

デジタルトランスフォーム 先進的メディア処理技術

AI

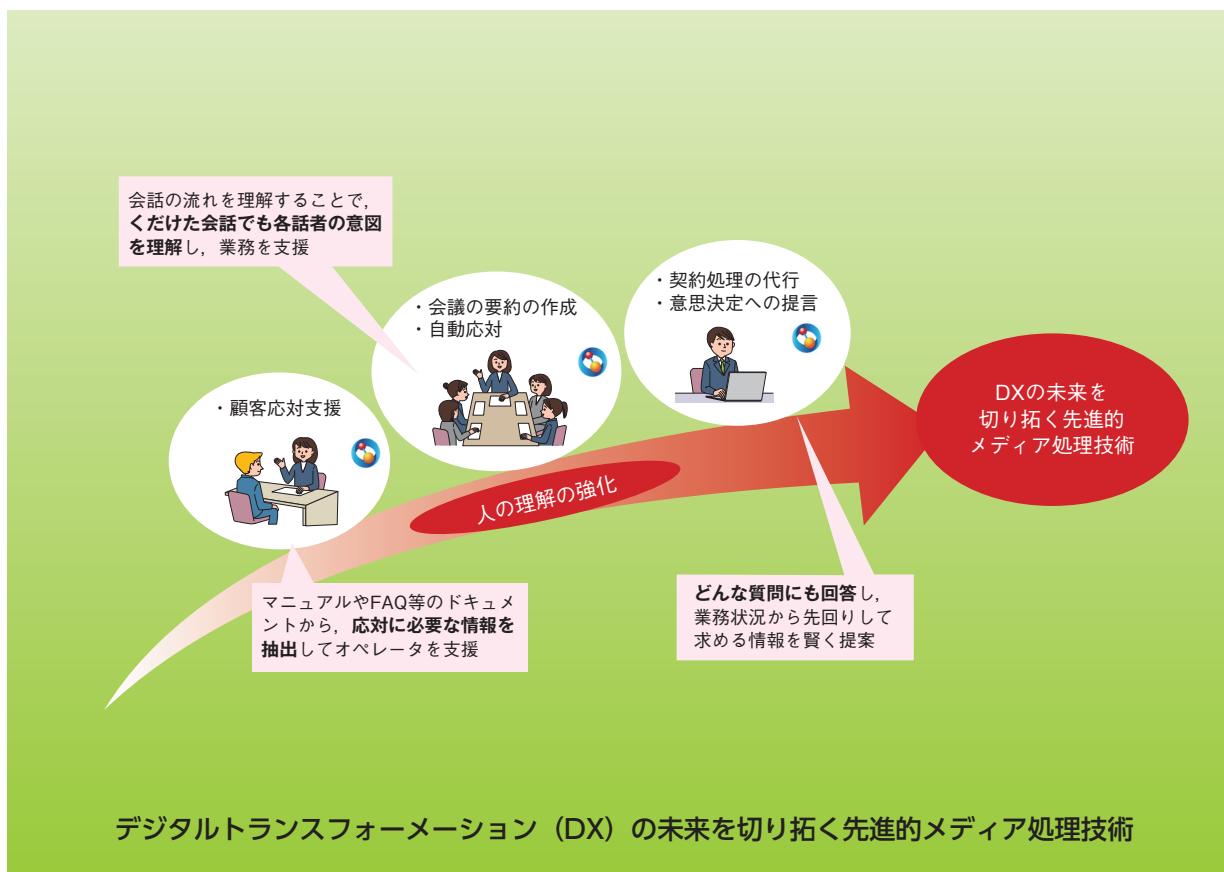
コンタクトセンタ

音声認識

深層学習

FAQ検索

近年、ICTによって既存の仕組みを変革して業務の効率化や新しい価値の創出を行うデジタルトランスフォーメーション（DX）に大きな期待が寄せられている。効果的なDXの実現においては高度なAI（人工知能）技術が必要であり、具体的には、音声処理や自然言語処理といったメディア処理に基づくAI技術が必須となる。本特集では、コンタクトセンタのDXを事例として、NTTにおける先進的なAI技術の研究開発について紹介する。



メーションの未来を切り拓く —コンタクトセンタAI

■ コンタクトセンタAIの先進的取り組み

6

音声・自然言語処理技術を活用した、コンタクトセンタのさまざまな課題を解決するための最新技術について紹介する。

■ 進化を続ける音声認識エンジン「VoiceRex®」

9

音声認識エンジン「VoiceRex®」の歴史、および最新のエンジンに導入されているコンタクトセンタでの活用が期待される技術について紹介する。

■ 機械読解による自然言語理解への挑戦

12

マニュアルに書かれた業務知識を理解し、お客様の言葉を理解して質問に適切に応答するエージェントの実現をめざして研究開発を進めている機械読解技術について紹介する。

■ オペレータの応対を支援する自動知識支援システム

16

業務経験の浅いオペレータを支援するために、応対中のオペレータに適切な情報(知識)を自動的に提示する自動知識支援システムを紹介する。

主役登場

西田 京介 (NTTメディアインテリジェンス研究所)

20

機械読解のState of the Artをめざして

コンタクトセンタAIの先進的取り組み

コンタクトセンタは、多くのお客様の声が集まる顧客接点としての重要性が増しています。NTTメディアインテリジェンス研究所では、AI（人工知能）の適用先に、コンタクトセンタを注力分野の1つとして研究開発を進めています。本稿では、私たちが長年培ってきた音声・自然言語処理技術を活用した、コンタクトセンタのさまざまな課題を解決するための最新技術について紹介します。

はじめに

第三次AI（人工知能）ブームが産業界で本格化してから3年以上が経過し、さまざまな分野でAIの実利用が始まっています。NTTグループでは2016年6月にAIブランドcorevo[®]を発表しさまざまな取り組みを進めています⁽¹⁾。第三次AIブームをもたらした主要な技術は、機械学習、とりわけ深層学習（ディープラーニング）であることは疑いようがありません。深層学習は基本技術としてさまざまな問題に適用することができますが、単に適用するだけでは実際のビジネスで利用できるだけの性能が得られないことがほとんどです。適用する問題に特化したネットワークモデルを創出し、適切なデータを用いて学習することで、初めて実利用に足りる性能を発揮することができます。

初期のAIの議論に汎用型AIと特化型AIというものがあります。汎用型AIは人間の頭脳のように1つのAIであらゆる問題を解こうとするのですが、まだまだ実現には遠い状況です。一方、特化型AIは特定の問題のみを解くことができるAIで、現在のAIはすべて特化型といえます。問題を特化

することで、人間並、あるいは人間以上の性能を発揮するまでになってきています。つまり、現在のAIをうまく使っていくには、適用分野を定めて、問題に適応していくことが重要になります。

NTTメディアインテリジェンス研究所では、AIの適用先としてコンタクトセンタを注力分野の1つとして研究開発を進めています。コンタクトセンタは多くの電話やチャット応対が日々行われ、多くのナレッジ検索や通話分析が行われています。つまり、私たちが長年培ってきた音声・自然言語処理技術を活かすことができるフィールドの1つといえます。NTTグループには多くのコンタクトセンタがあり、技術をブラッシュアップするための多くのフィールドとデータが存在します。これらの蓄積と環境を活かして、実ビジネスに適用可能な技術の創出をめざしています。

コンタクトセンタの課題

コンタクトセンタは、お客様への対応業務を電話、電子メール、チャットなどのさまざまなチャネルを通じて行なうことを専門とする部署です。もともとは、申込受付やお客さまサポート

たなか きみひと^{†1} やぎ たかし^{†1}

田中 公人 /八木 貴史

いいづか てつや^{†2}

飯塚 哲也

NTTメディアインテリジェンス研究所^{†1}

NTTメディアインテリジェンス研究所 所長^{†2}

を行う運営拠点として位置付けられていましたが、近年のビジネス環境の変化により、多くのお客様の声が集まる顧客接点としての重要性が増しています。国内コンタクトセンタを取り巻く環境は、深刻化する人手不足に悩まされており、オペレータの採用や定着化が大きな課題となっています⁽²⁾。また、応対品質、すなわちCX（Customer Experience）の向上も同時に求められており、特に人の入れ替わりの激しいセンタでは、少ない人数・初級者の多いメンバという限られた運営リソースの中、いかにCXを高めるかがコンタクトセンタ・マネジメントにおける大きな課題となっています。

コンタクトセンタにおけるAI

コンタクトセンタには、IVR（Interactive Voice Response）、CTI（Computer Telephony Integration）、ナレッジシステム、CRM（Customer Relationship Management）など、多くのITシステムが導入されています。AIはこれらのシステムを高度化するとともに、新たな機能を提供し、さらなる運営の効率化やCX向上を図ります。

