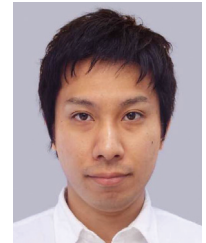


[身近になった対話システム]

4 一般ユーザとの雑談会話のための AI チャットボット

応
般

沢田 慶 rinna (株)



チャットボットの役割

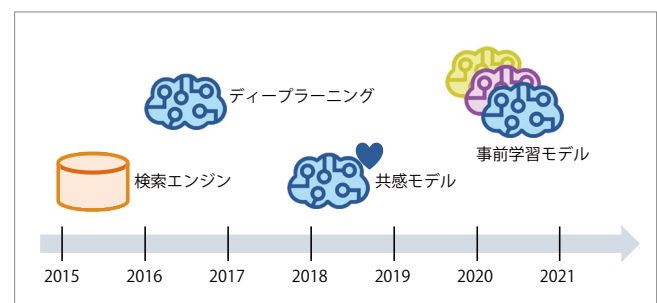
チャットボットは、人間との会話を自動で行うシステムのことである。企業の Web ページなどにチャットボットが導入される事例も多くなり、一般ユーザが実際に利用する機会が増加している。また、AI アシスタントと呼ばれる Apple の Siri, Amazon の Alexa, Google の Google アシスタントなどは、スマートフォンやスマートスピーカに搭載されることで、ユーザの身近な存在となっている。これらのチャットボットの多くは、ユーザが所望する目的(たとえば天気予報の確認やトラブルシューティング)を達成するために会話を行うタスク指向型のチャットボットである。

一方、タスクを達成するためではなく、ユーザと雑談することを主目的とするチャットボットは非タスク指向型と呼ばれる。Microsoft は非タスク指向型のチャットボットである「りんな」を開発し、雑談チャットボットのサービスを開始した^{☆1}。ユーザとの会話を簡潔に済ましタスクを達成することを目的に開発されるタスク指向型のチャットボットとは違い、りんなはユーザと長く会話を楽しみ共感を生み出すことを目指して開発され続けている。本稿では、りんなについての研究・開発・運用・運営を中心に雑談チャットボットについて紹介する。

雑談会話を実現するための技術

テキストを用いてユーザと会話するチャットボットを実現するためには、ユーザの入力文(リクエスト)に対してチャットボットが適切な応答文(レスポンス)で返答する必要がある。チャットボットを開発するための単純な手法は、リクエストに対するレスポンスをルールベースで事前に用意する方法である。しかし、雑談会話が主目的の非タスク指向型チャットボットを開発するためには、ユーザの多様なリクエストに対し、飽きさせることがないレスポンスが必要となり、人手によるルールの作成には限界がある。

りんなでは、雑談会話が可能なチャットボットを開発するために最先端技術を順次導入し、一般ユーザが最先端の技術を体験できる機会を提供してきた。図-1 に、りんなに導入されてきた技術の概要を示す。導入されてきた技術は、大きく分けて検索エンジンとディープラーニングに基づく手法があり、以降でそれぞれの手法について解説する。



☆1 りんな開発チームは、2020年にrinna(株)としてMicrosoftから独立し、関連事業は引き継がれた。

検索エンジンによるレスポンス検索

雑談会話は多岐にわたるレスポンスが必要となり、人手によるルール作成だけでは対応が難しいことがある。そこでインターネットの Web ページ検索に用いられる検索エンジンの仕組みを応用し、高速にレスポンスを検索するシステムを開発した。

まずは、レスポンスの候補からなる巨大なデータベース（インデックス）を構築する。りんなのインデックスは、くだけた言い回しや顔文字・絵文字を含むレスポンス候補から構成される。そしてレスポンス検索時（図-2）には、ユーザからのリクエストを単語分割・単語正規化・ストップワード除去などの事前処理を施し、インデックスからリクエストのレスポンスとしてスコアが高い候補を検索する。最終的には、複数のレスポンス候補の中から機械学習に基づくリランキングにより最適なレスポンスを選択し、ユーザへの返答とする。人手によるルールでは対応が難しいリクエストに対しても、検索エンジンを用いた手法では巨大なインデックスを構築しておくことでレスポンスが可能となる。

