

ドコモ AI エージェント・オープンパートナーイニシアティブ：多目的対話プラットフォームの開発

大庭 隆伸^{1,a)} 秋永 和計¹ 小田 哲¹ 吉川 貴¹ 藤本 拓¹ 角野 公亮¹ 山崎 光司¹ 鎌土 記良¹
中島 悠輔¹

概要：株式会社 NTT ドコモは中期戦略 2020 「beyond 宣言」で掲げている、お客様のライフスタイルを革新する新 AI エージェントの実現に向け、基盤の中核となる AI エージェントプラットフォーム (PF) を開発・公開した。AI エージェント PF は 3 つの主要なサブ PF から構成される。本稿では AI エージェントの取り組みとこれら PF を紹介し、特に多目的対話 PF について詳しく説明する。多目的対話 PF は 3rd パーティにオープンな PF であり、対話を行うエージェントを登録することで、ユーザに様々なエージェントとの対話を提供する。AI エージェントの構築のため開発環境としての Web インターフェースが提供されており、スロット対話を中心とした対話を容易に構築することができるよう設計されている。

DOCOMO AI Agent Open Partner Initiative: Development of Multi-purpose Dialogue Platform

OBA TAKANOBU^{1,a)} AKINAGA YOSHIKAZU¹ ODA SATOSHI¹ YOSHIKAWA TAKASHI¹
FUJIMOTO HIROSHI¹ KADONO KOUSUKE¹ YAMAZAKI KOJI¹ KAMADO NORIYOSHI¹
NAKASHIMA YUSUKE¹

1. はじめに

スマートフォンの普及と、音声認識および自然言語処理技術の向上により、音声対話を介してユーザをアシストするサービスが急速に普及した。Apple の Siri[1], NTT ドコモのしゃべってコンシェル [2], Yahoo!Japan の音声アシスト [3] などが代表例として挙げられる。また、Microsoft は基本 OS ソフト Windows10 に cortana を導入し、一般の PC 向けにも音声アシスト機能を提供している。

更に近年、スマートスピーカーと呼ばれる新しいハードウェアの登場に注目が集まっている。その代表格は Amazon の Amazon Echo である。Amazon Echo の中の人格 Alexa[4] に向かって話しかけることで、同種の音声アシストサービスを受けることができる。また、LINE はクラウド AI プラットフォーム Clova[5] を用いたスマートスピー

カー WAVE を発売し、スマートディスプレイ FACE の発売も発表している。Google は同社の音声アシスト機能を提供する google アシスタントに接続されたスマートスピーカー Google home を販売している。

このような中、NTT ドコモは 中期戦略 2020 「beyond 宣言」の中で、お客様のライフスタイルを革新する新 AI エージェントの提供を掲げ、その取り組みの第一弾としてドコモ AI エージェント・オープンパートナーイニシアティブを発表した [6]。これは音声対話型のアシスト機能およびハードウェアをパートナー企業との協創により作り上げていく取り組みである。本取り組みの中核に位置づけられるシステムが AI エージェントプラットフォーム (PF)^{*1} である。その機能はオープンに公開しており、音声アシストを中心とした機能を利用可能となっている。本稿では、AI エージェントの取り組みについて紹介し、本 PF の構成について説明する。特に、音声対話機能を提供する多目的音

¹ NTT DOCOMO, INC.,
Yokosuka, Kanagawa 239-8536, Japan
^{a)} takanobu.ooba.pr@nttdocomo.com

^{*1} 正式名称は AI エージェント API であるが本稿では正しくその役割を記述することを優先しプラットフォームと称する。

声対話 PF について詳しく取り上げる。

2. AI エージェント

2.1 AI エージェントの世界観

従来の音声アシストサービスでは、1つの人格を持ったキャラクター(以降、エージェント)がユーザのあらゆる要求に応えることを指向したデザインとなっている。全知全能のエージェントの構築を目指しているとも解釈できるが、現実には、1社で音声アシストを通して提供できる機能やサービスには限界がある。

