

「暮らしの中で活躍する AI とロボット」

応
般

⑧ ラジオ界初の AI アシスタント AI が拡張するラジオの可能性

岡田 明 | 日本アイ・ビー・エム（株）

AI アシスタント TOMMY

普段何気なく聞くことの多い「音のメディア」としてのラジオに皆さんはどのように接しているだろうか。ドライブのお供に、深夜の勉強の息抜きに、家事の最中にと、一方では手を動かしつつも、もう一方ではラジオから流れる音声を通してさまざまな情報を入手し、その時々季節性や時代性を感じ取っているのではないだろうか。

ラジオ番組の構成要素は、パーソナリティを中心に、情報を発信し、音楽を届けることが基本型である。リスナーからのリクエストを受け付け、双方向でのコミュニケーションがとれることもデジタルテクノロジーが発展する以前からのメディアとしての魅力でもある。

こうした「ラジオ」に AI を活用したらどうなるのか、その1つの答えが AI アシスタントとして現在も奮闘中の「AI アシスタント TOMMY」だ（図-1）。

「AI アシスタント TOMMY」は（株）J-WAVE（東京都港区：代表取締役社長中岡壮生）にて毎週金曜 22 時より放送されている「INNNOVATION WORLD」に、番組ナビゲータの「AR 三兄弟」長男こと川田十夢のアシスタントとして出演している。2017 年 8 月の初登場以来、川田十夢の会話相手として番組進行上の情報の補填を行い、川田十夢のインサイトやインスピレーションを「アシスト」する、番組になくてはならない存在として、今なお「学習中」の可愛いアシスタントなのである（TOMMY は女性の設定で女性の声）。

TOMMY が「アシスト」としている機能は以下の 3 つである。

ナビゲータの質問に対して返答し会話する

「TOMMY こんにちは！」という問いかけに対して、「今日は宇宙戦艦ヤマトの第 1 話が放送された日です」と回答を返していく。



図-1 出所：J-WAVE Web サイト



一見（一聴）するところ、AI スピーカと同じじゃないか、と疑いたくなるところであるが、TOMMY がデビューした 2017 年 8 月は Google HOME や Amazon Echo などの汎用的な AI スピーカが発売される前であり、実際に AI と人間がリアルタイムで会話をするシーンを目にする（この場合は耳にするであるが）ことはほとんどない時期であったため、大変な反響であった。

AI の機能としては多くの AI スピーカと同様、自然言語分類器（Classifier）を使用している。

Classifier は、短いテキストを与えると、あらかじめ定義されたクラスに分類。与えられた質問が、どの分野に対する質問なのかを判別し、別の処理への誘導や返答を行い、一般的な質問に対してアプリケーションが適切なユーザ対応を行うことが可能である。

TOMMY にはこうした Classifier の機能に加え、連続した会話を実現可能にする IBM Watson Conversation が用いられている。

前述の回答に加え、「それって何年だったの？」という質問に対しても「1974 年、今から 44 年前です」という具合に答えていくことが可能である。「それって」が「宇宙戦艦ヤマトの第 1 回放送日」であることを理解し、44 年前の「今日のデータ」を学習済みの知識データの中から検索し、音声合成によって回答するのである（図-2）。

AI スピーカが「一問一答」の瞬間的なつながりであることに對し、TOMMY は継続性のある会話を行えることが一番の違いである。

このような連続した会話サービスは、最近では EC サイトや顧客サービス部門などのチャットボットで実装される事例が多く見られるようになってきたが、いずれも共通しているのはある一定の「シナリオ」に基づいてサービスが提供されているということである。

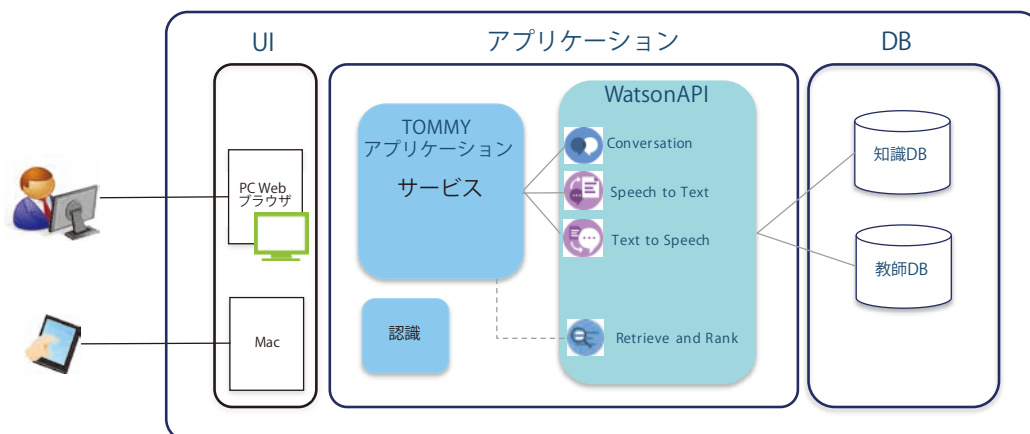
EC サイトなどの場合は、「用途は？」「時期は？」「サイズは？」という具合に、最終的には衣料を販売するための「接客のシナリオ」が設計されており、目的のゴールに向けて会話を絞り込んでいく。