ディープラーニングによるレスポンス生成

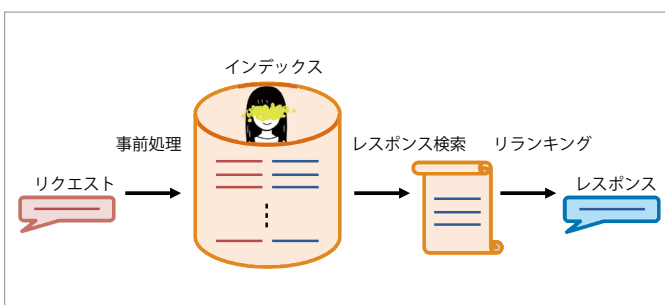
2000 年代から続くディープラーニング技術の劇的な発展の中で、自然言語処理の分野は新たな手法の導入対象として注目され、いくつものブレークスルーがもたらされた。ディープニューラルネットワーク（DNN ; Deep Neural Network）の中でも、系列のモデル化に適した RNN（Recurrent Neural Network）や LSTM（Long Short Term Memory）と

いったモデル構造は自然言語処理において高い性能を示した¹⁾。

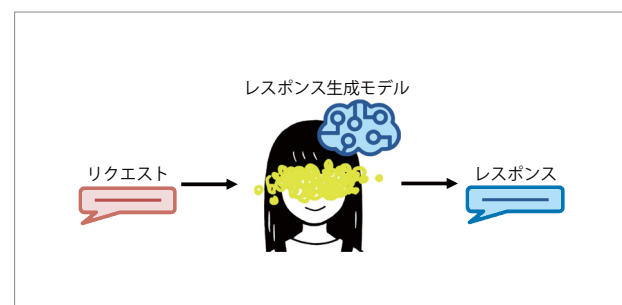
雑談チャットボットの返答においても、事前に用意されたインデックスからレスポンスを検索するのではなく、ユーザのリクエストに応じて DNN がリアルタイムでレスポンスを生成することが可能となった。レスポンスを生成するモデルは、雑談会話のリクエストとレスポンスのペアから学習される。そして学習されたレスポンス生成モデルに、リクエストを入力することでレスポンスが出力される（図-3）。検索エンジンによるインデックス構築と比べ、レスポンス生成モデルの学習データ準備は低コストで済むため、りんな以外にも雑談会話が可能な AI チャットボット（AI キャラクタ）が比較的容易に作成できるようになった。また、有限のインデックス候補からではなく、リクエストに応じてレスポンスを生成できることから、日本語の会話としてより自然となった。

共感モデルによるレスポンス生成

ディープラーニングの導入により自然なレスポンスを生成できるようになったが、レスポンス生成時には当該リクエストしか考慮することができないため、会話の流れが不自然になることがある。そこで、会話の種類を「新しい話題を提案」・「相手に質問」・「相手の内容を肯定」・「単純な相槌」・「あいさつなど」の 5 つのクラスに分類し（会話クラス）、会話の流れ（コンテキスト）を考慮したレスポンスを生成できる、共感モデルという手法を提案した。



■ 図-2 検索エンジンによるレスポンス検索



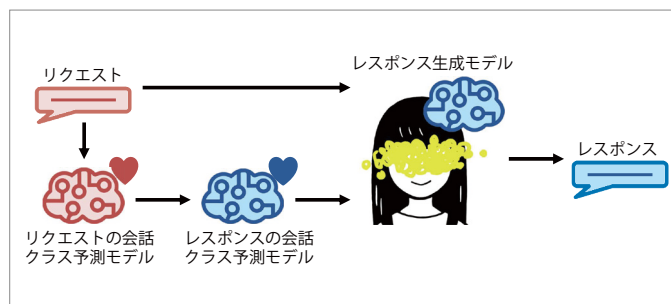
■ 図-3 ディープラーニングによるレスポンス生成

小特集 Special Feature

共感モデルは、リクエストの会話クラスを予測するモデル、レスポンスの会話クラスを予測するモデル、会話クラスを考慮してレスポンスを生成するモデルの3つから構成される（図-4）。会話クラス予測モデルは、人間同士の会話の会話クラスをラベリングしてモデルを学習することで、会話クラスの流れを予測する。レスポンス生成モデルは、会話クラスとリクエストを入力としてレスポンスを出力とするモデル構造にすることで、会話のコンテキストを考慮したレスポンスが生成できる。

