小特集
Special Feature

［身近になった対話システム］

1 対話システムを知ろう
―自然言語による機械と人間とのコミュニケーション―

基 般

中野幹生　（株）C4A 研究所

いろいろな対話システム

近年「対話システム」という言葉を目にする機会が増えてきたのではないだろうか？　対話システムの研究は 50 年以上も前から行われてきており，カーナビゲーションの音声インタフェースなどの製品も 1990 年代から発売されているが，一般的に使われるようになったのは，2011 年に Apple の iPhone で Siri が利用可能になったころではないかと思われる．現在では，対話ロボットや AI スピーカ，チャットボットなど，さまざまな対話システムの製品が販売されたりサービスが提供されたりしている（図 -1）．

対話システムにはさまざまなタイプがあり，いくつかの観点から分類できるが（表 -1），共通しているのは，人間と言葉で情報をやりとりするというこ

■図 -1　対話システムの例：左から LINE（株）の AI スピーカ CLOVA Desk，（株）ミクシィの対話ロボット Romi，（株）Nextremer・東急（株）・東急電鉄（株）・（株）Huber. の AI 案内サイネージ．写真・画像は各社提供．

■表 -1　対話システムの分類

分類の軸	分類	説明
タスクの有無・種類	タスク指向型	対話によって遂行すべきタスク（情報検索，説明，インタビュー，説得，交渉，クイズ，ゲームなど）が明確なシステム．
	非タスク指向型	タスクが明確ではないシステム．雑談システムや傾聴システムなど．タスクは遂行しないが，ユーザから情報を取得したり，ユーザの感情を良くしたりするといった目的はある．
	混合型	タスクが明確な対話と明確でない対話の両方を行う．
話題の範囲（ドメイン）	クローズドドメイン	天気，レストラン，旅行など，話題を限定することで，深い対話を行うシステム．多くの場合，バックエンドに構造化データベースを持つ．
	オープンドメイン	話題を限定しない対話を行う．
入出力のモダリティ	ユニモーダル	入出力がテキストのみまたは音声のみ．
	マルチモーダル	テキスト・音声以外の入出力を用いる．入力は，画像，ポインティング，その他のセンサ，出力は画像，CG エージェント，ロボットなど．
対話参加者の数	1 対 1	ユーザ 1 人とシステム 1 体での対話のみを行う．
	マルチパーティ	複数のユーザと同時に対話できる．または複数のシステムが 1 人または複数のユーザと対話する．

© 2021 Information Processing Society of Japan

とである[1]．これによって，人がデータベースや機器の操作をしたり，対話自体を楽しんだりすることができるようになる．

現状の対話システムはどのような対話ができるのだろうか？　図 -2 に NTT コミュニケーション科学基礎研究所が開発したオープンドメインのテキスト雑談対話システム[☆1] と人間との対話を示す．このシステムは，大量の対話データから学習したモデルを用いて，文脈にあった対話を生成することができる[2]．

また図 -3 には，大阪大学駒谷研究室の武田龍氏らが構築した対話ロボットの対話例を示す．このシステムは旅行先を薦めて選んでもらうというタスクを行うタスク指向型のマルチモーダル対話システムであり，対話によってユーザの情報や意向を聞きながらタスクを達成する．このシステムは，ユーザ発話の音声理解がうまくいかなくても対話が進むようにデザインされたルールを用いている．これらの例のように，ある程度の長さの自然な対話が行えるよ

☆1　対話システムのタイプについては表 -1 を参照されたい．

S:　こんにちは。よろしくお願いします。早速ですけど、何か趣味ってありますか？
U:　趣味じゃないんだけど、アイクリームを探してて。資生堂の化粧品がいいんだって
S:　資生堂の化粧品はいいですよね。わたしもアイクリーム使っています。どんなアイクリームがいいですか?
U:　友達がつかってるやつが、6 万もするんだって
S:　6 万?!それは買えないです
U:　だよね。効果あってもなかなかかえないよね
S:　そんなにするんですか！ それはなかなか手が出せないですね。 コスメがお好きなんですか?
U:　あんまり化粧はしないけど、基礎化粧はしないとって思ってる。子どもたちがおばさんってうるさいし
S:　お肌に合う合わないもありますからね。 基礎は大切ですよね!
U:　うん。子どもの若さには勝てないけど。子どもなんて、かき氷ばっかり食べてるのに、おなかも冷えないし
S:　なるほど。 かき氷ばかり食べてるとおなか壊すんですか?
U:　じゃない？若いと壊さないの？
S:　確かに！ 若さには勝てませんね。 お腹は壊さないですね。
U:　そうなんだ。すごい
S:　ふわふわしたのばかり食べてると壊すって聞いたんですけど、そうでもないんですかね?
U:　へー。ふわふわだといいの？しらなかった。安いからだめなのか
S:　安いからっていうのはあるかもしれないですね。 油っこいものは胃もたれするっていいますよね。

