

NTTドコモ
サービスイノベーション部 担当部長

吉村 健 Takeshi Yoshimura

音声認識・自然言語処理・自然対話技術で日常生活をより便利に

検索エンジンや自動車の運転アシスト機能等をはじめ、日常のさまざまなシーンでAI（人工知能）が活用され、世の中がより安全、便利になってきています。さらに生成AIであるChatGPTが話題になり、その秘められた可能性に期待が高まっています。一方、こうしたAIを使うにあたっては、まだ人が機械に合わせて使わなければならない場合が多く、自然言語処理技術により自然なかたちでの利用に近づけることができます。NTTドコモ サービスイノベーション部の吉村健氏に、音声認識・自然言語処理・自然対話技術、そして、開発者としての姿勢を伺いました。



幅広く、奥深い音声認識・自然言語処理技術

現在、手掛けている開発の概要をお聞かせいただけますか。

私は、1999年にNTTドコモに入社し、当時のマルチメディア研究所で、VoIP（Voice over IP）のヘッダ圧縮、ストリーミング、MBMS（Multimedia Broadcast and Multicast Service）等の技術検討・国際標準化、および音響QRコード技術（音波による情報伝送技術）の開発、フィーチャーフォン向け音声認識システムの開発等を行ってきました。その後、データマイニングを手掛け、異動となったサービス開発部門で、データマイニング、Twitter検索、「しゃべってコンシェル」、自然対話プラットフォーム（PF）、画像認識PFの開発を手掛けてきました。現在は、音声認識・自然言語処理技術に軸足を移して、①自然対話PF、②自然言語処理AI（人工知能）、③音声DX（デジタルトランスフォーメーション）基盤、④マルチモーダル認識等をテーマに開発を行っています。

自然対話PFは、「しゃべってコンシェル」や「my daiz」で使

用されているAIエンジンであり、シナリオベースの対話機能を中心に、発話内容の意図を機械学習により解釈したり、雑談のやり取りやFAQへの回答、天気情報等の外部情報を自然言語で返すことが可能となるPFです。音声認識、音声合成の機能をAPI（Application Programming Interface）でつなぐことで、非技術者でも簡単に対話エージェントの構築と自然言語による対話が可能となります（図1）。自然対話PFでは、xAIMLという言葉で表現される、さまざまなシナリオをつくり込んでいくことが重要となりますが、これを簡易に行えるようにするために「SUNABA」という開発環境を有するWebサイトも開設しました（<https://docs.sunaba.docomo.ne.jp/>）。「SUNABA」では、マニュアル等をまとめたドキュメントサイトや、xAIMLを開発・実行できるWebオーサリングツール（ボットエディタ）を提供しています。

対話中心の自然対話PFに対して、自然言語処理技術を対話以外の分野に応用することを目的として開発したものが自然言語処理AIです。さまざまなテキストを入力とし、AIがテキストの内容を自動分類したり、自動でのラベリング、あるいは新たに文章を自動生成するような技術です。どのような文書かを自動で判断・

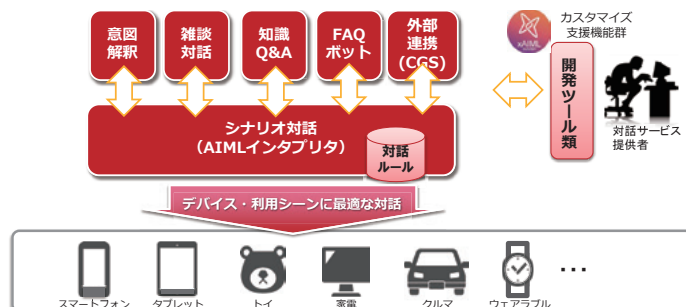


図1 自然対話プラットフォーム

分類する文書分類、個人情報・誹謗中傷を自動で判定し隠すマスキング、固有名詞や業界用語を抽出する固有表現抽出、文章を所定の文字数に要約する文章自動要約等、さまざまなAIが存在し、業務シーンでの用途に応じて使い分けことができます。

音声DX基盤は、話し言葉音声認識、会議・会話分析、非言語情報認識、マルチモーダル認識といった音声処理関連技術アセットを組み合わせて、音声字幕作成機能、議事録作成機能、注釈表示機能、対話可視化機能、反応可視化機能等の業務効率化・価値創造に貢献する機能を実装しています。リモート環境における会議や共同作業、ブレインストーミング、そしてオンライン講義、1 on 1面談、窓口対応、コールセンタ等の対面対応、さらに聴覚障がい者支援等、さまざまな領域への展開を検討しています(図2)。