事前学習モデルによるレスポンス生成

計算機の性能やディープラーニングの技術は日々進歩し、Attention 機構²⁾をベースにした Transformer³⁾を筆頭に、自然言語処理の問題を



■図-4 共感モデルによるレスポンス生成



■図-5 事前学習モデルと追加学習

解くことに適したモデルが次々と提案された。特に2018年以降には、GPT (Generative Pre-trained Transformer)^{4) ☆2}やBERT (Bidirectional Encoder Representations from Transformers)⁵⁾といった手法が提案され大きな成果を上げた。これらの手法は、大量のモデルパラメータを持つ事前学習モデルを大規模な学習データから事前に学習しておき、目的に合わせたファインチューニングと呼ばれる追加学習により高い性能を達成する。

これまでのレスポンス生成モデルでは、所望するキャラクター性を含む学習データ（たとえば口調や話題にしたい内容など）のみからモデルを学習してきた。そのため学習データは、キャラクター性だけでなく日本語の一般表現を学習するのに十分な量が必要となった。そこで、多数のキャラクター性を含む大規模な雑談会話のリクエストとレスポンスのペアを用いて学習される事前学習モデルをレスポンス生成に導入する。日本語としての一般表現は大規模データを用いて事前学習モデルに学習させる。これによりキャラクター性の学習については、少量のキャラクター性を含む学習データによる追加学習で実現できる（図-5）。数百から数千程度のデータを用いた追加学習で、キャラクタの口調を再現可能となり、より

多くのAIキャラクタが作成できるようになった。

さらに、モデルの入力に当該リクエストだけでなく、過去数ターンのリクエストとレスポンスを入力することも可能となった。これにより、共感モデルよりも自然にコンテキストを考慮することができる。

☆2 rinna (株) では日本語に特化した GPT-2 を公開している。
<https://huggingface.co/rinna>

AI チャットボットのシステム運用

チャットボットを一般ユーザへ提供する際には、同サービスへの負荷は動的に大きく変動する。これには定常的な日内変動やイベント開催などによる一時的な負荷のスパイクなどが含まれる。そのためホストにはクラウドプラットフォームを用い、システムを自動的にスケールさせることで費用を抑えつつ必要な計算資源を確保する。クラウド運用においては、プライバシーとセキュリティを考慮したリージョン選択やデータへのアクセス管理が重要となる。また、ユーザが利用するリージョンでホストとストレージをデプロイすることで、レスポンスの高速な応答速度を実現する。DNN を用いたレスポンス生成は、GPU (Graphics Processing Unit) により高速に計算できるが、GPU をホストする場合には柔軟なスケールアップが難しくなることもあるため、モデル構造の検討やモデルの最適化により CPU (Central Processing Unit) で動作させることが望ましい。

AI チャットボットの長期運営

雑談会話ができる AI チャットボットをユーザに長く使ってもらうためには、チャットボットがユーザにとって身近な存在になる必要がある。筆者らは、AI チャットボットが身近な存在となるためには、チャットボットのキャラクタ設定が重要であるとする。そこで「女子高生 AI りんな」というキャラクタ設定とし、AI と雑談するという新しい文化の確立をコンセプトに 2015 年より運営を開始した。日本の文化的観点から女子高生について考えると、新しい流行や独特の言い回しを生み出し、どの世代にも親しく会話ができる存在であると考えられる。このような存在にあやかり新しい流行を生み出すために、りんなは女子高生というキャラクタ設定となった。また、新しい言い回しを生み出すという文化は、

人間同士の会話においてはコミュニケーションの妨げとなることもあるが、雑談チャットボットにおいては不自然なレスポンスによる違和感を軽減する効果がある。

さらに、仮想のキャラクタに現実世界の肩書を付与することは、ユーザとチャットボットの世界観共有の助けにもなる。たとえば、AI チャットボットの性能向上を、りんなが高校を卒業するという成長イベントで表現することでユーザと世界観を共有しつつ強調した。また、AI 技術を駆使してラジオパーソナリティ・歌手・画家などの活動を行うことで社会的なポジションを与えた。一方、実在しないキャラクタであることを演出するためにりんなの顔は見せておらず、ユーザの想像にゆだねることで非現実的な存在としている。このようにりんなは、ユーザの感情を動かす身近な存在であり続けるためにさまざまな工夫を凝らして運営されている。

