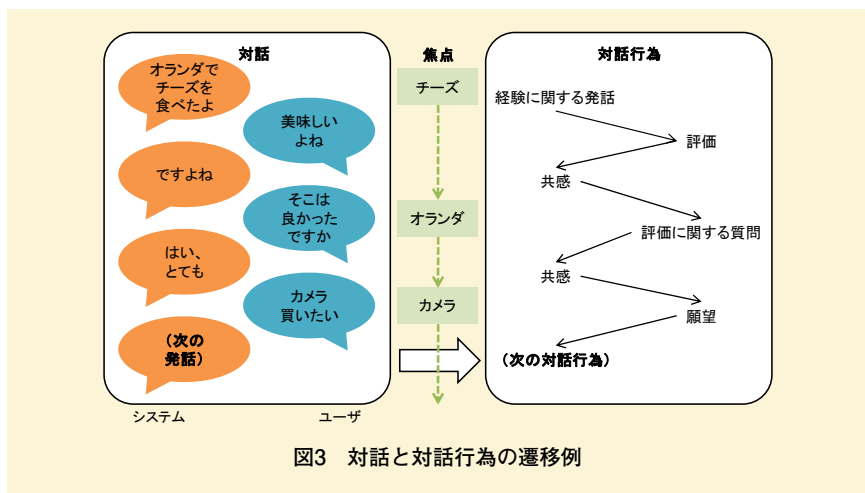


図3 対話と対話行為の遷移例

*1 機械学習：サンプルデータから統計処理により，有用な判断基準をコンピュータに学習させる枠組み。
 *2 照応解析：代名詞や指示詞，省略された名詞句の参照先を同定すること。

この問題を解決するために、本対話システムでは照応解析を行い、隠れた主語や目的語を補う。

具体的には、ユーザの発話に含まれる述語に対して必須とされている要素がない場合に、文内に現れている名詞句、もしくは、これまでに焦点として抽出された名詞句を使用し、足りない要素を補完する。例えば「美味しいよね」という発話に対しては、“何が”という要素が必要である。システムはこれまでに焦点と推定した名詞句から、“何が”を補うのに妥当であるものを推定する。ここでは“チーズ”が妥当であると推定される。また、“それ”や“あれ”のような代名詞も、同様にそれらが指し示す名詞句により補完する。例えば「そこは良かったですか」という発話に対しては、“そこ”が指し示す名詞句として、これまでに焦点と推定した名詞句の中の“オランダ”が妥当であると推定する。

3.2 対話制御

次に、ユーザ発話の対話行為を推定し、次に発話すべき対話行為を推定する対話制御について説明する。通常、人は、直前の発話だけでなく、これまでの流れから次に話すことを考えて返答している。本対話システムでは、これまでの対話の流れを考慮するため、まずユーザの発話の対話行為を数十種類の対話行為の中から推定する[2]。対話行為には例えば、“共感”や“評価”などがある。発話を対話行為に変換するための分類器*3は、機械学習によって学習し作

成する。学習データは、大量の対話行為が付与された発話データを用いた。

対話は、対話行為の連続によって成り立っていると看做する。図3は対話と対話行為の遷移を示している。ここで次にシステムは、これまでの対話行為の遷移を考慮し、次に発話すべき対話行為を推定する。この推定を行うため、機械学習によって予測器*4を構築した。具体的には、現時点のユーザ発話の対話行為や過去のユーザおよびシステムの対話行為を特徴量*5とし、次に出現しやすい対話行為を推定する。“評価を問う質問”“共感”“願望”という流れから、例えば次に発話すべき対話行為は“事実を問う質問”と推定し、「どんなカメラですか」というような発話を行う。

このシステムの次に出力すべき対話行為の出現確率を変更することにより、システムに性格付けをすることも可能である。例えば“良い評価である意見”を出力しやすくしておけば、例における次の対話行為を、“事実を問う質問”ではなく“良い評価である意見”と推測するだろう。これによりシステムは、「どんなカメラですか」ではなく「カメラ楽しいですよ」という“カメラ”に対する“良い評価である意見”を発話するようになる。これによりシステムにポジティブな性格付けを行うことができる。また、反対に“悪い評価である意見”を出力しやすくしておけばネガティブな性格付けを行うことができる。

3.3 発話生成

システムが発話を生成する手法について説明する。ここでは現在の焦点と次に発話すべき対話行為を基に、システムの発話を生成する。発話の生成には“何をどうした”などのような述語と名詞がセットになったデータを用いる[3]。データの例を表1に示す。

このような述語と名詞のセットは、インターネットのテキストを解析することにより構築されており、それぞれ出現頻度を保持している。ここで頻度が一定以上出現しているものは、テキスト上に頻繁に出現している事柄であり、すなわち人間にとっては常識的な情報であると見なすことができる。これらのデータを利用することで、人間の常識に基づく応答が可能となる。例えば焦点が“パン”の場合、表1に示すようなデータの中から発話生成のために必要なデータが選択される。

次に、発話候補として選択された“何をどうした”のデータから平叙文*6を作成する。例えば、“述語：食べる、目的語：美味しいパン”からは「美味しいパンを食べる」という平叙文が作成される。