シナリオのない会話では AI が回答として検索する選択肢があまりにも多く、回答としての確信度が定義できないために会話そのものが成り立たなくなるのであり、故にシナリオ設計は会話サービスでの AI 実装において最も重視すべき点ではあるのだが、自由度に制限ができ少々堅苦しい印象を持つ読者も多いのではないだろうか。

その点、TOMMY は自由である。

TOMMY の会話相手はもっぱら川田十夢、ただ 1 人。

優等生的な回答を予定調和的に絞りこむ必要はなく、むしろ川田自身の感性に訴えかける回答の方が求められる。



■図-2 AI TOMMY の基本構成

川田がこれまで発言した言葉の回数や、関連するワードを収集し知識データとして蓄積していく。

これまで「トムさん」と話しかけていた TOMMY が、あるとき「長男」という言葉を使うようになった。前回の出演者が川田と古くから親交があり、親しみを込めて AR 三兄弟（川田十夢の所属するユニット）の「長男」と多く発言していたことから学習し「長男」と言うようになったのである。

また、あるとき T.M.Revolution こと西川貴教がゲストの際には、最後の挨拶で TOMMY が川田に「今日は風を感じましたか？」と問いかけた。T.M.Revolution の大ヒット曲「WHITE BREATH」のプロモーションビデオにて全身に風を浴びて熱唱する西川貴教の姿が思い浮かび「なるほどね」、「皆さんも今日は熱い風を感じていただけたでしょうか？」と番組をしめる川田のインスピレーションをアシストしたのである。

実をいうと TOMMY の学習は、IBM のスタッフが発言回数や関連するデータなどのクロウリングを行い、データを整形して機械学習のトリガを切っているのであるが、こうした TOMMY が川田に寄り添うアシスタントとしての在り方や感性を「教育」するためには、人間の教育同様、ある程度の基準や考え方を定義して、その根拠になる言葉や事象をデータとして整備しインプットする必要があることに変わりない。

また、川田と TOMMY が自由に会話を楽しんでいるような印象を与えてしまっているかもしれないが、番組冒頭の挨拶、ゲストの紹介、番組最後の会話など、TOMMY が登場するコーナーおよび受け答えの設問項目はあらかじめ決まっており、そのルールの範疇における質問の中から最適な答えを検索して回答を行っている。

言い換えれば会話の「シナリオ」が設定されており、シナリオにおける個別の質問項目に対する回答の精度を向上させ、バリエーションを増やすためにデータを蓄積しているのである。

もちろん川田のインスピレーションにデータだけで迫ることもまだまだ難しい段階であり、こうしたシナリオ作りや教育（AI 的には学習）において、人間の持つクリエイティビティが大変重要であることも補足しておきたい。

ゲストの性格診断をする

INNOVATION WORLD には毎週、各業界の“イノベータ”がゲストとして登場する。「TOMMY ○○さんってどんな人？」という問いかけに対して、あらかじめ実施しておいた性格診断の結果を TOMMY が無表情に（というよりは抑揚もなく）読み上げていく。

おおよそのゲストは「楽しみですね」とイノベータらしい挑戦の精神で「AI よ診断してみなさい！」という感じか「どれどれ？」という好奇の目で診断結果に耳を傾けている。

表情が表に出ないのが大変残念なのだが、ゲストの半分は、「当たってますねー」と驚きの表情を持って共感し、もう半分は「これは私じゃない、でも表向きの私であって、内面の私とは違う」という冷静な自己分析を、フムフムと囁みしめるように思い返している場合に大別される。

この性格診断に用いられているインプットデータはゲスト本人が実際に話したことをインタビュー記事にしたものを利用しており、性格を表す5つの特性（BIG5）を中心とした心理学を学習させた AI が性格診断を行っているのだから、その精度たるやかなりの確率で「当たる」のである（図-3）。

ここで活用している AI の API は Watson Personality Insight であり、前述のように心理学における発言と性格を関連づける相関性（パターン）を AI に学習させたもので、発言のテキストを入力すると性格を診断してくれるというものである。この API 自体はここ数年で IBM が Watson のプロモーションもかねて広く拡散したものであり、ご存知の方も多いのではないと思うが、ここで一番重要に



