機械が会話のパートナーになる日——大規模深層学習で拓く雑談対話システムの新境地

近年，人との自然な雑談を通して人の対話欲求を充足させることを目的とする，雑談対話システムに注目が集まっています．NTTでは以前より雑談対話システムの研究を続けており，複数の対話ロボットを有機的に連動させ対話のスムーズさ・快適さを向上させる研究など，特色豊かな研究を幅広く進めています．本稿では，最新の深層学習モデルを利用した高性能な雑談対話システムについて，NTTが構築したシステムの詳細や現状の到達点・課題を解説します．

杉山 弘晃（すぎやま ひろあき）　水上 雅博（みずかみ まさひろ）
有本 庸浩（ありもと つねひろ）　成松 宏美（なりまつ ひろみ）
千葉 祐弥（ちば ゆうや）　中嶋 秀治（なかじま ひではる）
NTT コミュニケーション科学基礎研究所

一問一答型の雑談対話システム

私たちは日々，周りの人々との雑談を楽しみながら生活しています．このような人の対話欲求を充足させることを目的とするシステムを雑談対話システムと呼びます．システムとの対話は場所や時間に制約されないため，独居の人々や，高齢者施設・病院等で生活している人のコミュニケーションの機会を増やすことができます．さらに，人と異なり遠慮や気兼ねもいらないため，対人コミュニケーションを苦手とする人が安心して話せる相手として，また人に話すのははばかられるようなプライベートな話題の相談相手としてなど，多様な用途に有望であると考えられています．

こうした雑談対話システムを実現するうえで，私たちはまず，雑談対話の特徴に着目しました．雑談対話は，それまで研究が進められてきたタスク対話（スケジュール管理・レストラン予約など）と異なる，２つの大きな特徴があります．１つは，対話に出現する話題の幅が広い（オープンドメイン）という特徴です．雑談対話に現れる話題の幅について，私たちが収集したテキスト対話を調べたところ，3680対話中に18000もの話題が含まれていました．一番多い話題であった「旅行」でさえ，全体の0.7%しか含まれておらず，極めてロングテールな分布になっていたのです．これだけの話題をカバーするのは，容易なことではありません．もう１つの特徴は，対話の流れが多様であるということです．スケジュール管理のような対話であれば，ある程度決まった流れが存在するため，あらかじめ流れを想定してシステムを設計することができます．しかし，雑談対話はその場その場で流れが大きく異なるため，あらかじめ流れを想定しておくことが困難です．

これらの特徴のため，当初の雑談対話システムの開発では，ユーザ発話に対して一問一答で応答を生成する，というアプローチが採られました．あらかじめ入力と出力のペア（応答パターン）を大量に用意しておき，ユーザ発話と類似するパターンを検索して応答する，という方法です．もちろん対話は発話の連なりで構成されますので，文脈を無視した一問一答で良いのか，という議論も当然ありました．しかし実際問題として，文脈を考慮しようにも，発話の組合せが膨大すぎて全く扱いきれなかったのです．

この応答パターンの代表的な作成方法として，ルールベース，抽出ベース，生成ベースの３種類の方法があります．１番目のルールベースは，人手で応答ルールを作成する方法です．ユーザが「こんにちは」と言ったら「こんにちは」と返す，「眠い」と言ったら「おやすみ」と返す，というように，想定されるユーザ発話に対するシステム応答を設計者が作成します．人手で応答文をつくるため，制御性が高く不適切な発話をする危険性が低い，時事ネタのようなユーザを楽しませる発話を用意しやすい，といった利点があります．こうした利点から，SiriやAlexa等，現在の商用システムのほとんどは，ルールベースで実現されています．一方ルールベースの欠点として，人手で発話を構築するため，幅広い話題に対応できるシステムの構築が難しいことが

挙げられます．

2番目の抽出ベースは，大規模データから文（用例）を抽出・検索して発話に利用する方法です．2種類のアプローチがあり，類似文自体を用例として利用する方法（新聞記事，ブログ，ツイート単体など）と，発話の対を用例として利用する方法（対話ログ，ツイートリプライ，QAなど）があります．構築コストが非常に低く，ほぼすべての話題に何かしら応答可能という利点がありますが，類似文を返す場合はオウム返しになりがち，発話の対を返す場合は文脈が合わず関連性の薄い発話が出力されがちという欠点があります．

3番目は，生成ベースという，用例ベースの話題の幅を活かしつつ応答発話の品質を向上させるアプローチです．あらかじめ大量のテキストから，係り受け関係を手掛かりに関連する話題をペアとして抽出しておき，ユーザ発話中の重要な部分に対応するペアを利用して発話を生成します．こうすることで，抽出ベースでの問題点であった，無関係な文やオウム返しを抑制し，高品質な発話を生成することができます．