コンタクトセンタにおけるAIの代表格として、音声マイニング技術が挙

げられます。音声マイニング技術は、深層学習の適用による音声認識精度の飛躍的な向上により、実用の領域に入り、コンタクトセンタにおいてもっとも導入が進んでいるAIの1つです。NTTでは、NTTテクノクロスを通じて、音声マイニング技術 (ForeSight Voice Mining: FSVM⁽³⁾) の商用展開を行っています。音声マイニング技術は、「通話音声を音声認識によりテキスト化」し、「大量通話の中からコンタクトセンタ運用の課題やその解決方法のヒントを得るための統計分析・見える化」を実現します⁽⁴⁾。統計分析を行うことで、例えば高スキルオペレータの応対ノウハウを抽出し、センタ内に水平展開して全体のスキル向上を図ったり、センタ全体の統計情報を一覧表示することで、異常な事象（例えば、特定内容の問合せの急増）を即座に把握することができます。また音声がテキスト化されているので、通話が

終わった後に会話の詳細を素早く見直すことができ、応対履歴の投入作業が早く・正確に行えるようになります。これらの機能はすでにNTTグループ内外のさまざまなコンタクトセンタで活用されており、業務効率化や売上向上などの成果を達成しています。

音声マイニング技術の構成例を図に示します。音声マイニング技術では、入力された通話音声データに対して音声認識が行われ、その結果は通話分析部にてアーリストが所望する分析に利用されます。音声マイニングは、コンタクトセンタにおけるAIのプラットフォーム的な役割を担い、他のAIモジュールを追加することによって音声認識結果をより高度に活用する機能を追加できます。NTT研究所では、音声通話から感情などのテキスト以外の情報を抽出するAIの研究開発を行っています。これらのAIにより音声認識部を拡張することで分析の幅を広げ

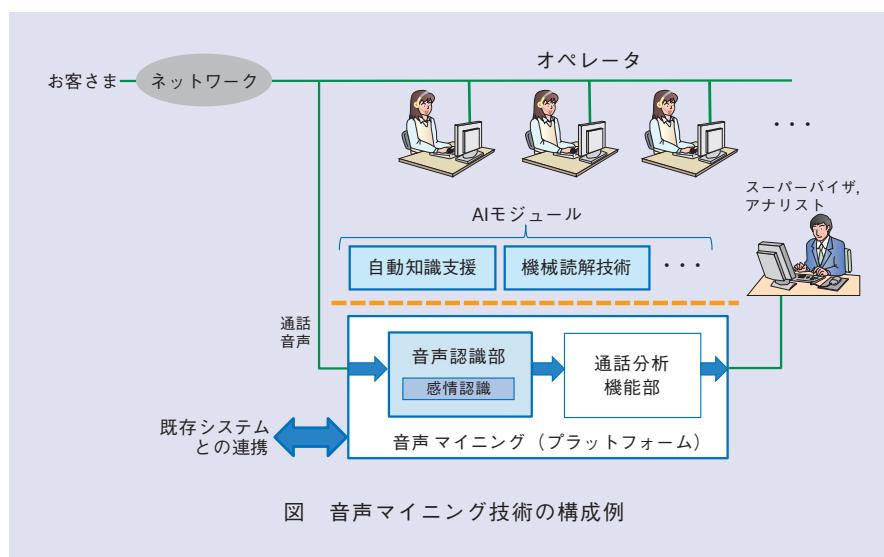
ることができます。また、オペレータの応対支援に向けて、音声認識結果のテキストを用いてナレッジ提示や生成を行うAIの研究開発を進めています。これらのAIは、音声マイニング技術をプラットフォームとして活用して構築できるように構成されています。これにより、早期にコンタクトセンタの現場で動作させることができ可能になっています。

さらなる高度化に向けて

音声マイニング技術の商用展開を通じて、新たな課題がみえてきています。

- ① お客様音声の認識精度のさらなる向上
- ② 音声認識結果を活用したさらなる価値創造
- ③ 導入時のAIチューニング等のコスト削減・時間短縮

現在、コンタクトセンタにおけるオペレータ音声の認識精度は90%を超えます。しかしながら、お客様音声の認識精度はこれよりおおむね10ポイント程度低くなります。お客様音声は、話す場所によっては背景ノイズが大きく、携帯電話であれば符号化による音質劣化などがあります。また、オペレータと比較すると発声が不安定で話し方もくだけており、音声認識にとって困難な条件が多くなります。現在のお客様音声の認識精度は、認識結果を人間が見て理解するという用途や、単語レベルの統計分析を行うなどの用途では利用可能なレベルにあるといえます。しかしながら、認識結果テ



キストをナレッジ検索や生成など、高
度に活用していくためには、より正確
な認識が求められています。本特集記
事『進化を続ける音声認識エンジン
「VoiceRex®」』では、①を解決するた
めの最新技術について紹介します。

音声マイニング技術の導入に伴い、
音声認識結果をより高度に活用した、
さらなるオペレータの支援の高度化や
コスト削減の要望が増しています。②
に対して、NTT研究所では、オペレー
タ応対に必要なナレッジの獲得を容易
にするナレッジ支援に関する研究開発
に力を入れています。

本特集記事『機械読解による自然言
語理解への挑戦』では、質問に対する
答えをマニュアルなどから導き出す技
術について紹介します。コンタクトセ
ンタで用いられるナレッジには、
FAQ (Frequently Asked Questions)、
マニュアル、約款などがありますが、
FAQが十分に整備されていないコン
タクトセンタが少なからず存在しま
す。機械読解は、FAQ検索のよう
に事前に質問と回答のペアを準備する
必要がなく、ナレッジ整備コストの削減
が期待されます。また、オペレータへ
のインタビューによると、「普段から
研修などで使っているマニュアルのほ
うがFAQよりも慣れていて使いやす
い」「FAQに記載のない問合せは、結
局マニュアルを見なくてはいけない」
といった意見もあり、オペレータの利
便性という観点でも期待が高まっ
ています。

本特集記事『オペレータの応対を支

援する自動知識支援システム』では、
音声認識とFAQ検索の連携により、
適切なタイミングで適切なFAQを、
オペレータが応対中に自動提示する自
動知識支援に関して紹介します。自動
知識支援はオペレータがナレッジを調
べる時間を削減します。また、音声認
識結果を用いたFAQ整備作業の効率
化に関する取り組みについても合わせ
て紹介します。

音声認識などのAIは、適用する現
場ごとにチューニングを行うことで精
度を高めることができますが、特に小
規模なコンタクトセンタなどへの導入
に向けては、チューニングに必要とな
るコストの削減、時間の短縮が重要と
なります。③について、NTT研究所
では新技術創出による精度向上と合
わせて、ベースとなるAIモデルの強化
による現場ごとのチューニング作業な
どの軽減にも力を入れています。これ
までに、通話音声などの学習用データ
を大量に収集し、音声認識やナレッジ
検索の業界別ベースモデルを作成して
います。業界別ベースモデルを構築し
た主要な業界では、少ない学習データ
で追加学習するだけで、短時間・安価
に高い精度を実現することができるよ
うになっています。

AIの利用シーン拡大に向けて

深層学習の出現やコンピュータの性
能向上による大量データ学習等によ
り、音声認識や自然言語処理の精度は
飛躍的に向上し、ビジネス現場への普
及が拡大しつつあります。今後は、本

稿で紹介した技術をはじめ、さまざま
なメディア処理技術を活用してパート
ナー様とビジネスにおける成功事例を
積み上げるとともに、AIの利用シー
ン拡大に向けてさらなる革新的な技術
の研究開発を推進していきます。

■参考文献

- (1) <http://www.ntt.co.jp/corevo/>
- (2) 月刊コールセンタジャパン編集部：“コール
センタ白書2018”，リックテレコム，2018。
- (3) https://www.ntt-tx.co.jp/products/foresight_vm/
- (4) 河村・町田・松井・坂本・石井：“コールセ
ンタにおけるAIの活用。”NTT技術ジャーナ
ル，Vol.28，No.2，pp.35-37，2016。



(左から) 飯塚 哲也/ 田中 公人/
八木 貴史

新しい技術は時として既存の手段よりも
劣っている部分が目につくことがあります
が、うまく使いこなした企業は、技術の進
化とともに競争優位を獲得します。ぜひコ
ンタクトセンタAIをご活用ください。

◆問い合わせ先

NTTメディアインテリジェンス研究所
知識メディアプロジェクト
TEL 046-859-5305
E-mail kimi.tanaka@hco.ntt.co.jp