■図 -2　対話システムシンポジウムで開催された対話システムライブコンペティション 3（https://dialog-system-live-competition.github.io/dslc3/index.html）のオープントラック（話題を限定しない対話を行う）で最優秀賞を獲得したシステムの対話例の一部．S はシステム，U はユーザの発話．

S:　こんにちは，いらっしゃいませ。（お辞儀）本日はお忙しいなか，ご来店ありがとうございます。（軽くお辞儀）
U:　こんにちは
S:　わたくし，お客様の観光をご案内させて頂く，ショーコと申します。（軽くお辞儀）お客様はこれまで，アンドロイドと会話したことはございますか？
U:　ありません
S:　ないですよね。実物を見るのも初めてだと思います。
U:　はい
S:　少し緊張もあると思いますので，
U:　はい
S:　いくつか簡単な質問をさせて頂きますね。
U:　はい
S:　まずは，お客様のお名前を伺ってもよろしいでしょうか？
U:　はい ○○△△です
S:　○○さま，でよろしいですね？
U:　はいはい
S:　ありがとうございます。（軽くお辞儀）○○さまは観光の際，自家用車か公共交通機関，どちらを主にご利用される予定ですか？
U:　公共交通機関です
S:　電車などの公共交通機関をご利用の予定ですね。それでもスムーズに観光できると思います。
U:　はい
S:　観光は，お一人様で
U:　はい
S:　もしくは，ご家族やご友人とご一緒に，と考えておられますか？
U:　えーとひとりです
S:　お一人様ですね。承知しました。
U:　質問へのご回答，ありがとうございました。（軽くお辞儀）
S:　はい
U:　事前にお答え頂いた情報によりますと，
S:　はい
U:　○○さまは，箕面大滝と，国立民族学博物館に，興味をお持ちですね。
S:　こちら，ディスプレイ左側の写真が，箕面大滝，
U:　はい
S:　右側の写真が，国立民族学博物館になります。二か所とも良い雰囲気ですね。（ユーザの顔を見る）パッと見て、右か左の写真，どちらの印象が良いですか？

■図 -3　対話ロボットコンペティション（https://sites.google.com/view/crobotcompetition/home）の風景と予備予選で最も体験者から評価の高かった対話例の一部．写真は対話ロボットコンペティション実行委員会提供．

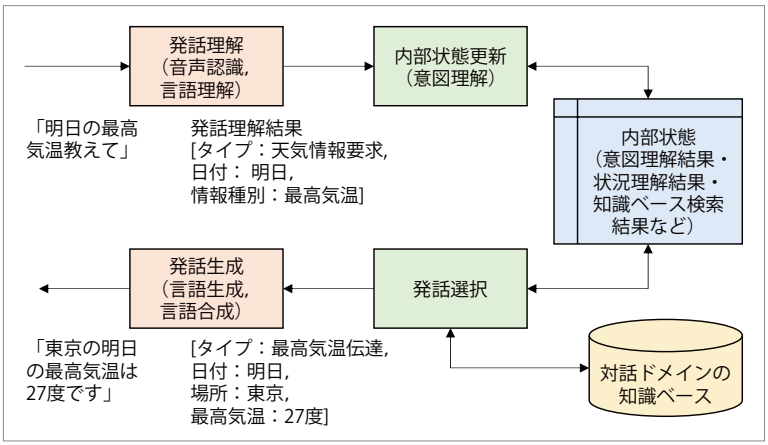
© 2021 Information Processing Society of Japan

うになってきている．

対話システムの構成

対話システムはさまざまな要素技術を統合して構築する必要があるため，システム開発において重要なのは，開発しようとしているシステムに適したアーキテクチャを選択することである．図 -4 に対話システムの一般的なアーキテクチャを示す．ユーザ発話が入力されると，発話理解部が意味を表す表現に変換し，それをもとに内部状態更新部が内部状態を更新する．内部状態には，ユーザの意図理解結果や知識ベース検索結果などが蓄えられている．この内部状態をもとに発話選択部がシステムの発話の内容を決定して発話生成部に送ると，発話生成部が文の形でユーザに提示する．図の例で，システムは，ユーザ発話にはない「東京」の最高気温を答えているが，これは，対話の履歴や位置情報などから推定して決めることができたと想定している．