マルチモーダル認識は、音声だけでなく表情、動作といった動画情報から、対話相手の興味や発話タイミング等を認識する技術です。現在、オンラインコミュニケーションを円滑化できるようなサポートをめざして、対話相手の興味の度合いを認識し、それを分かりやすく表示する「マルチモーダル興味認識」、口の動きに基づく発話の先読みにより、頭切れや音声の遅延が生じない「発話検知技術」について取り組んでいます。

音声認識・自然言語処理技術は幅広く、奥深いんですね。すでに実際のサービス等に組み込まれているのでしょうか。

自然対話PFは、「ドコモAIエージェントAPI」の中に組み込まれています。本サービスは、音声認識・音声合成技術と自然対話PFを活用したSaaS (Software as a Service) 型の音声対話サービスで、法人のお客さまが、スマートフォン、ロボット、家電、Webブラウザ等さまざまなUI (User Interface) を通して自然な対話を行う対話エージェントを構築・展開を可能とするものです。

自然言語処理AIは、「お客さまアンケート・スタッフの声分析」「ニュース記事自動要約」「音声認識IVR (Interactive Voice Response: 自動応答システム)」等において利用されています。

「お客さまアンケート・スタッフの声分析」では、CS (Customer Satisfaction) 部門がさまざまな顧客接点から収集した意

見・要望をAIエンジンにより自動分類・分析し、後段の処理へのエスカレーションや、アンケート中に含まれる個人情報・誹謗中傷を含む文の仕分け・マスキングを行います。

「ニュース記事自動要約」は、NTTドコモ北京研究所との共同プロジェクトとしての取り組みで、要約長さ指定機能、タイトル利用機能、ヒント・マイナスヒント機能、要約の原文引用位置可視化機能等を提供しています。また、事前学習、強化学習、負例を用いた対照学習等を導入し、流暢性や文法、冗長性観点で高い精度を達成するとともに、文圧縮と動的組合せによる要約字数の制御も行います。

「音声認識IVR」は、NTTドコモの総合受付センタから故障受付等の専門センタにIVRにより転送される電話の集中を避けるために、IVR上に自然言語処理AIを導入し、お客さまのご要件から適切なご案内を自動判別し、専門センタへの転送や自動回答によるご案内を実施します。

音声DX基盤は、ファシリテーションを支援しながら、会議参加者の発言率・交差数・質疑回数などを分析するとともに、ブレインストーミングなどにおける活用を想定して、音声認識結果からキーワードを抽出し関連ワードと併せて可視化すること、相手が聞き取りやすい話し方を支援する、話速・音量の可視化等を行う会議・会話分析をめざして、これから具体化に入る段階です。

マルチモーダル認識については、マイクON/OFFを自動制御する「マルチモーダル発話検知」を展示会等に出展し、これから本格化させようとしている段階です。この技術は「発話を先読みしマイクミュート制御をするマルチモーダル発話検知技術」として学会発表も行い、複数の賞もいただきました。

こうした受賞もさることながら、やはり自分が携った技術が世の中に出て多くの人に使ってもらえるということは、非常にうれしくもあり感慨深いものです。以前担当していた「しゃべってコンシェル」は、どうかすると間違ったことを言うこともありました。それでも多くの方々にご利用いただけたのは、大変ありがたかったし、精度を上げようとの思いにもつながりました。また、当時「しゃべってコンシェル」のテレビCMを大々的に放映しており、テレビの俳優さんに話しかけてもらうと出てくる回答が、機械学習モデルの学習内容により、次の瞬間には異なった回答に

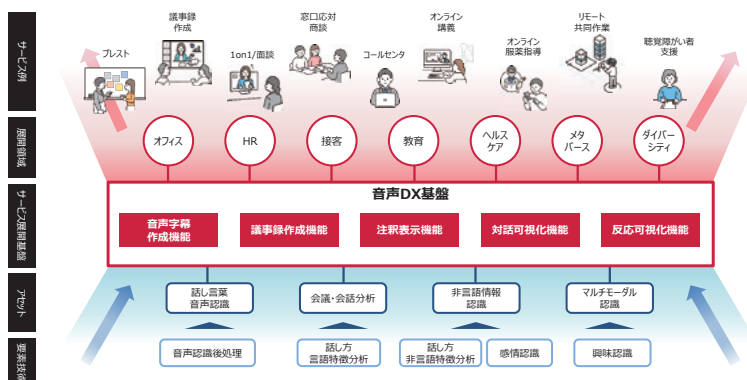


図2 音声DX基盤

になってしまうのではないかとヒヤヒヤしたこともありましたが、とはいえ、それくらい世の中に注目されていたという証でもあり、うれしさもひとしおでした。

ChatGPTと自然言語処理技術の連携

開発者としてスキルの維持、スキルアップはどのようにしていますか。

研究部門在籍のころは、研究を実践する中でスキルの幅と奥行きを広げてきました。ところが、サービス開発部門に異動したときは、分野が異なったうえにマネージャーとしての役割も求められるようになり、当初は技術の専門性はほとんどない状態で、日々の業務を通じたスキルアップもなかなかできる環境にはありませんでした。