進化を続ける音声認識エンジン「VoiceRex®」

コンタクトセンタ向けのAIを支える基盤技術の1つに音声認識があります。今や多くの場面で利用されるようになり、ビジネスをさまざまなかたちで支えています。その研究開発の歴史は長く、さまざまな技術を積み重ね、現在の水準に至っています。NTTメディアインテリジェンス研究所で開発する音声認識エンジン「VoiceRex®」の歴史を軸にこれらを紹介し、最新のエンジンに導入されているコンタクトセンタでの活用が期待される技術について紹介します。

音声認識エンジン「VoiceRex®」の歴史

コンタクトセンタ向けのAIを支える基盤技術の1つに音声認識技術があります。音声認識技術とは入力信号の中に含まれる音声をテキスト化する技術です。NTT研究所における音声認識の研究開発の歴史は長く、半世紀程度に及びます。NTTメディアインテリジェンス研究所では、こうした長年にわたる研究成果を基に、幅広いサービス分野に適用可能なものとして音声認識エンジン「VoiceRex®」を開発し、グループ各社に提供しています。

音声認識をコンタクトセンタ通話の分析に用いるというアイデアは研究開発の当初から存在しました。当時のレベルでは夢のまた夢でありましたが、何十年後かに到達すべき目標でもありました。VoiceRex®（の前身となる音声認識ライブラリ）は、1990年代に入って初めてリリースされましたが、当時のものはキーワードを認識することしかできないものでした。それから現在のように会話等の長い発話を認識可能になったのは2000年のことでした。とはいって、その性能は人間どうしの会話を認識しようとしても到底正し

く認識できないレベルでした。新聞のような書き言葉を、はきはきとした口調で読み上げたような音声でなければ正しく認識できず、認識可能な単語数も極めて限られたでした。

そこから「VoiceRex®」は何段階かの技術革新を経て、性能が飛躍的に向上してきたという経緯があります。まず2008年、日本で初めてのWFST（Weighted Finite State Transducer）と呼ばれる技術を採用しました。従来よりも約100倍の言葉を覚えられるようになり、およそ1000万語の中から最適な単語を選ぶことが可能になりました。この進化は2009年の衆議院での議会録作成システムへの採用につながりました。議会は1人がしゃべる一問一答形式であり、こういう場では90%程度の高い認識性能を達成することができました。本会議場や各委員会等で、人手で行う速記に代わるものとして存在を知らしめることになったのです。

その後も、音声データベースの拡張と整備の進展や、計算機能の向上を背景とした大規模なデータベースを効果的に活用する技術の創出などがあり、音声認識の性能は徐々に向上を続けてきました。そして、コンタクトセンタの通話の音声認識はいよいよ現

お お ば たかのぶ た な か と も ひ ろ
大庭 隆伸 / 田中 智大

ますむら りょう
増村 亮

NTTメディアインテリジェンス研究所

実の問題設定となり、2014年、NTTソフトウェアから「ForeSight Voice Mining」*というコンタクトセンタ向けの製品が発表されることになりました。

さて、その数年前、音声認識の研究者コミュニティの中では、ある技術が注目を集めていました。深層学習（ディープニューラルネットワーク）の登場です。深層学習は音声認識にとって大きなパラダイムシフトとなりました。音声信号、すなわち空気振動を「あいうえお」という音の並びに直す音響モデルの性能が極めて高まり、これにより通話の認識率も飛躍的に向上しました。深層学習を採用したVoiceRex®は2014年に商用リリースされています。さらに2015年にニューラルネットワークの一種CNN-NIN（Convolutional Neural Network - Network In Network）を採用し、騒がしい公共エリアでモバイル端末を使った音声認識を行う、「CHiME3」という技術評価国際イベントで参加研究機関中1位を獲得しました。スマートフォンの普及により、外で電話をする機会が増えましたが、こうした人混みなどの周囲雑音がある音声信号に対しても精度良く認識ができるようになりました。

* ForeSight Voice Miningは現在NTTテクノクロス株式会社より販売されています。

こうした技術革新を経て音声認識の用途は急速に拡大しました。今やVoiceRex[®]は多くの商品・サービスで利用されています。特にコンタクトセンタ向けAI関連の商品導入事例は急速に数を増やしており、音声認識を用いた商品・サービスの中でも中核的な存在になってきています。

一方、数多くのお客さまにご利用いただくことで、新たな課題も生まれてきています。その1つは話題の多様性です。現在の音声認識技術では、入力される通話の話題が既知であると精度面では有利に働きます。例えば、電話窓口を提供している企業ごとに取り扱っているサービス名は違っています。また、同じ会社の電話窓口でも入会・退会を受け付ける窓口、苦情を受け付ける窓口、技術的な質問に答える窓口などに分かれています。そのため、現状では会社ごと、もしくはコンタクトセンタの単位で、話題に関するモデル（言語モデル）を個別にチューニングする作業を行っています。

もう1つの課題は、話の流暢さへの対応です。人間どうしの会話では明確に発音しないという現象が頻発します。例えば、文末の「～します」をアルファベットで書くと「shimasu」ですが、「shim」くらいまでは発音するが、残りの「asu」は音はあるものはっきりとは発音されず、リズムだけ表現するといったことが頻発します。こうした音の不明瞭さ（場合によっては発音をしていない現象）に対しては、言葉の文脈的な観点から予測するような枠組みが必要になります。以降では、最新版VoiceRex[®]に搭載されている、これらの課題を克服するための新しい技術について紹介します。

会話コンテキスト言語モデル

言語モデルとは単語のつながりを予

測するモデルです。直感的には文が日本語として正しいかどうかを判定するといった役割を果たします。音声認識エンジンには、ある単語は直前のN-1個の単語に依存すると考えてモデル化を行うN-gram言語モデルと呼ばれる確率モデルが利用されています。Nが大きいと単語の組み合わせが爆発的に増加することからNはたかだか3～4程度です。そのため、局所的なコンテキスト（文脈情報）のみを考慮したモデル化となってしまいます。短い発話であればN-gram言語モデルでも十分ですが、発話が長くなればなるほど、より長いコンテキストを考慮することが必要になるのです。

そこで近年、ニューラルネットワークを用いた言語モデルが注目を集めています。特にリカレントニューラルネットワーク（RNN）言語モデルと呼ばれる長距離のコンテキストを扱うことが可能なモデルが注目を集めています。音声認識で用いる際は、一度音声認識をして得られた複数の音声認識結果候補の文に対して、RNN言語モデルで算出したスコアを加味することで、最終的な認識結果文を決定するという使い方をします。これをリスクアーリング法と呼びます。

さて、RNN言語モデルの問題点の1つは発話単位のコンテキストの利用にとどまっている点です。コンタクトセンタの通話では、発話をまたぐコンテキストの利用は極めて重要です。例えば、オペレータがお客さまの問合せに対して回答することを考えると、その回答内容はお客さまの問合せ内容に関連する内容のはずです。そこで私たちはそのような発話をまたぐ長期的なコンテキストを考慮した、会話コンテキスト言語モデル技術⁽¹⁾を開発しました。実際に音声認識する際には、逐

次的に前の発話の音声認識結果をコンテキストとして利用します。VoiceRex[®]では、会話コンテキスト言語モデルを発話ごとにリスクアーリング法により適用することで、より良い文（正解に近い文）を選び直しながら、それをコンテキストとして与え、次の発話の音声認識を行うことができる実装になっています。そのため、各発話の入力ごとに、会話コンテキスト言語モデルの恩恵が得られるのです。

ニューラル誤り訂正言語モデル

話の流暢さにより明瞭に発音されない音があることは述べましたが、そこには一定の傾向があります。前述のような「～します」といった文末表現、助詞、「ありがとうございます」といった頻出表現などで多くみられ、それぞれで音声認識は毎回同じような誤りを起こしてしまいかがちです。発音のあいまいさに起因する事象以外にも、頻出の誤りパターンがいくつか存在します。こうした誤りの偏りを見つけて、まるごと修正するアプローチを誤り訂正と呼びます。

私たちはRNN言語モデルに対して音声認識誤りを考慮できる枠組みを取り入れたニューラル誤り訂正言語モデル技術⁽²⁾を開発することで、認識精度の向上に成功しました。具体的にはEncoder-Decoderモデルと呼ばれるニューラルネットワークを導入し、音声認識誤りを含む音声認識結果から正解文を推定するといった機構を与えました。ニューラル誤り訂正言語モデルの学習には、一度音声認識して得られた音声認識誤りを含む文と、それに対応する正解文を用意して、前者と後者の関係を学習させます。これにより誤りの傾向とそれを修正する方法を同時に獲得できるのです。