システムのタイプによって，このアーキテクチャに変更が加えられる．画像入力の結果からユーザの感情を推定して応答を変更するようなマルチモーダル対話システムでは，画像認識や感情推定のモジュールが加わる．また，オープンドメインのテキスト雑談対話システムでは，全体を 1 つのニューラルネットワークで構成しているものもある[3]．



■図 -4　対話システムの一般的な構成

良い対話システムとは？

ところで，どのような対話システムが良い対話システムなのだろうか？　筆者は，以下の値が大きいシステムが良いシステムであると考えている．

＜ユーザの利益・満足度＞
＋＜システム販売者・運用者の利益＞
－＜システム構築コスト＞
－＜システム運用コスト＞

これらの各項やその重みは，作ろうとしているシステムやビジネスモデルに応じて変わってくる．

ユーザの利益・満足度は，ユーザが何かの予約ができたり情報を得られたりしたかどうか，すなわちタスクが達成できたかどうかや，対話によって良い感情を得られたかどうかなどで測る．これを大きくするには，単に対話の言語的な内容からユーザの意図を理解して応答するだけではなく，ユーザの感情やパーソナリティ，ユーザの置かれた状況などを認識し，それに合わせて応答したり説明したりすることが必要になる．音声対話システムやマルチモーダル対話システムならば，音声や表情などから得られる情報を利用することも有効である．

システム販売者・運用者の利益は，ビジネスモデルによってさまざまである．コールセンタの自動化など人が行っている仕事をシステムで置き換えることでコストを削減する場合，対話システムを組み込んだ製品を売ったり使用契約を結んだりする場合，対話によって何かの商品を販売する場合，ユーザから情報を得て広告を表示するなど他のビジネスにつなげる場合などがあり得る．

システムの構築コストも重要な要因である．近年では，Google Cloud Dialogflow[☆2]，Amazon Lex[☆3]，NTT ドコ

☆2 https://cloud.google.com/dialogflow
☆3 https://aws.amazon.com/jp/lex/

© 2021 Information Processing Society of Japan

モの AI エージェント API[☆4] など，対話システムを Web 上で簡単に構築できるようにするサービスや，Rasa[☆5] や Botpress[☆6] などのオープンソースの対話システム構築ツールが利用可能である．

システムの運用コストも無視できない．性能が同じならばハードウェアのコストが低い方がよい．また，対話システムは一度作ってしまえば終わりではなく，ユーザとの対話のログを分析しながら，モデルや知識ベースを改良していく必要がある．このようなシステムの継続的改良にかかるコストを減らす仕組みも重要である．

対話システムの利点は何か？

対話は人間にとって最も基本的なコミュニケーションの様式であり，人間と同じような対話ができるシステムがあれば非常に役に立つことは間違いない．しかしながら，現状の対話システム技術では人間のような対話ができるわけではないし，近い将来も難しいだろう．まだ不完全な対話しか行えない対話システムの存在価値は何だろうか？　人間と情報を授受するシステムはほかにも数多くあるが，それらに対して，対話システムにはどのような利点があるのだろうか？

メニューから選ぶグラフィカルユーザインタフェースなどと比べた場合の対話システムの利点は，ユーザがメニューのどれを選ぶべきか分からない場合や，メニューのどれでもないと思った場合でも，思いついた言葉で入力することができることである．その入力から答えるべき内容をシステムが決定できる場合もあれば，回答を絞り込むために聞き返す場合もある．このように曖昧な要求でも複数のやりとりで理解できるのが 1 つの利点である．また，その対話ではユーザの要求が理解できなかったとしても，対話ログをもとにシステムを改良することもできる．

また，音声入出力を用いるシステムは，車の運転中など，手がふさがっている場合やディスプレイに注意を向けられない場合でも使うことができる．

さらに，バーチャルエージェントを用いた研究から，人は，自律的に動作するシステムと対話しているときの方が，人が操作しているシステムと対話しているときよりもより個人的な情報を開示しやすいということが知られている[4)]．この知見を活かし，人間のインタビュアーやカウンセラーには話しにくいことを話してもらうことも考えられる．