一方、チームのメンバーにはそれぞれ得意な技術分野があるので、こうしたメンバーから吸収するとともに、社内でその分野のトップの技術者をめざすように指導しながら、私自身も一緒に勉強してきました。また、これまでの経験から私には広い分野の経験があったので、それによりメンバーを側面支援しながら、チームのスキルアップを図ってきました。こうした努力が次の開発につながったことはいうまでもありません。

私個人としては、外部のセミナー等に積極的に参加しています。幸いなことに情報処理学会の委員や研究会の主査もやっているので、自然言語処理とは関係が薄い研究会ではありますが、やはり見ている面白い技術も多くあり、それがどういう社会課題への取り組みにつながっているのかということも参考になります。もちろん、そこから得られるものがあれば積極的にチームに展開しています。

さらに、こうした机上の知識、理論だけではなく、その技術を具体的に理解し、本質に迫るためにも、自分の手を動かしていくことを実践しています。

こうしたことを繰り返していく中で、自身ばかりではなく、チーム全体のスキルアップを図っています。

事業会社は異動がありますが、こうしたスキルを活かして将来的に何を経験したいのでしょうか。

これまでの経験を活かして、自然言語処理に関する研究開発の道を究めていきたいと思っています。自然言語処理技術とその応用は、大きな広がりを持っており、これにより人間に近い対話ができれば、日常生活の一部としてあらゆるモノ、コトが便利になると思います。例えば、昨今の検索エンジンにはAIが活用されていますが、利用に際してはまだ機械に人が合わせている状態です。これに自然言語処理を活用すれば、人が自然なかたちでコンピュータとコミュニケーションを取れ、誰にとっても使いやすくなると思っています。

さて、最近ChatGPTをはじめとするLLM（Large Language Model：大規模言語モデル）が話題になっていますが、当面これ

に関するテーマに取り組んでいくつもりです。ChatGPTを使うと、回答が毎回変わるとか、間違った回答が出てくる等の話がよくありますが、ChatGPTの良さをうまく使いこなしていくには、使う側のリテラシーに依存している部分が多いと思います。ChatGPT単体だけではなく、その周辺に自然言語処理技術を整備・適用することで、使う側のリテラシー依存度を大幅に下げることができ、たとえ検索利用であってもより自然に近いかたちでコミュニケーションできるのではないかと考えています。タイミング的にも今が1つのチャンスではないかと思っていますので、これを開発のテーマとして取り組んでいきたいと思っています。

そのために、まずはChatGPT等のLLMを多くの人に使ってもらい、それで自然言語でこういったことができるんだということを知ってもらうための環境整備を進めています。併せて、汎用的な言語モデルで分類やマスキング等、どこまで精度を向上できるか、試行錯誤を通じて見極めながら研究開発を進めていきたいと思っています。

時には「技術の無駄遣い」も必要

後進の方やパートナーへのメッセージをお願いします。

いくら良い技術であっても、それだけで実用化に持っていくことは難しいものです。市場のニーズはもちろん、既存技術からのマイグレーションに要するコスト等、周囲の環境を含めて実用化に向けたシナリオをつくっていくことが重要です。もちろん技術そのものを追究する姿勢も重要です。

そして、技術の追究から実用化に向けた流れの中で、時には「技術の無駄遣い」も必要になります。ここでいう「技術」とは、実用化に向けて進んでいる道の脇にある技術と、それとは全く関係のない自分の興味のある分野の技術の両方のことで、実用化に直接関係ない技術に取り組むことで見識を広げることができ、新たな発見や、実用化の方向性確認や行き詰まりの解消に役立つかもしれないので、「無駄遣い」と言いながらも決して無駄にはなりません。マネージャーとして「技術の無駄遣い」を推奨する環境を準備・提供していきますので、ぜひご参加ください。

それから、スキル維持・向上の話の中でも触れましたが、私が携わってきたデータ分析、データマイニング、機械学習等を通して感じたことですが、自分で手を動かしているという経験していくことが重要です。頭の中で考えていたとおりにいかないこともよくあり、自分で手を動かしているとその予兆や原因が肌感覚で分かることが多々あります。

さて、私たちは一緒になって伴走してもらえたいパートナー様を探しています。いきなり確実に使えるものであればおそらく誰かがすでにやっているはずですが、使えるかどうか分からないところからR&Dとして、1つの目標に向けて一緒に伴走していただいて、新しい技術をつくってきたいので、ぜひ一緒にいただきたいと思います。