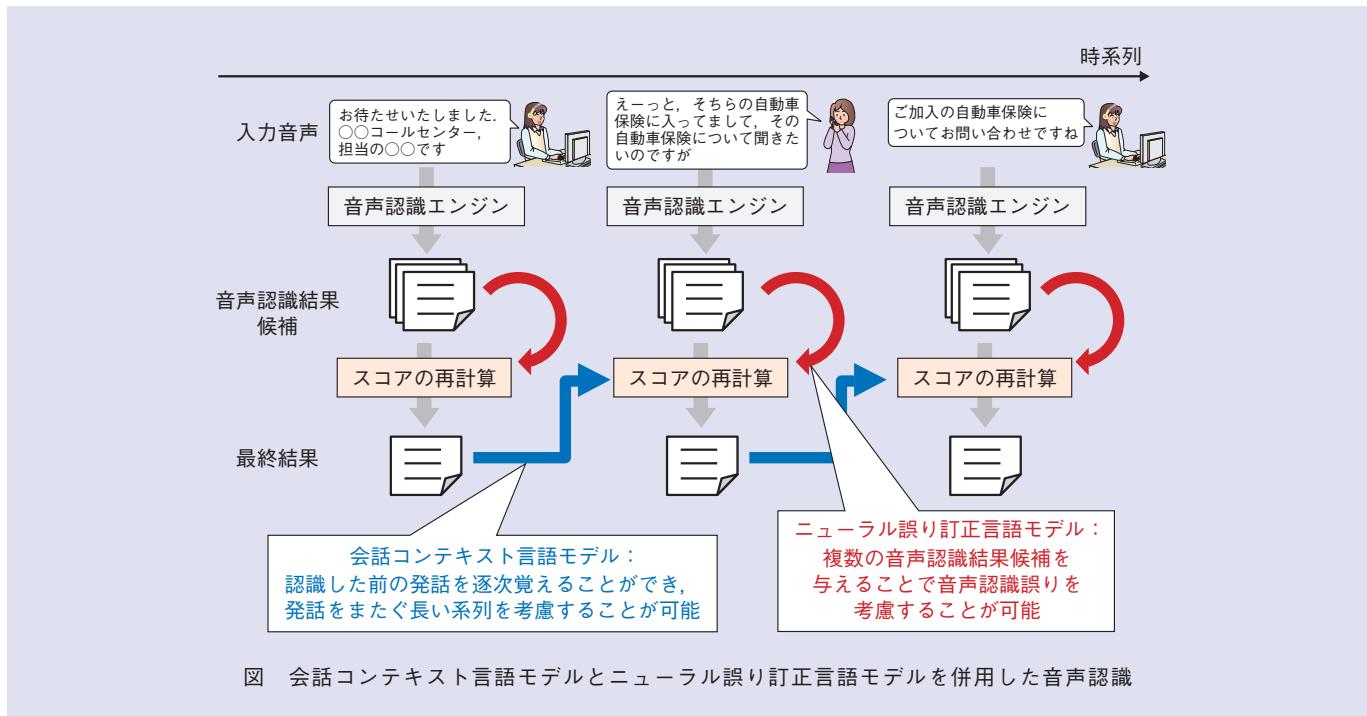


図 会話コンテキスト言語モデルとニューラル誤り訂正言語モデルを併用した音声認識

実際に音声認識を行う際には、会話コンテキスト言語モデルと同様にリストアリング法を適用します。会話コンテキスト言語モデルと併用する事が可能で、両言語モデルの恩恵を受けた認識結果を発話ごとに得ることができます。この様子を図に示します。通話の中で入力信号から発話区間が切り出されるたびに、認識結果候補を複数出力し、両言語モデルを適応して最終的な認識結果を求めます。会話コンテキスト言語モデルに関しては、その最終的な認識結果を次の発話を認識する際のコンテキストとすることを繰り返していくかたちをとっています。

今後の展開

これまで対象としていたコンタクトセンタの多くは、お客様のほうからコンタクトセンタに通電するという形式がほとんどでした。その場合、オペレータとお客様は1度きりの会話をするので、お互いに比較的丁寧に話をしようとします。それは音声認識精

度の観点でも都合の良い現象といえます。しかしながら、コンタクトセンタのAI関連製品の導入が進むにつれ、会社側からお客様に連絡をするとといったシーンにも用いられるようになってきました。この場合、お客様ごとに担当者がつき、複数回にわたって同じ人どうしで会話をを行うため、よりフランクな会話が行われており、音声認識の観点では課題となっています。これまで、音声認識は適用先を拡大にするのに伴い、従来よりも認識の難しい状況に直面してきました。そのたびに、それを打開するための研究開発を行うことで進化を続けてきました。今後も、実際に音声認識をご利用いただいている方々が直面している、新たな課題に挑戦しながら、VoiceRex®は進化を続けていくでしょう。

■参考文献

- (1) R. Masumura, T. Tanaka, A. Ando, H. Masataki, and Y. Aono : "Role Play Dialogue Aware Language Models Based on Conditional Hierarchical Recurrent Encoder-Decoder," Proc. of Interspeech 2018, pp.1259-1263, Hyderabad, India, Sept. 2018.

- (2) T. Tanaka, R. Masumura, H. Masataki, and Y. Aono : "Neural Error Corrective Language Models for Automatic Speech Recognition," Proc. of Interspeech 2018, pp.401-405, Hyderabad, India, Sept. 2018.



(左から) 増村 亮/ 田中 智大/
大庭 隆伸

音声認識は多くのビジネス案件で活用されており今後も拡大が期待されます。そこで、いくつかのグループ各社を通して、音声認識をお試しいただける環境を提供しています。ぜひ「NTT 音声認識」と検索してみてください。

◆問い合わせ先

NTTメディアインテリジェンス研究所
音声言語メディアプロジェクト
TEL 046-859-2943
FAX 046-855-1054
E-mail takanobu.oba.ec@hco.ntt.co.jp

機械読解による自然言語理解への挑戦

AI（人工知能）の文章読解力に注目が集まっています。AIに人間が用いる言葉（自然言語）を理解させることはとても難しい課題ですが、近年では深層学習の適用により大幅にAIの読解力が向上しました。本稿では、マニュアルに書かれた業務知識を理解し、お客様の言葉を理解して質問に適切に応答するエージェントの実現をめざして研究開発を進めている機械読解技術について紹介します。

にしだ きょうすけ
西田 京介 / 斎藤 いつみ
おおつか あつし
大塚 淳史 / 西田 光甫
のもと なりか
野本 済央 / 浅野 久子

NTTメディアインテリジェンス研究所

機械読解と自然言語理解

コンタクトセンタにおけるお客様応対をAI（人工知能）技術により支援するために、NTTメディアインテリジェンス研究所（MD研）では機械読解について研究開発に取り組んでいます。機械読解とはAIがマニュアルや契約書などのテキストを読んで質問に応答する技術で、人間が日常的に用いる言葉を理解する「自然言語理解」の実現に挑戦するものです。コンタクトセンタでは、検索用にマニュアルからFAQを事前に準備しなくとも、マニュアルに書かれた内容を正確に読み解いてピンポイントに回答を発見する

ことでオペレータを支援することをめざしています（図1）。

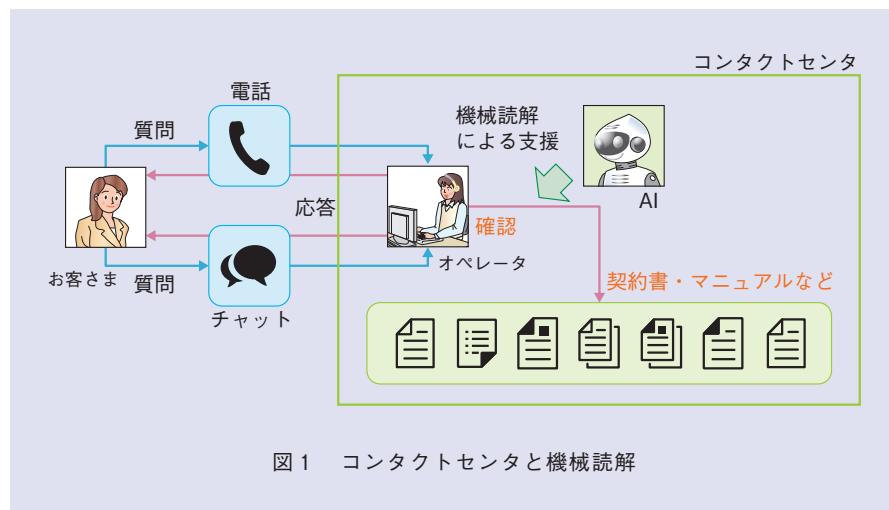
機械読解は新しい研究分野ですが、深層学習の発展および大規模なデータセットの整備により急速に発展しています。特に、スタンフォード大学が作成した機械読解データセットSQuAD⁽¹⁾において、2018年1月にAIが人間を上回る回答精度を達成したことで、この研究分野に大きな注目が集まりました。しかし、SQuADの問題設定は比較的単純であり、より難しい問題設定においてAIはまだ人間の読解力には及びません。MD研では、研究用のデータセットを用いた学術的な競争の中で技術を磨きながら、コンタクトセンタ