このように対話システムにはさまざまな利点があり，それをうまく活かしたアプリケーションの開発が重要である．

対話システムに関する誤解

関連分野の研究開発者にとって対話システムは比較的イメージしやすいアプリケーションであることから，対話システムに興味を持つ人は多い．しかしながら，対話システム技術に関する誤解があると思われる場合も多い．以下では，よくある誤解の解消を試みる．

対話システムは自然言語処理の応用？

対話システムは自然言語でユーザとコミュニケーションするので，自然言語処理の技術を用いる．特に，発話の理解や生成については，テキストを対象とした自然言語処理の技術と共通点が多い．しかしながら，対話システムは基本的にはヒューマンコンピュータインタラクションを実現するシステムであり，そのために自然言語処理と人工知能（知識表現・意図推定・意思決定など）の技術を用いるものである（図 -5）．自然言語処理技術の向上だけで良い対話システムができるわけではない．発話理解がうまくいかない場合でもユーザの印象を悪くすることな

☆4 https://www.nttdocomo.co.jp/biz/service/ai_agent_api/
☆5 https://rasa.com/open-source/
☆6 https://botpress.com/

© 2021 Information Processing Society of Japan

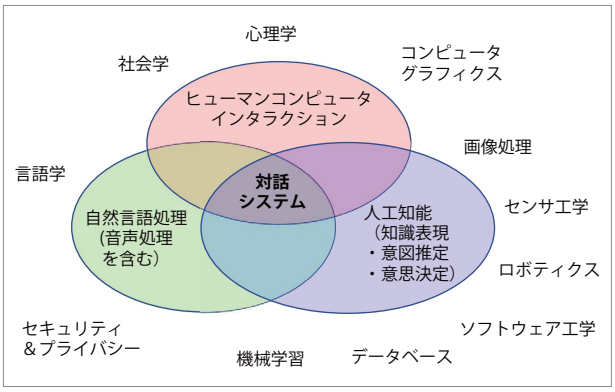
く対話を進めることや，発話のタイミングの決定など，一般的に自然言語処理では扱わない技術も重要である．

対話システムは評価法が定まっていないのが問題？

機械学習を用いた人工知能技術の多くの研究では，音声認識率や画像認識率のような，コミュニティで合意がとれた客観的な評価指標があり，それを最大化することが研究の目標となる．しかしながら，対話システムはユーザインタフェースの一種なので，主観評価を行うことも必要である．また，要素技術というよりは製品や実サービスに近いため，さまざまな観点から技術を評価する必要があり，研究段階でも必ずしもみんなが同じ評価指標を使う必要はない．実際，対話システムのいろいろなサブ課題に応じて評価法が考案され用いられている．なお，対話システム研究が，人のように振舞うシステムを目指していると誤解されることもあるが，役に立ちさえすれば，人間と同じように振舞う必要はない．また，すべての人間が常に適切な対話を行うわけではないので，必ずしもシステムのお手本になるとは限らない．

人間同士の対話のデータがあれば対話システムが作れる？

対話システム研究は，人間同士の対話のデータから学習したモデルを与えると動作する汎用対話エンジンのようなものを作っていると考えられている場合がある．実際，先に紹介したNTTコミュニケーション科学基礎研究所の雑談システムのように，人間同士の対話のデータから対話システムを構築する手法が研究されている．しかしながら，あらゆるタイプの対話システムでそのような学習を行うためには，さまざまな話題・タスクで，対話相手に合わせて適切なことを話している対話データが必要である．さらに対話相手の表情・姿勢や，時間・季節・天候・場所など周囲の状況によっても対話は変化するので，十分なバリエーションのあるデータを収集する必要がある．そのコストを考えると，対話の理論などを参考にしながら手作業でルールや対話シナリオを書いた方が良い場合も多い．実際，実用的なシステムの開発では，多くの場面でルールやシナリオの記述が行われており，そのような作業を効率的にするための技術や知見の蓄積も対話システム研究の重要な課題である．