NTTではこれらの方法を組み合わせ，利点・欠点をうまく補い合うことで，安定して高品質な応答を生成するシステムを実現しています．

一問一答型雑談対話システムの問題点とNTTのアプローチ

ところが，実際にこれらの方法で構築した対話システムと話してみると，どうにもうまく話がかみ合わず，対話が破綻してしまうことがあります．調べてみると，発話単体への応答としては妥当だけれども，対話の文脈にそぐわない応答，というものが多く観察されたのです．例えば**図1**のように，「ラーメン食べに行きたいなあ」というユーザ発話に対する，「ラーメンはスープが命だよね」というシステム応答は，文脈を見なければ自然な応答です．一方，このユーザ発話が，ユーザが糖質制限でラーメンをなかなか食べられない，という文脈で発話されていた場合，上記のシステム発話は，ユーザの意図を無視した，意味の通らない発話となってしまいます．しかしながら，文脈を考慮して応答発話を出力したくとも，発話履歴の組合せが膨大となるため，これまで紹介した方法では考慮しきれません．

そこでNTTでは，3種類のアプローチで改善を図りました．1番目のアプローチは，話者切替による文脈のリセットです．複数のロボットを連携させて対話を進め，2体目のロボットが適宜割り込んで発話し文脈をリセットすることで，比較的小さなデータを使う場合でも，文脈との齟齬に対するユーザの違和感を軽減する方法を考案しました[1]．また，ロボットどうしの自然な対話をあらかじめつくり込んでおくことで，対話が破綻しかけたときでもロボット間対話を挟み込んで対話を自然に継続させたり，ユーザ発話を起点として展開される話の流れをつくり出せたりすることも分かりました．この機能を応用し，京都市動物園で「物知りAIロボット」の実証実験を行い，来場者と動物について知識を深める対話を実現しています[2]．

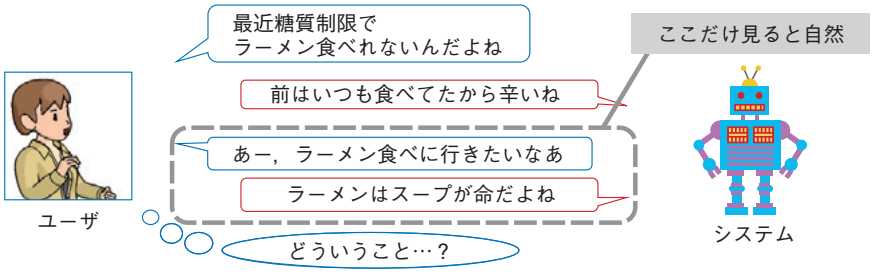

図1　一問一答型雑談対話システムの問題点

2番目のアプローチは，特定話題に限定した，高密度応答パターン列の構築です．当初のオープンドメインな雑談というコンセプトからはいったん離れますが，話題を限定すれば文脈を考慮しきれるのではないか，という真正面からのアプローチです．とはいえ，通常の雑談対話では，どう話題が展開されるか分かりません．そのため，話題を限定しやすいタスク対話と雑談対話の中間点として，議論対話を対象とすることにしました．ある特定の命題（例：永住するなら田舎か都会か）について，論点（例：老後の快適さ）を20個用意し，それぞれの論点について支持する意見・反対する意見を木構造としてどんどんつなげていく，という方法で，議論構造という高密度な応答パターン列を構築しました（**図2**）．このシステムを米国のオースティンで開催されたSxSW（サウス・バイ・サウス・ウェスト）という展示会で大阪大学・ATRの石黒研究室と共同で披露し，文脈を踏まえた議論対話を実現しています．