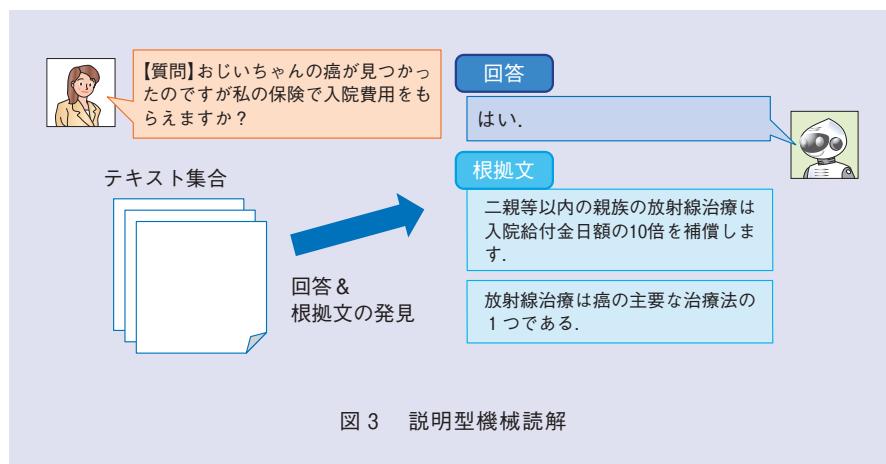
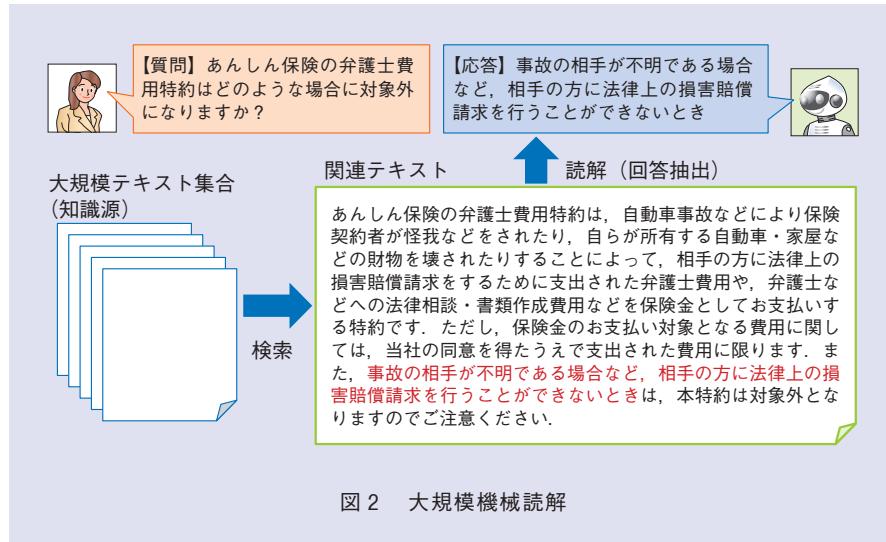
における実用化に向けた課題の解決に取り組んでいます。本稿では、私たちの機械読解に関する研究成果について紹介します。

多数のテキストから回答を発見する「大規模機械読解技術」

初期の機械読解研究の問題設定では、知識源として扱うテキストは1つに限定されていました。しかし、コンタクトセンタをはじめとした実際の応用シーンでは、多数のテキストから質問に対する回答を発見する必要があります。すべてのテキストを機械読解モデルにより詳細に読み解くのではシステム動作が遅くなってしまうので、回答に必要な文書を高速かつ正確に絞り込まなければなりません。

そこでMD研では、大規模機械読解技術を確立しました^{(2), (3)}。この技術は、まず高速に動作するキーワード検索技術を用いて質問に関連するテキストを大まかに絞り込み、次にニューラル検索モデルにより正確にテキストを絞り込み、最後にニューラル読解モデルにより回答を発見します（図2）。この際、ニューラル検索・読解を同時に1つのモデルで学習することで、検索精度が大きく向上しました。これに





より、英語Wikipediaの500万記事を知識源とした機械読解タスクにおいて、発表当時世界最高の回答精度を達成することができました。

複数箇所に書かれた根拠を理解し提示する「説明型機械読解技術」

機械学習に基づくAIでは回答が正しいことを100%保証することは難し

いため、機械読解モデルが提出した回答の妥当性を人間が確認できることは重要な機能となります。人間が回答の妥当性を確認するためには、回答の根拠となる情報がテキスト中のどこに書いてあるのかを機械読解モデルが提示することが有用となります。しかし、回答の根拠となる情報はテキスト中の複数の文に分かれて記述されている

ケースがあるため、これらすべてを理解し抽出することが重要な課題として知られています⁽⁴⁾。

そこでMD研では、文書要約の技術を拡張したモデルを機械読解モデルと組み合わせることによって、説明型機械読解技術を確立しました^{(5), (6)}(図3)。文書要約技術ではテキスト中の重要な文を要約として抽出しますが、提案手法では回答のために重要な文を根拠文として抽出します。このニューラル根拠抽出モデルが根拠文を抽出すると同時にニューラル読解モデルが回答を発見することで、根拠が複数の文に分かれてしまうような難しい質問に対しても、回答・根拠双方を精度良く提示することを実現しています。本技術によって、英語Wikipedia記事への質問に対して回答とその根拠を出力する機械読解タスクHotpotQA⁽⁴⁾において、発表当時リーダーボードの1位を獲得することができました。

質問があいまいな場合に問い合わせる「改定質問生成技術」

機械読解の研究用データセットでは回答が一意に特定できるように作成されたものが主流ですが、コンタクトセンターなどで実際にお客様が入力する質問では、質問の意図があいまいで、回答を特定できない場合があります。このような場合に人間のオペレーターのようにAIが適切に質問の意図を問い合わせることができれば、より自然なコミュニケーションが実現できます。

そこでMD研では、あいまいな質問を機械読解が回答できる具体的な質問に書き換える改訂質問生成技術を確立しました^{(7), (8)}(図4)。提案手法では、質問に対する機械読解の回答の候補を抽出し、回答候補ごとに質問のあいまい性を解消した改訂質問を生成します。例えば、「弁護士特約の限度額はいくら」という質問をした際、改訂質問生成では、回答を直ちに提示するの

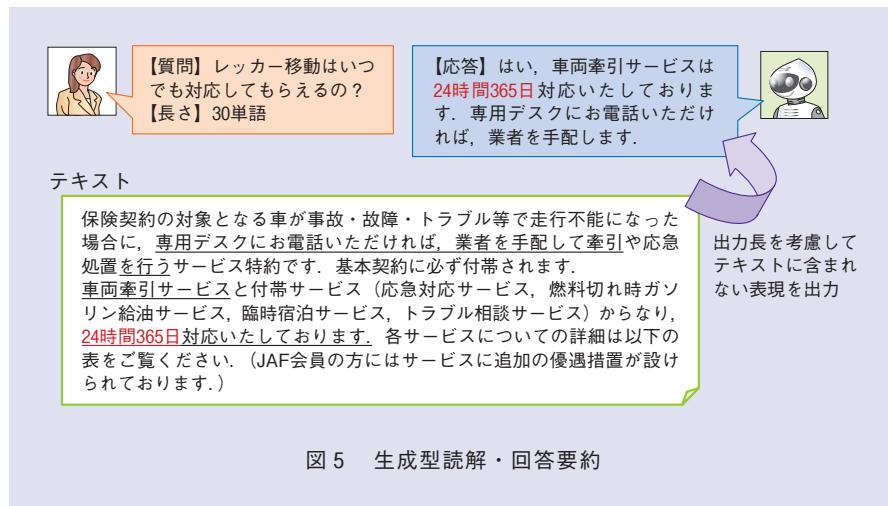
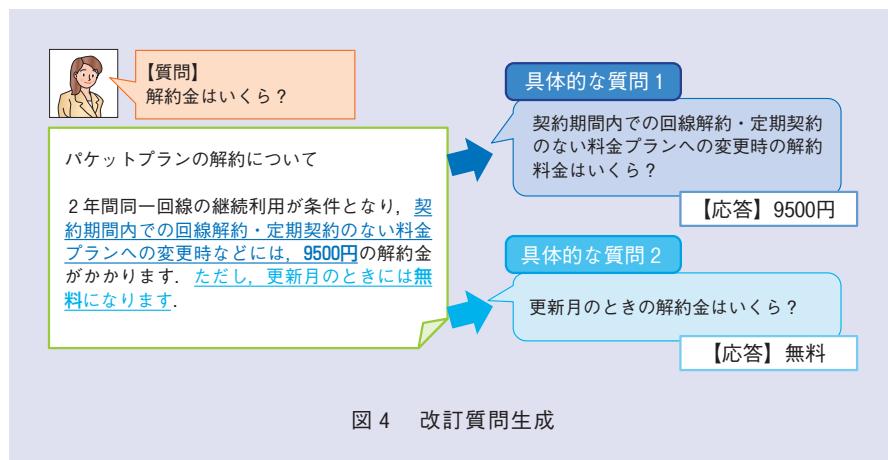
ではなく、「1回の事故による弁護士特約の保険金の限度額はいくら」と「法律相談・書類作成についての弁護士特約の限度額はいくら」といった複数の改訂質問を生成し、質問者に自身が聞きたいことに近い質問がどちらであるかを選択させるように問い合わせます。この例では、質問者が前者を選択した場合の回答は「300万円」、後者は「10万円」といったように、選択した改訂