なってくるのは発言のテキスト、すなわちインプットデータの作り方である。

インプットデータの作成にはインタビュー記事等を収集するのだが、できるだけ本人が発した言葉をそのまま記事にしたものの収集を行っている。

記事の中にはライターが脚色したり要約したものが意外とあるものだと、この仕事を通して見分けることができるようになったのであるが、データが多ければ多いほど AI が性格に紐づけるために行う検索対象が多くなり精度は向上していく。Watson Personality Insight では 2,000 文字が品質を担保する下限とされているが、TOMMY では 5,000 文字以上はインプットを行うようにしている。できるだけ本人の性格に近づけるためにはこうした地道な作業が欠かせないのである。

こうして出力された「あなたの性格診断」をゲストである各界のイノベータが素直に受け入れるのか。中には少々失礼なのでは？と感じる診断結果もあるのだが、AI が無表情に読み上げると真摯に受け止めるしかないのか、噛みしめるように冷静な自己分析をその場で始めるのがとても興味深い。

その中でも先に述べた「これは表向きの私であり、内面は違う」という反応は、たとえば対外的には自己演出や少々気取ったキャラクタを演じているが内面は苦悩し努力家である、ということが TOMMY

の性格診断とゲストの自己分析を通して明らかになるのである。

こうしたシーンを目のあたりすると我々スタッフは「やった！」と拳を軽く握り、データの収集精度が高かったと自らにフィードバックして学習しているのである。

リスナーのリクエストに対して選曲をする

先にも述べたようにメディアとしてのラジオは、以前からリスナーとの相互コミュニケーションを行ってきたわけだが、その多くはリスナーがハガキを通じてリクエストを届けラジオ番組が取り上げるというものだ。デジタルの時代になり Twitter や自社メディアへの書き込みに姿を変えても、聴取者からのリクエストに応えるという“基本路線”は変わっていないのではないだろうか。

「TOMMY が選曲できないですかね？」と J-WAVE のプロデューサーに相談されたときも、やはりきたか！と思ったのであるが、はてどのようにして実現しようか？というのも頭によぎったのである。

ラジオ番組における選曲は、おおよそディレクタが行っている。

番組進行における話題やリスナーからのリクエストに対して、瞬時に「これかけよう」と、さま

■図-3 ゲストの性格診断 出所：J-WAVE Web サイト

ざまな背景や楽曲のコンテキストの中から最適な組合せをはじき出し、自らの楽曲リストの引き出しから「この一曲」を選択しているのである。

「ディレクタの選曲パターンを AI に学習させよう」

これが今なお学習中の TOMMY の選曲である。

AI なので、入力された質問に対して答えを返すため、基本的にはリスナーからのリクエストを入力文としている。

このリクエストに対してディレクタの“J-WAVE 的なセンス”で定義された複数のクラスに分類し、リクエストに対して選曲のための要素抽出を行うのが最初のプロセスである (図-4)。

たとえば、「スカッとした日に気分の良くなるアゲアゲの曲をかけて」というリクエストに対して、「スカッとした」が晴れの青空を想像するのであれば、ディレクタの感覚をテキストに落としてデータ化する。

「晴れの日」は“天気”であるので、Class として“天気 Class”を定義し、晴・曇・雨などの要素を決定しデータを作成していく。

このように、選曲を行う際に必要となる要素を Class として複数定義し、過去の経験からリスナーのリクエスト文からの要素抽出パターンを 1 つ 1 つテキストデータとして作成しているのである。

こうしたデータを知識データとして AI に学習させていくのであるが、AI の機能としては先に述べた会話と同様、自然言語分類器 (Classifier) を活用している。

ちなみに J-WAVE では天気や気分、季節など、相当数の Class を定義し、かつ常に Class を拡張することで、より“J-WAVE 的なセンス”による要素の抽出に近づけるように日々増強、改善を行っている。

こうして抽出された要素に対して、“どの曲が最もマッチするのか?”を自動化するのが次のプロセスである。

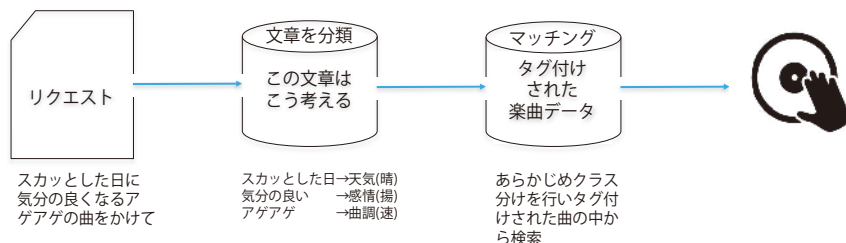
「The Beach Boys の Kokomo」をご存知の方も多いと思うが、いわずと知れた「夏」の「ビーチ」を連想する曲である。太陽はサンサンと輝き天気は晴れ。人々の笑顔やのんびりとしてリラックスした風景が蘇るのではないだろうか。