3番目のアプローチは，ユーザ発話の誘導による特定の文脈への引き込みです．1番目の物知りAIロボットの

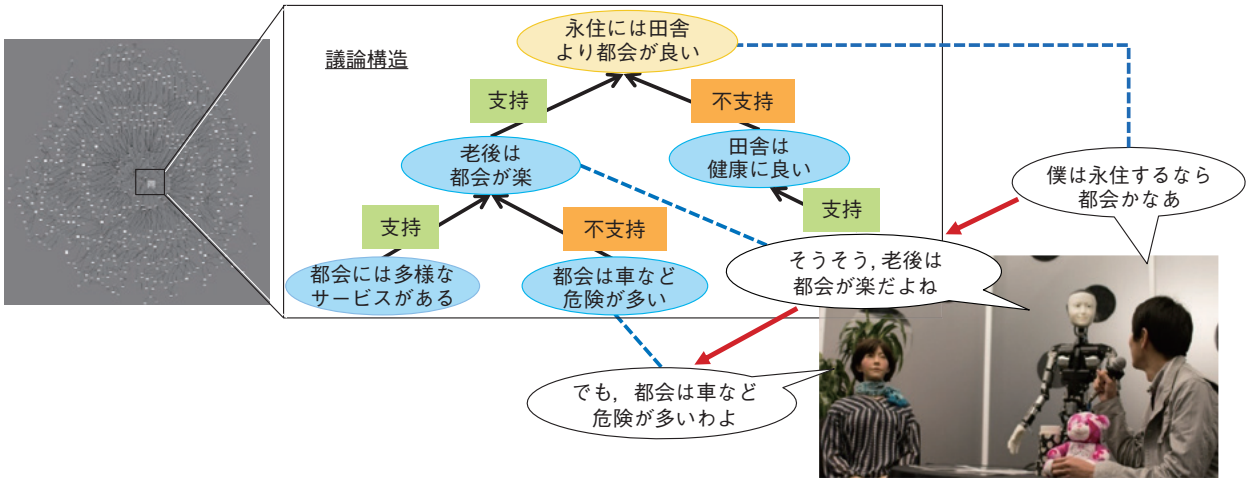

図 2　議論構造に基づく議論対話システム

実験や 2 番目の議論対話システムの実験を通して，タスク対話のように自明な流れが存在しない雑談対話であっても，うまくユーザ発話を誘導できれば，特定の対話の流れに引き込み続けられることに気付きました．これを利用し，タスク対話システムと同様の設計方法で旅行についての雑談を行うシステムを構築したところ，ほとんどのユーザを想定した流れに引き込むことができました．クラウドソーシングを利用した実験でも，非常に小規模なルール量で自然な雑談を実現できることを確認しています．

大規模深層学習による急速な性能向上

さて，ここまで紹介したシステムは，いずれも人がつくった発話や用例を選択するか，単語を組み合わせて人手で構築したテンプレートに当てはめるといった方法で発話していました．一方で近年，深層学習と呼ばれる技術が急速に進展しており，自然言語処理研究全体に非常に大きな影響をもたらしています．特に，大量のテキストデータで文の自然さや基本的な構造をあらかじめ学習する事前学習と呼ばれる方法が肝になっており，これを用いて学習された汎用言語モデルは，翻訳や質問応答など特定の目的に沿った少量のデータで転移学習（Fine-tune）することで，非常に高い性能を得られることが知られています．

雑談対話システムも例外ではなく，2020年に相次いで，深層学習を利用した高性能な雑談対話システムが英語圏で提案されました[(3)]．NTT でも，Twitterから収集した21億の発話ペア（数発話の文脈を入力，それに後続する 1 発話を出力とするペア）を用いて事前学習し，これまでの研究で蓄積してきた高品質な対話データ20万ペアを利用してFine-tuneすることで，非常に自然な雑談を行うシステムを実現しています[(3)]．このシステムは，雑談対話システムの性能を競う「対話システムライブコンペティション 3（ライブコンペ 3）」で圧倒的な性能で最優秀賞を獲得しました．ライブコンペ 3 では，システムが対応できる話題の幅広さを評価するため，ユーザは 2 つの固有名詞を話題として選択し，それらに触れるよう対話する，というレギュレーションがありました．**図 3** にライブコンペ 3 でのユーザとの対話（左がシステム）を示します．このユーザは，「水曜どうでしょう」というバラエティ番組と，「渡辺麻友」というタレントを話題として選んでいました．従来のシステムではここまで細かい話題に適切に応答することは困難だったのですが，私たちの構築したシステムは，渡辺麻友についてはよく知らない，としつつも，見事に応答し続けることに成功しています．

今後の展望

深層学習によって，非常に自然な発話文を生成できるようになった現在でも，課題は山積しています．例えば現在のシステムは，発話の一貫性や事実性について考慮せず，文の自然さ（生成確率）のみを用いて学習されているため，過去の発話と矛盾したことや，事実と異なる「嘘」をよく発話します．さらに，対話の内容や相手のことを記憶していないため，数カ月にわたって繰り返し対話し続けることは困難です．人の対話欲求を継続的に満たす雑談対話システム実現に向けて，これら