■図-5　対話システムに関係する研究分野

対話システムは言葉の意味を理解していない？

機械が言葉の意味を理解するとはどういうことかに関しては哲学的な議論が必要だが，最近多く研究されているオープンドメインのテキスト雑談対話システムなどでは，言語表現を何かの意味空間にマッピングしているわけではないので，意味を理解していないと言ってもよい．これに対し，多くのクローズドドメインのタスク指向型対話システム，たとえば，レストラン検索システムなどは，ユーザ発話中のキーワードをもとに知識ベースを検索するが，「イタリアン」「イタリア料理」など，異なる言語表現を知識ベースの同じ値や属性にマッピングする．また，「高い」「安い」などの表現も具体的な値の範囲にマッピングしている．これは，知識ベースの値の空間を意味空間と考えると，それに言語表現を（いわゆるシンボルグラウンディングと同様の意味で）グラウンドしていると言える．このグラウンディングをもって対話システムが「言葉の意味を理解して

© 2021 Information Processing Society of Japan

いる」と言ってもよいのではないだろうか．

対話システムの今後

この 10 年は，対話システムが一般に普及した時期だと言える．この間，さまざまな製品・サービスが作られてきたが，現状の技術での可能性を試すフェーズはそろそろ一段落つくのではないだろうか．今後は，上述したような，対話システムならではの利点を活かしたシステムにフォーカスが当たっていくと考えられる．

今後の対話システム開発で重要になってくるのが倫理的・社会的側面だと考えている[5]．2016 年 3 月に Microsoft のチャットボット Tay が社会的に不適切な発話を行うようになり停止することになった．このような不適切な発話を避けるためのシステム構築技術やテスト方法などが必要だろう．また，先に，人間は自律的に動作する対話システムに対して個人的な情報を開示しやすいと述べたが，これはユーザの意に反して個人的な情報を収集し，悪用することにもつながる．悪意を持ったシステム開発をどう防ぐかは重要な課題である．

逆に社会的にポジティブなインパクトのある対話システムも重要である．たとえば，南カリフォルニア大学で開発され米国ホロコースト博物館に設置されているシステム "New Dimensions in Testimony" は，ホロコーストの生存者へのインタビュー映像を用いて，ユーザの質問に答えることができる（図 -6）[6]．このシステムによって，生存者から体験談を直接聞くことができなくても，インタラクティブに歴史を学ぶことが可能になった．ほかにも，教育，医療，福祉，行政などの分野への対話システムの適用が試みられており，社会の役に立つ対話システムの発展が期待できる．

■図 -6　南カリフォルニア大学のシステム "New Dimensions in Testimony"[6]

最後に，対話システムの，情報技術の教育や普及への応用可能性に触れたい．対話システムはさまざまな技術や知見を総合して構築する必要があるため，システムを作ったり評価したりするときに，さまざまな情報技術を用いる．たとえば，ほとんどの対話システムで必要となるヒューマンコンピュータインタラクション，アルゴリズム，自然言語処理，人工知能，機械学習，データベース，ソフトウェア工学の技術に加え，システムのタイプによっては，音声処理，画像処理，ネットワーク，セキュリティなどの技術に実践的に触れることができるだろう．

対話システムは，情報技術の専門家だけではなく，心理学，言語学，社会学，ロボット工学などの関連分野の専門家，さらには，対話システムが扱うドメインの専門家と協力し，その知見を活かして研究開発を進めていく必要がある．このプロセスが，さまざまな分野の方に情報技術をもっと知っていただく機会になるのではないかと期待している．

参考文献

1）中野幹生，駒谷和範，船越孝太郎，中野有紀子（著），奥村学（監修）：対話システム，コロナ社（2015）．
2）杉山弘晃，成松宏美，水上雅博，有本庸浩，千葉祐弥，目黒豊美，中嶋秀治：Transformer encoder-decoder モデルによる趣味雑談システムの構築，人工知能学会 SLUD 研究会 B5（02），pp.104-109（2020）．
3）東中竜一郎：AI の雑談力，KADOKAWA（2021）．
4）Lucas, G. M., Gratch, J., King, A. and Morency, L.-P. : It's Only a Computer : Virtual Humans Increase Willingness to Disclose, Computers in Human Behavior, 37, pp.94-100（2014）．
5）東中竜一郎：対話システムと倫理，人工知能 31（5），pp.626-627（2016）．
6）トラウム デービッド：双方向型歴史学習の支援のための対話システム技術の活用，人工知能，31（6），pp. 806-812（2016）．

（2021 年 7 月 9 日受付）

■中野幹生（正会員）　mikio.nakano@c4a.jp

1990 年 東京大学大学院理学系研究科修士課程修了．NTT，HRI-JP を経て，2020 年（株）C4A 研究所を設立，現在同社代表取締役，大阪大学招へい教授，名古屋大学特任教授．博士（理学）．

© 2021 Information Processing Society of Japan