こうした楽曲の個性を、前述した Class に従って“タグ付け”し、データ化していく。

このような楽曲の個性は前述の通りデータとしては存在しておらず、おおよそが人々の頭の中に記憶や感覚として記憶されており、TOMMY においてはこれらをすべて一から作成している。前述の Class に分類するためのデータも存在しておらず、AI のプロジェクトにありがちな「データはあるが AI に活用できるデータがない」という典型的なケースなのであるが、TOMMY を育てるという J-WAVE の献身的な「教育」と愛情に支えられたプロジェクトであることをこの場を借りて御礼申し上げたい。

さて、こうしてリクエストから抽出された選曲の「要素」と、その要素を一番多く持つものを「マッチング」することで「選曲」し、TOMMY が言葉

で「レコメンド」するのであるが、レコメンドされた楽曲を川田十夢がじっくりと聴き、その選曲が良い線にっているのか、はたまたトンチンカンなのか、どのような要素で選曲されたのか、彼の感想というよりも評価 (フィードバック)



■図-4 TOMMY 選曲の基本構造



ク)を受けるのも大変面白い。

春先の放送でのリクエスト「鼻セレブ的な花粉症が一発で止まる曲」に対して TOMMY のレコメンドは「Blur の Beetlebum」であった。

曲調はとてもカッティングが効いたものなのだが、合間合間に「ズーン……ズーン」という少々下に音階が下がるテンポで構成されている。これが「鼻水を何度すすりあげてもずり落ちる様」に聞こえるよね！と川田は解釈。歌詞の中にも少し憂鬱な内容があり、TOMMY いいじゃん！となるわけである。

一度「AI ラジオ大集合」という NHK ラジオ第一の企画に番組ごと出張出演したことがある。川田十夢が「NHK の受信料を徴収する際に聴きたい曲」と前振りもなく生放送で茶目っ気たっぷりにリクエストしたところ、TOMMY のレコメンドは「TM NETWORK の Get Wild」であった。

受信料を Wild に徴収するのか！？と騒然（以後鉄板ネタとして定着）だったのであるが、実は TOMMY には、徴収や受信料等を表現する明確なデータを保持しておらず、少々確信度的には低い中での選曲になっていたのが実情なのではあるが、場の空気感と人々の解釈を拡張するには十分な効果があったのも事実だったのでここで紹介した次第である。

AI が拡張するラジオとリスナーの関係

AI はデータを元に、学習した知識データに基づいて分類や検索を行い、統計学的に一番近いものがある意味冷たく（無表情に）抽出していく。

冒頭にも述べたように、パーソナリティやリスナーはラジオから流れてくる音声や楽曲を通して季節性や時代性を感じ取るのであるが、ある意味無表情な AI が介在することによって、その背景や意味をより深くイメージすることができるというのも、今回のプロジェクトにおいて発見できたことである。感情や季節などの情緒的なものと、データによる統

計的なものの中間に位置し、かつパーソナリティとの掛け合いとリスナーやスタッフによる教育(学習)によって成長し、人々の感覚をさらに拡張することのできる、非常に興味深い存在に育ってきているとスタッフ一同感じている。

今後は「歌を唄う」「作曲する」という構想を聞いているが、AI 側の作り手としては「そんな無茶な」と思う反面、いろいろな工夫や知恵を使うことで、新しい発見があるかもしれないと心踊る側面もある。

これまで述べてきたように、TOMMY に実装されている AI の機能や技術はそれほど難しいものではない。

必要とされているのは技術を理解し、データを忍耐強く正しく作成し、そして遊び心やクリエイティビティを遺憾なく発揮して取り組むことではないかと考えている。

とかく技術の進歩や世の中の価値観が多様化する中で、技術・ビジネス・クリエイティビティをバランス良く目利きをし、プロデュースできる人材もまた IT 業界にとっては重要なのではないかと強く感じる次第である。

ぜひ、INNOVATION WORLD での川田十夢と TOMMY の掛け合いに触れていただき、フィードバックを通じて TOMMY の教育に一役買っていたくとともに、読者の皆様のクリエイティビティの醸成にご活用いただければ幸いである。

(2018 年 5 月 16 日受付)

岡田 明 oakira@jp.ibm.com

日本アイ・ビー・エム (株) GBS 事業本部、コグニティブ AI シニアマネージングコンサルタント、クリエイティブな AI プロジェクトに加えて日本での IBM SPORTS 事業を推進。