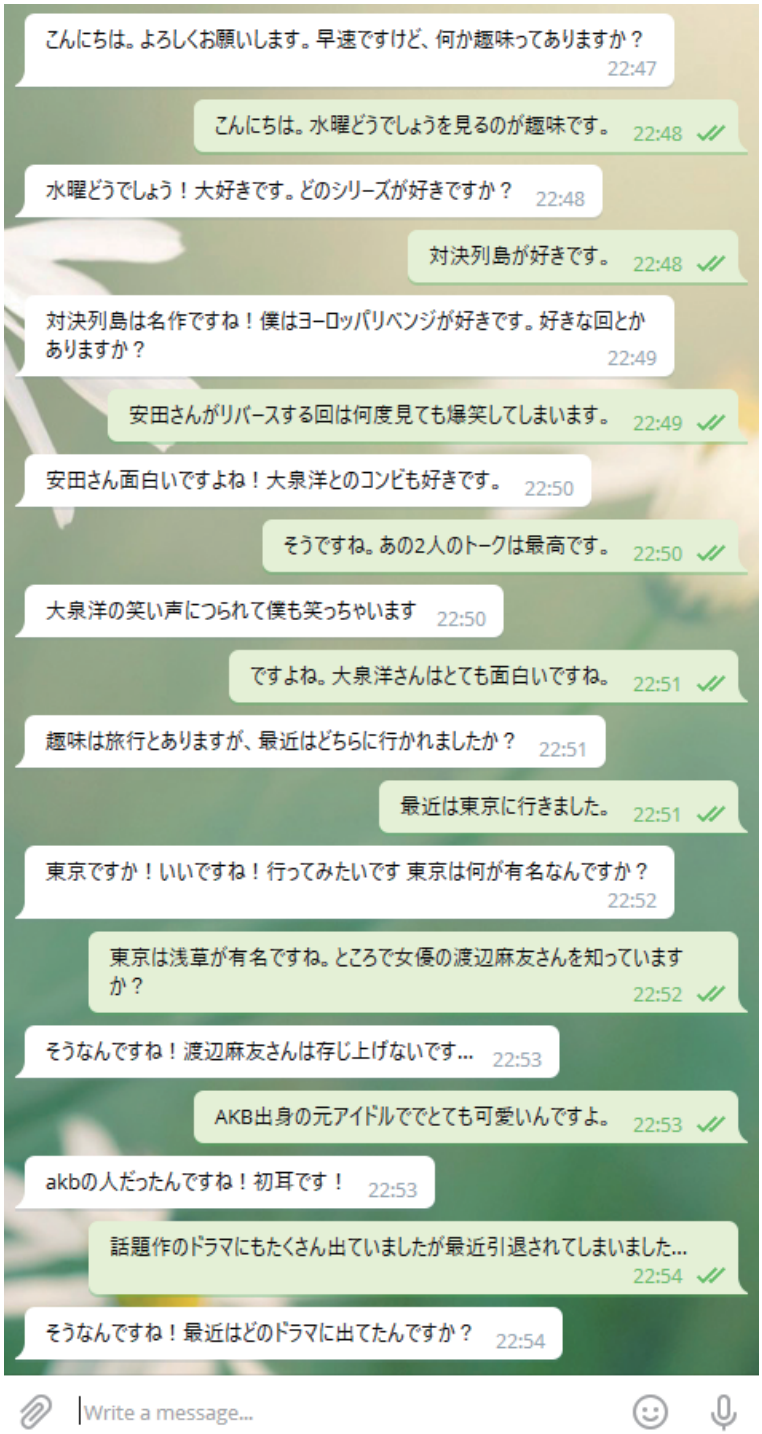

図3　ライブコンペ3での対話例

の課題に取り組んでいく予定です．

■参考文献

(1) H. Sugiyama, T. Meguro, Y. Yoshikawa, and J. Yamato: “Improving Dialogue Continuity using Inter-Robot Interaction,” Proc. of IRO-MAN 2018, pp. 105-112, Nanjing, China, August 2018.

(2) 杉山・水上・成松：“複数ロボット協調による一問一答型雑談対話からの脱却，” 第32回人工知能学会全国大会，2018.

(3) S. Roller, E. Dinan, N. Goyal, D. Ju, M. Williamson, Y. Liu, J. Xu, M. Ott, K. Shuster, E. M. Smith, Y. L. Boureau, and J. Weston: “Recipes for building an open-domain chatbot,” Proc. of EACL 2021: Main Volume, pp. 300-325, April 2021.

(4) 杉山・成松・水上・有本・千葉・目黒・中嶋：“Transformer encoder-decoder モデルによる趣味雑談システムの構築，” 言語・音声理解と対話処理研究会（SIG-SLUD），Vol. B5, No. 02, pp. 104-109, 2020.

（上段左から）杉山 弘晃/ 成松 宏美
（中段左から）千葉 祐弥/ 有本 庸浩
（下段左から）水上 雅博/ 中嶋 秀治

大規模深層学習に基づく対話モデルによって，雑談対話システムの品質が基礎研究から実サービスに急接近しています．システムと当たり前に会話する未来がすぐそこまで来ていると感じていますので，機を逃さず研究開発に取り組んでいきたいと思います．

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
協創情報研究部
インタラクション対話研究グループ
TEL　0774-93-5020
FAX　0774-93-5026
E-mail　cs-liaison-ml@hco.ntt.co.jp