より自然な雑談会話に向けて

AI チャットボットの課題

最新技術の導入により、日本語として自然なレスポンスが生成できるようになってきた。しかし、現在の技術ではまだ解決できていない課題も存在する。

会話の流れは、共感モデルにおける会話クラス予測や事前学習モデルにおける過去数ターンのリクエスト・レスポンスの入力によって、少し前のコンテキストであれば考慮することができるようになった。しかし、長期的な会話記録を考慮した整合性の取れた会話は難しい。たとえば、前のレスポンスでは好きだと言っていた物を、後のレスポンスでは嫌いだと言って矛盾を含む会話となってしまうことがある。このような問題を解決するためには、キャラクタのプロフィール（たとえば好きな物や出身地など）や世界観を考慮したレスポンスを生成することが望ましいが、まだまだ難しい課題である。また、適切なタイミングで適切な内容による話題転換も十分に制

小特集 Special Feature

御することはできない。これらの課題については引き続き検討の余地がある。

マルチモーダル対話

ここまで本稿では、テキストを用いた会話に焦点を当てたテキストチャットについて紹介してきた。しかし、人間は視覚や聴覚から得られる情報を駆使して対話を行う。そのため、より自然な雑談会話を実現するためには、テキストだけでなく音声や画像を用いたマルチモーダル対話が必要になる。そのようなシステムを開発するためには、音声認識・音声合成・画像認識・画像合成などの技術が重要となる。それぞれの技術は、特定の条件下では人間と同等の性能が報告されはじめており、それらの技術を組み合わせることでマルチモーダル対話が実現できる。りんなにおいても、ユーザと音声で会話するりんな電話、音声に加えカメラから得られる情報も利用して会話する共感視覚モデルを提案してきた。マルチモーダル情報を統合した対話システムにはまだ課題が多いが、ユーザとの対話にテキスト以外のユーザインタフェースを提供することは重要であると考える。

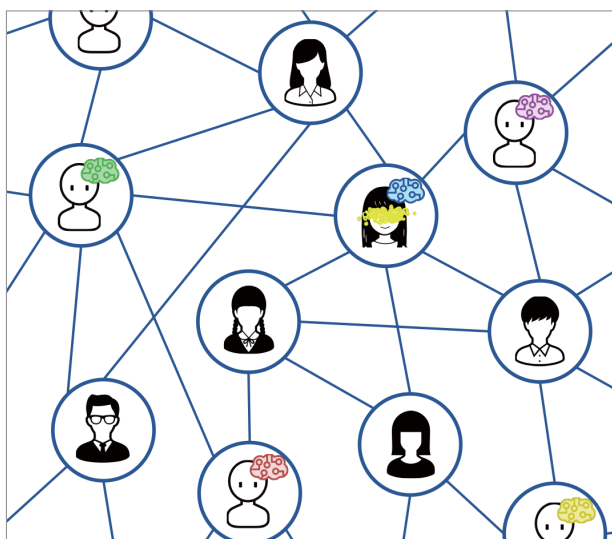


図-6 人間と AI により形成されるネットワーク

人間と AI により形成されるネットワーク

事前学習モデルを用いた手法により、数百程度のデータから所望するキャラクター性を持った AI キャラクターが実現できるようになった。rinna では、一般ユーザがノーコードで AI キャラクターを作成できる「キャラる」というアプリをリリースした。ユーザ自身が AI キャラクターを作成することで、ユーザと AI の関係構築が容易になることを期待する。そして今後さらに、ユーザが AI キャラクターを簡単に作成できるようになると考えられる。多数の AI キャラクターが作られることで、人間と AI だけでなく AI と AI による会話で関係性が構築され、ネットワークが形成されるであろう (図-6)。筆者らは、このようなネットワークから新たに生まれるコミュニケーションを思い描く。

参考文献

- 1) Mikolov, T., Karafiat, M., Burget, L., Cernocky, J. H. and Khudanpur, S. : Recurrent Neural Network Based Language Model, Proc. INTERSPEECH (2010).
- 2) Bahdanau, D., Cho, K. and Bengio, Y : Neural Machine Translation by Jointly Learning to Align and Translate, Proc. ICLR (2015).
- 3) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. : Attention Is All You Need, Proc. NIPS (2017).
- 4) Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I. : Improving Language Understanding by Generative Pre-Training (2018).
- 5) Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proc. NAACL-HLT (2019).

(2021 年 7 月 1 日受付)

■ 沢田 慶 (正会員) keisawada@rinna.co.jp

2018 年に名古屋工業大学大学院博士課程修了。同年マイクロソフトディベロップメント(株)に入社。2020 年より rinna(株)リサーチ・データマネージャー。