質問によって回答が変わります。これにより、従来よりも多様な質問に対し機械読解を適用することができます。

スタイルを指定して回答を生成する「生成型機械読解技術」

現在の機械読解の問題設定は、回答を知識源となるテキストから抽出することが主流です。しかし、スマートデバイスやチャットボットでは、より自然な文章を生成することが求められます。しかし、回答の生成は難易度が高く、さらに学習データが不足しているため、これまで世界的に研究は進んでいませんでした。

そこでMD研では、テキストに書いていない表現を生成して応答可能な生成型機械読解技術を確立しました^{(9), (10)}(図5)。例えば、「レッカー移動の受付時間は?」と質問した際に、「レッカー移動の受付時間は24時間365日です」という、テキストのみでなく質問の内容を適切に含めた自然な応答を作成できることが特徴です。また、提案手法は、複数の異なる回答スタイルの機械読解データを同時に学習し、生成時に回答スタイルを選択できるようになりました。これにより、学習データが不足する問題を解決しています。提案手法は、検索エンジンの実ログを用いたオープンドメインなQAを行うMS MARCO⁽¹¹⁾の回答スタイルの異なる2タスクにおいて、発表当時リーダーボードの1位を獲得することができました。



回答の長さを制御する 「回答要約技術」

お客様からの質問に応答するチャットボットにおいては、機械読解モデルが長い回答を出力した場合にそのまま提示するとお客様が読み難いため、回答の長さを適切に調整することが望れます。また、スマートフォンで回答を読む場合とPCで回答を読む場合など、状況に応じて回答の長さを適切に調整することで、よりお客様や状況に合わせた柔軟な回答が可能になると考えられます。

そこでMD研では、ニューラルネットを用いて長さをコントロール可能な回答要約技術を確立しました⁽¹²⁾(図4)。本技術では、テキスト中の重要な個所を特定するニューラルネットのモデルと、特定した個所から回答を要約(生成)するニューラルネットのモデルを組み合わせることによって質問に対して適切に回答を要約する機能を実現しています。回答を要約する際に、長さの情報をベクトル化して与えることによって、任意の長さで回答を出力することが可能となります。

今後の展開

機械読解技術をコンタクトセンタにおけるマニュアルを対象とした質問応答に適用する中で実用化に向けた課題を発見し、さらに技術改善にフィードバックしていくことで、AIによるオペレータの支援、さらには、自動応対

の実現をめざしていきます。機械読解は人間が用いる自然な言葉の理解という難しい挑戦であり、コンタクトセンタをはじめ、NTTグループのさまざまなエージェントAIサービスにおいてイノベーションを起こせると考えています。MD研では、人間と自然にコミュニケーションが可能なAIの実現をめざして、今後も研究開発に取り組みます。

■参考文献

- (1) P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang : "SQuAD: 100,000+ questions for machine comprehension of text," Proc. of EMNLP 2016, pp.2383-2392, Austin, U.S.A., Nov. 2016.
- (2) 西田・齊藤・大塚・浅野・富田：“情報検索とのマルチタスク学習による大規模機械読解,” 言語処理学会第24回年次大会, pp.963-966, 2018.
- (3) K. Nishida, I. Saito, A. Otsuka, H. Asano, and J. Tomita : "Retrieve-and-read: Multi-task learning of information retrieval and reading comprehension," Proc. of CIKM 2018, pp.647-656, Torino, Italy, Oct. 2018.
- (4) Z. Yang, P. Qi, S. Zhang, Y. Bengio, W. W. Cohen, R. Salakhutdinov, and C. D. Manning : "HotpotQA: A dataset for diverse, explainable multi-hop question answering," Proc. of EMNLP 2018, pp.2369-2380, Brusseles, Belgium, Oct. 2018.
- (5) 西田・西田・永田・大塚・齊藤・浅野・富田：“抽出型要約との同時学習による回答根拠を提示可能な機械読解,” 言語処理学会第25回年次大会, pp.25-28, 2019.
- (6) K. Nishida, K. Nishida, M. Nagata, I. Saito, A. Otsuka, H. Asano, and J. Tomita: "Answering while Summarizing: Multi-task Learning for Multi-hop QA with Evidence Extraction," arXiv, 1905.08511, 2019.
- (7) A. Otsuka, K. Nishida, I. Saito, H. Asano, and J. Tomita : "Specific question generation for reading comprehension," Proc. of AAAI 2019 Reasoning for Complex QA Workshop, Honolulu, U.S.A., Jan. 2019.
- (8) 大塚・西田・齊藤・西田・浅野・富田：“問い合わせ可能な質問応答：読解と質問生成の同時学習モデル,” 第11回データ工学と情報マネジメントに関するフォーラム, A3-3, 2019.
- (9) K. Nishida, I. Saito, K. Nishida, K. Shinoda, A. Otsuka, H. Asano, and J. Tomita : "Multi-style generative reading comprehension," arXiv, 1901.02262, 2019.
- (10) 西田・齊藤・西田・篠田・大塚・浅野・富田：“回答スタイルを制御可能な生成型機械読解,” 言語処理学会第25回年次大会, pp.17-20, 2019.
- (11) P. Bajaj, D. Campos, N. Craswell, L. Deng, J. Gao, X. Liu, R. Majumder, A. McNamara, B. Mitra, T. Nguyen, M. Rosenberg, X. Song, A. Stoica, S. Tiwary, and T. Wang : "MS MARCO: A human generated machine reading comprehension dataset," arXiv, 1611.09268v3, 2018.
- (12) 齊藤・西田・大塚・西田・浅野・富田：“クエリ・出力長を考慮可能な文書要約モデル,” 言語処理学会第25回年次大会, pp.497-500, 2019.



(後列左から) 大塚 淳史/ 野本 済央/
西田 京介

(前列左から) 齊藤 いつみ/ 浅野 久子/
西田 光甫

機械読解の研究を通じて学術的に世界で一番の技術を創るとともに、実社会の中でたくさんのお客さまの役に立つAIサービスの実現をめざして研究開発に取り組んでいます。

◆問い合わせ先

NTTメディアインテリジェンス研究所
知識メディアプロジェクト
TEL 046-859-2674
FAX 046-859-2116
E-mail kyosuke.nishida.rx@hco.ntt.co.jp

オペレータの応対を支援する自動知識支援システム

近年、コンタクトセンタでは取り扱う商品やサービスの種類や複雑さが増大し、それに伴うオペレータに求められる知識量の増加はオペレータにとって大きな負担となり、オペレータの定着率も低下しています。本稿では、特に業務経験の浅いオペレータを支援するために、応対中のオペレータに適切な情報（知識）を自動的に提示する自動知識支援システムを紹介します。

コンタクトセンタの課題

コンタクトセンタは企業の顧客接点として重要な役割を担っています。短時間で適切に応対することによって顧客満足度を向上させることが大きな目標の1つとなっています。しかしながら、取り扱う商品やサービスの種類や複雑さは年々増大し、それに伴うオペレータに求められる知識量の増加はオペレータにとって大きな負担となり、オペレータの定着率も低下しています。このような状況の中で、特に業務経験の浅いオペレータを支援するため、応対中のオペレータに適切な情報（知識）を自動的に提示する自動知識支援システムを開発しています。

一方、オペレータに提示する情報（知識）の作成・維持には大きなコストがかかっています。これらのコストを下げるために、情報（知識）の1つであるFAQ（Frequent Asked Question）の整備を支援する技術にも取り組んでいます。本稿では自動知識支援システムおよびFAQ整備支援技術について紹介します。

自動知識支援システムの概要

自動知識支援システムは、コンタク

トセンタ（コールセンタ）に電話をかけてきたお客様の用件に基づいて適切な情報を応対中のオペレータに提示することで、業務経験の浅いオペレータを支援するシステムです。オペレータが閲覧・操作するシステムの画面を図1に示します。画面の左側にはオペレータとお客様の発話がテキストとして表示され、画面の右側にはお客様の用件を伝える発話あるいはオペレータの用件を確認する発話から自動で検索されたFAQからスコアの高い