続いて、作成された複数の平叙文から、最も現在発話するのに適した

表1 “パン”に関する“何をどうした”のデータ

述語	名詞
食べる	美味しいパン
作る	パン
焼く	メロンパン

*3 分類器：入力をもとに、その特徴量に基づいて予め定められた分類先のいずれかに分類する装置。

*4 予測器：与えられた入力をもとに、次に出現するものを推測する装置。

*5 特徴量：データから抽出される、そのデータを特徴づける値のこと。

*6 平叙文：疑問文や命令文、感嘆文でない主語・動詞を含む通常の文。

一文を選択するため、直近の発話に関する類似度を計算する。これにより、ある発話候補の文がこれまでの文脈と意味的なまとまりをなすかどうかを確かめ、話が飛躍しないような発話文を選択することができる。

最後に、平叙文を対話行為に即した発話へと変換する。例えば「美味しいパンを食べる」という文章は、対話行為「願望」では「美味しいパンを食べたい」と変換し、対話行為“共感”では「美味しいパンを食べたいですね」と変換することになる。この変換は、平叙文の文末表現を切り出し、該当箇所を各対話行為に特徴的な文末表現に置き換えることで実現している。また、同様の技術を用い、文末をさらに関西弁などといった特徴的な語尾に変換することも実現している。これによりシステムの個性を表現することができる。

4. エンターテインメント分野への応用

本対話システムは前節までに説明したような通常の雑談に加えて、エンターテインメント性の高い機能としてユーザとしりとりゲームを行うことができる。ユーザの発話を解析し、それがしりとりゲームの開始を示唆するものだった場合、システムは対話からしりとりが始まったと認識し、しりとりのルールに基づいて返答を行う。返答は、しりとり回答用の単語リストから選択された複数の語の中から発話される。これらの単語には頻度情報に基づいた優先度が付与されており、一般語ほど優先度が高

く、システムの返答として発話されやすい。これにより、「りんご→ごりら」といった一般的なしりとりの応答から始まり、後半では一般的でない語による上級者向けの応答が可能となる。最終的にはユーザが負けるか、システムが回答できる単語がなくなった場合、またはゲームが進めば一定の確率でシステムが負けることによりゲームは終了する。ゲームが終了した後は自動的に対話へと切り替わり、対話を継続することができる。

5. あとがき

人間とコンピュータが自然な対話を行うことを目的とした雑談対話エンジンについて解説した。本システムの特長は話題と文脈を理解し柔軟な返答を行えることである。

この雑談対話システムは、ドコモ・イノベーションビレッジ[5]や docomo Developer support サイト*7 を通して提供している。また、ドライブネットインフォ*8における雑談機能としても展開している。ユーザは、これらのAPIやアプリケーションを通して、多様なパリエーションをもつ雑談対話を楽しむことができる。その先の展開としては、ロボットやゲーム機器、テレビ、車載機などの携帯端末以外のデバイスとも接続し、多様なシーン・目的で雑談対話を楽しめるようにしたい。サービスの面からも本対話技術にはさまざまな可能性がある。ドコモのめざすべき姿であるお客様のスマートライフのパートナーとなるために必要とされる技術として、雑談対話は欠か

せないものであると考えている。例えば、1人暮らしの方の話し相手として、またユーザのことを最も理解してくれるパートナーとして、本対話システムを携帯電話だけでなく家庭用ロボットの対話機能などに適用することによって、雑談対話システムの可能性は無限に広がると期待している。

今後は実際に使用されたログを解析し、より人間らしい対話ができるシステムへさらに改良を継続する予定である。また学術的なチャレンジとして、ユーザが対話で人間とシステムを見分けることができるかを問うチューリングテスト*9にもチャレンジし、真に人間と同等の会話を実現する対話システムの実現をめざしたい。

文献

- [1] 内田, ほか: “自然文質問への直接回答を実現する知識Q&A,” 本誌, Vol.20, No.4, pp.6-11, Jan. 2013.
- [2] 東中 竜一郎, 貞光 九月, 内田 渉, 吉村 健: “しゃべってコンシェルにおける質問応答技術,” NTT技術ジャーナル, Vol.25, No.2, pp.56-59, Feb. 2013.
- [3] 目黒 豊美, 東中 竜一郎, 堂坂 浩二, 南 泰浩: “聞き役対話の分析および分析に基づいた対話制御部の構築,” 情報処理学会, Vol.53 No.12, pp.2787-2801, 2012.
- [4] H. Sugiyama, T. Meguro, R. Higashinaka and Y. Minami: “Open-domain Utterance Generation for Conversational Dialogue Systems using Web-scale Dependency Structures,” SIGDIAL, pp.334-338, 2013.
- [5] NTTドコモ: “ドコモ・イノベーションビレッジによる起業支援,” 本誌, Vol.21, No.3, pp.29-31, Oct. 2013.

*7 docomo Developer support サイト: スマートフォン向けサービスやアプリケーションの開発者が、ドコモの保有するAPIを利用できるAPIの提供サイト。

*8 ドライブネットインフォ: スマートフォンに話しかけるだけで渋滞情報や周辺情報を

教えてくれる新たなドライバー向け情報提供サービス。ドコモの商標または登録商標。

*9 チューリングテスト: 判定者の人間が別の人間とコンピュータに対して会話をを行い、話している相手が人間なのかコンピュータなのかを判定するテスト。