はせがわ たかあき せきぐち ゆういちろう

長谷川 隆明 / 関口 裕一郎

やまだ せつお たもと まさふみ

山田 節夫 / 田本 真詞

NTTメディアインテリジェンス研究所

類似質問とその回答が表示されます。

本システムは以下のステップからなります。

- ① 音声対話のテキスト化：「話し終わり判定」により、オペレータとお客様の音声対話を認識してオペレータが見やすいかたちでテキストにします。
- ② 音声対話の構造化：「応対シーン推定」により、コンタクトセンタの音声対話にみられる特徴から音声対話を構造化します。



図1 自動知識支援画面

③ FAQ検索の自動化：「FAQ検索発話判定」により、お客様の用件を含む発話あるいはオペレータが用件を確認する発話を抽出し、それをFAQ検索のクエリとして利用することであらかじめ作成されたFAQを自動検索します。音声認識から発話判定までの具体的な様子を図2に示します。

■音声対話のテキスト化

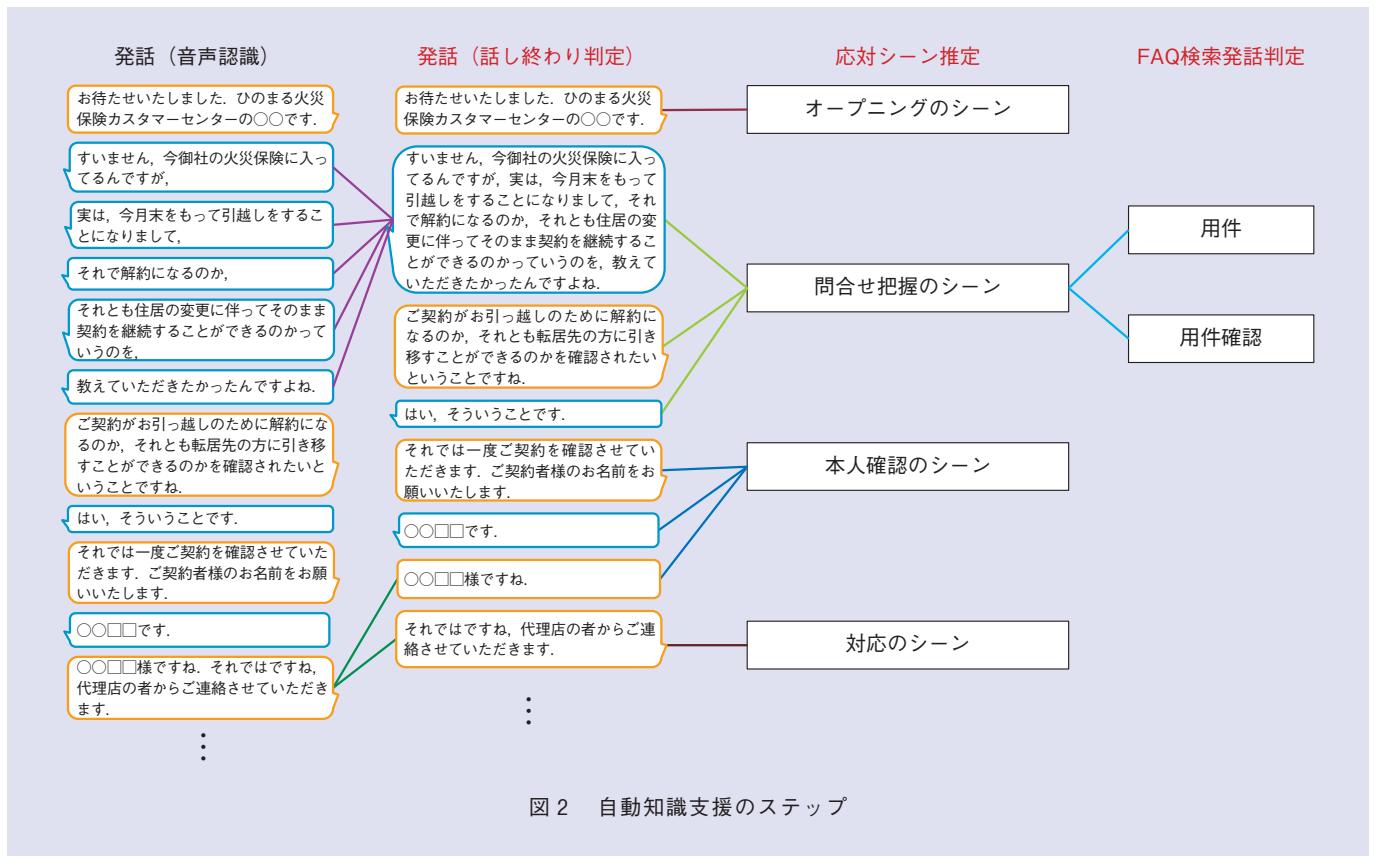
音声をテキスト化するには、最初に音声認識技術を用いますが、音声認識の結果をそのまま表示するだけではオペレータとお客様の2人の対話を見やすくなることはできません。お客様

まは電話を掛ける際に考えながら用件を伝えるため、発話がゆっくりになったり、間が空いたり、つっかえたりします。そのため、お客様の発話の音声認識結果をそのまま表示すると、ポーズが空いたところで区切れてしまい、本来は意味的にはひとまとめにして表示するべき発話が分割され、分かりにくくなってしまいます。自動知識支援システムでは「話し終わり判定」という機能を実現し、比較的長い単位で発話を表示しています。本機能は、音声認識の結果を受けて、オペレーターやお客様が話し終わったかどうか、言いたいことを言い切ったかどうかを

判定する機能です。1000通話程度の教師データからDNN (Deep Neural Network) によって学習されたモデルによる判定に加え、相手のあいづちを無視しながら、話者が変わる話者交代の時点で直前の話者は話し終わったとみなすヒューリスティックスによる補完を行うことで、高い判定精度を確保しています。

■音声対話の構造化

コンタクトセンタの対話は、特定の商品に関する問合せや特定のサービスに対する手続きの依頼など、業務で扱うタスクに偏った対話であるため、対話の流れには典型的なパターンがみら



れます。例えば、ある保険商品に関する問合せを受け付けるコールセンタでは、オペレータによる自身の名前を名乗るあいさつから始まり、お客さまが電話をかけてきた用件を確認し、契約者や契約内容を確認したうえで、用件への対応を行い、最後にお礼を述べて対話が終了します。オペレータとお客さまの対話を理解するために、このような対話の流れを大局的につかむことは非常に重要な手掛けとなりますが。私たちはこの対話の流れをつかむ機能を「応対シーン推定」と名付け、インバウンド型のコールセンタの応対シーンにふさわしいラベルを設計し、1000通話程度のコールセンタの対話ログから教師データを作成し、DNNによりモデルを学習することで、高い精度で応対シーンを推定する技術を確立しました。

■FAQ検索の自動化

FAQを自動で検索するためには、検索のためのクエリを適切に選ぶことと、検索のタイミングを適切に計ることの2点を満たすことが重要になります。検索のクエリについては、音声認識の結果をそのまま用いるのではなく、前述の「話し終わり判定」の結果を利用することで発話の分断を防ぐことができ、発話のまとまった単位を検索のクエリとすることができます。例えば、図2の例にある「実は、今月末をもって引越しをすることになります、」という発話で分断されると、クエリとして「引越し」というキーワードだけで検索することになり、後に続

く発話に含まれる「解約」や「住所変更」といったキーワードがなければオペレータにとって有効なFAQを検索することができません。このように「話し終わり判定」の結果を利用することで、発話から適切なクエリを選ぶことが可能になります。

また、お客さまやオペレータの発話があるたびに検索するのでは、検索の頻度が多くなり過ぎて、オペレータが検索結果を確認することが難しくなります。そこで、どの発話が用件であるのかを判定する必要があります。お客さまの発話が用件を述べているかどうかを判定する「用件発話判定」と、オペレータの発話が用件を確認している発話かどうかを判定する「用件確認発話判定」という局所的な分類機能を機械学習により実現しました。このときに、前述の大局的な「応対シーン推定」と組み合わせることによって、「問い合わせ把握のシーン」の発話だけに絞って「用件発話判定」と「用件確認発話判定」を適用しています。これによって、例えば「本人確認のシーン」の発話から誤って用件を抽出することがなくなるため、用件の抽出精度を向上させることができになりました。このようなプロセスで、用件として判定された発話から抽出したキーワードをクエリとすることで、FAQに対する適切な検索タイミングを実現しました。

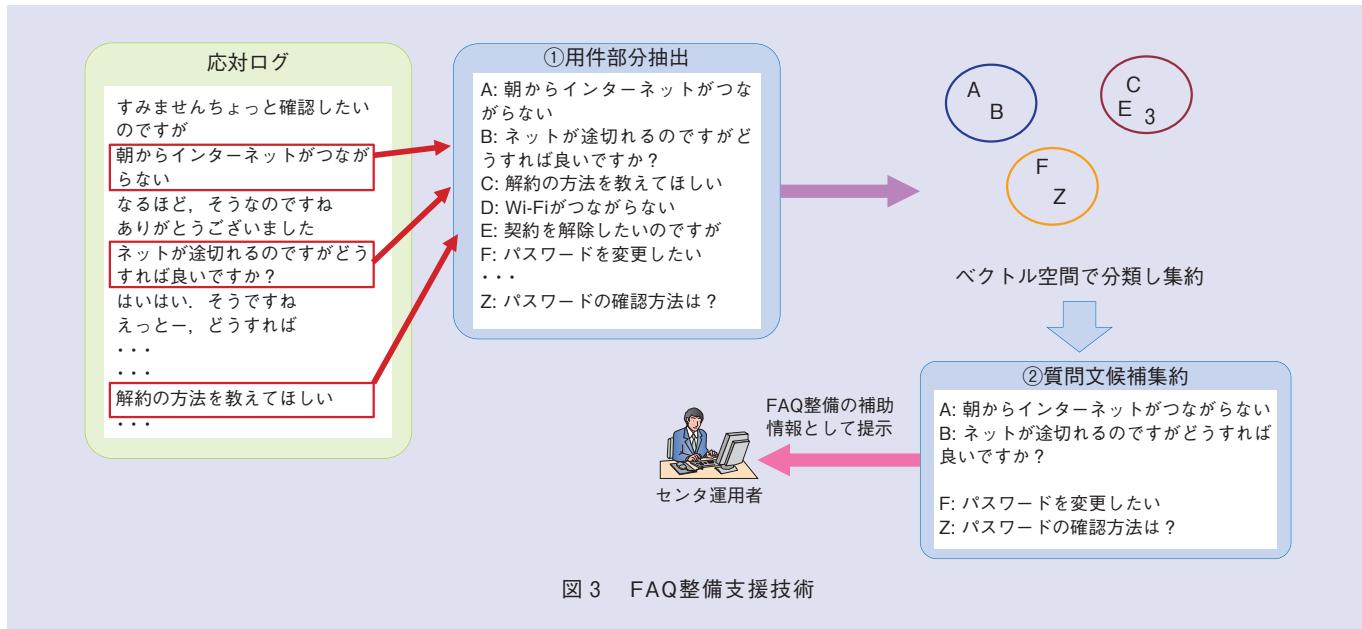
FAQの作成支援

用件発話もしくは用件確認発話から

FAQ検索を行っても、それに対応するFAQが検索対象として整備されていなければ、適切な検索結果を返すことはできません。FAQは質問文と回答文から構成されますが、回答文についてはオペレータが不正確な応対をした場合のリスクが高いため、適用先コンタクトセンタの業務に精通したセンタ運用者が作成・整備する必要があります。

一方で質問文については実際にお客さまから聞かれる内容に基づいて作成する必要があるため、実際の応対においてよく聞かれる事柄である必要があります。コンタクトセンタに蓄積された過去の通話内容（応対ログ）を解析することにより、FAQとすべき高い頻度で聞かれている事柄を質問文の候補としてセンタ運用者に提示するFAQ整備支援技術を実現しました（図3）。

まず、応対ログの中から用件となる部分を上述の「用件発話判定」を用いて抽出します。お客さまから聞かれる用件は多種多様ですが、FAQに追加すべき用件は過去に複数回にわたって質問されているものであるため、得られた用件を集約し頻度の高いものに絞り込みます。その際には、含まれている語句を手掛けりに各用件を多次元ベクトル空間に落とし込み、似た用件どうしをクラスタリングする手法を用いています。その後、集約された用件集合のうち、一定以上の規模を持つものを選び出すことにより、質問文とすべき候補の提示を行います。



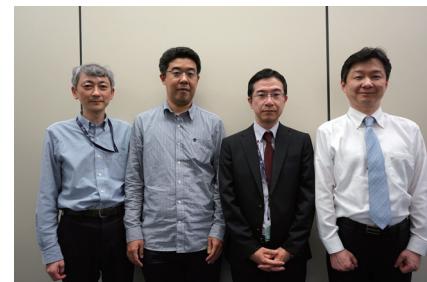
従来は自動知識支援システムの構築時に、作業者が大量の応対ログを確認して質問文の候補をセンタ運用者に伝えていましたが、この部分を自動化することによりシステム導入コストの低減に貢献します。

今後の展開

本稿では、コンタクトセンタにおけるオペレータとお客様の対話から用件を抽出しFAQを検索する技術を紹介しました。紹介した機能の多くは機械学習により実現されているため、本システムの導入にあたっては、コンタクトセンタごとに大量の教師データを低コストで作成したり、すでに作成されたモデルをさまざまなタスクに適用できるように汎用化することが重要な課題です。

今後もFAQの精度向上に加えて、

導入コストの削減の課題に対しても継続して取り組んでいきます。



(左から) 山田 節夫/ 関口 裕一郎/
長谷川 隆明/ 田本 真詞

人と人の自然な音声対話を理解したり背景知識を必要とするドキュメントを理解することは非常に難しいのですが、コンタクトセンタのようにタスクを限定すれば音声対話を理解したり文書を検索することで経験の浅いオペレータを支援できるようになってきました。今後はもっとオペレータを支援できるレベルを上げていきたいと考えています。

◆問い合わせ先

NTTメディアインテリジェンス研究所
知識メディアプロジェクト
E-mail keicho-ml@hco.ntt.co.jp



主役登場

機械読解のState of the Artをめざして

西田 京介

NTTメディアインテリジェンス研究所
特別研究員

私たちが今取り組んでいる機械読解（テキストを知識源とした自然言語理解に基づく質問応答）は、言語処理分野における流行のテーマで、過去に類をみないスピードで研究が進んでいます。最近ではコンペティション形式で取り組まれる研究課題（データセット）が多く、オンラインでリアルタイムに順位を確認可能なリーダーボード上で、世界中の企業・大学が機械読解モデルの質問応答精度を競っています。参加者の多いコンペティションでは、最高の性能（State of the Art）を達成してリーダーボードのトップに立てたとしても数日で他の研究者に抜かれてしまうことがあります。いわゆるレッドオーシャンに分類される研究領域となります。

レッドオーシャンと知りつつも機械読解に取り組むことにしたのは、機械読解の論文を初めて読んだとき「こんなことができるようになったのか！」と感動し、これを実用化できれば大きく世界が変わるだろう、自分でも取り組んでみたい、という気持ちを強く持ったためです。しかし、いざ機械読解の研究を開始してみると、私の予想を超えて競争は激化しました。リーダーボードが物凄い勢いで更新されるのを見ながら、この世界のスピードに追いつくことができるのか？と不安になりましたし、自分の考えていたアイデアを先にやられてしまうこともありました。その一方で、流行に真正面から取り組むことの楽しさと重要なことを知りました。最新の知見が日々発表されるテーマを学ぶことは刺激的ですし、社会に必要とされている技術である

からこそ機械読解が世界中で取り組まれている研究テーマになっていることを確信できました。

激しい競争の中1人で戦うのは難しいと感じ、周囲のメンバーに協力してもらいチームとしての研究スタイルを大きく変えました。個々人が関連の少ないテーマに取り組むのを止め、「全員が機械読解に近いテーマに取り組んで情報を共有しつつ、それぞれが個別にインパクトのある研究を行う」という方針に変更しました。毎週読んだ論文の内容をお互いに共有したり、プログラミング言語や深層学習のフレームワークもできる限り統一して知見やテクニックを共有しています。また、社外からも情報が入るように、勉強会への参加やTwitterでの情報収集にも力を入れています。こうした「チーム戦」「情報戦」は優秀なチームメンバーに恵まれたことで期待以上の成果を上げ、この半年間で、まだ公開になっていない物も含めて、私たちの技術が5つのデータセット・7つのタスクで世界一の精度を達成しました。私たちもやればできる、世界と闘える技術力があることを十分に示せたと思います。

機械読解の研究と実用化にはまだギャップがあります。世界一の技術でもすべてのテキストを理解して質問応答することはできないのが現状で、たくさんの課題が残っています。今後もチームとしてコンペティションなどで技術を磨きつつ、着実に実用化を果たしてNTTのAI（人工知能）事業へ貢献できるよう取り組んでいきます。