この問題に対し、オープンな PF を提供することで解決を目指す取り組みが進められている。Alexa では、skill と呼ばれる個別の機能やサービスを Alexa に登録するフレームワークが提供されている [7]。これにより任意の開発者が Alexa の機能を拡張することができる。

これに対し、AI エージェントの構想では、多くのエージェントが力を合わせてユーザをサポートする世界観を形成している。これは、分からないことは知っている人に聞く、各企業やお店のサービスはその店員から受ける、といった人間社会のアナロジーを取り込んだものである。

AI エージェントでは、エージェントをメインエージェントとエキスパートエージェントの2つに大別している。メインエージェントはユーザに寄り添うパートナーとしての位置づけであり、スマートスピーカー、スマート家電等のデバイスにユーザが話しかけた際に最初に登場するエージェントである。

エキスパートエージェントはメインエージェントから呼び出される形で登場する。エキスパートエージェントは特定の分野の専門家として作成される。直観的には各企業やお店の店員エージェントである。各社が各々のサービスを提供するエージェントを作成すれば、ユーザはそれら呼び出すことで、各社のサービスを受けられるようになる。

エキスパートエージェントは Alexa の skill に類似した概念であるが、エージェントが切り替わるという点が大きく異なる。エージェントが切り替わるという機構には、パートナーとの協創を目指す上で、幾つかの大きな利点がある。1つはパートナーがメインエージェントの人格とは独立に、パートナー自身の実現したい人格を持ったエージェントを作成することができる。企業やサービスのブランドイメージに沿ったエージェントを作成可能であり、特に日本ではマスコットを持った企業が多いが、そうしたマスコットを登場させることもできる。

2つ目は、メインエージェントをドコモ以外の企業のサービスとして提供できる点にある。メインエージェントとエキスパートエージェントは独立な存在なので、多数のメインエージェントを作成可能である。デバイスメーカーが独自のメインエージェントを作成することができ、ドコモのメインエージェントの仕様に制限されることなく、デ

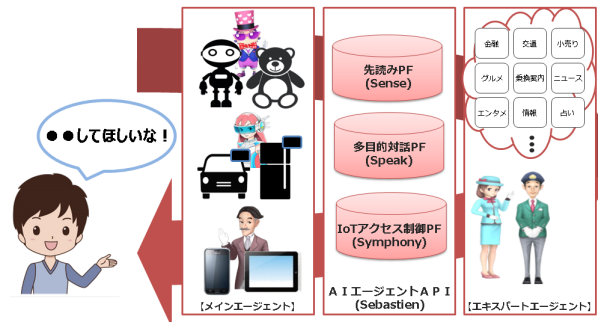


図 1 AI Agent service and its platforms.

バイスに機能を搭載することができる。

3つ目はエージェントおよび付帯サービスの提供元の切り替わりが明確になる点である。ユーザから見ると、エージェントの切り替わりによって自分が今どの企業のエージェントと会話しているのか、つまりどの企業のサービスを受けようとしているのかが明確である。これにより予期せぬ企業からのサービス提供を防止し、アクシデント発生時の問い合わせ先も明確になる。

2.2 AI エージェントプラットフォーム

AI エージェント PF は、多目的対話 PF、先読み PF、IoT アクセス制御 PF の3つの PF から構成される(図 1 参照)。多目的対話 PF は、音声およびテキストでの対話機能を提供する PF であり、前節に記載の世界観を実現するための中心的な役割を果たす。多目的対話 PF の詳細は次章に記載する。

先読み PF は、多種多様な情報を集約・解析し、その結果を各ユーザに配信するための PF である。収集する情報には、ユーザの行動に紐づく情報や、天気、公共交通機関の運行情報、災害情報などといった公共性の高い情報も含まれる。こうした情報を分析して、タイムリーに必要な情報を必要な人に届けることが目的である。

適用例の1つは宅配便の配送時間変更である。スケジュール等から事前に配送予定時間が得られているとすれば、ユーザの現在地点、現在時刻、配送予定時間の関係から、配送の受け取り可否を予測できる。もし受け取りがかなわないと推定されれば、配送時間変更の依頼をユーザに発信する。

IoT アクセス制御 PF は、宅内の IoT デバイスをインターネット網を通して屋外からでも利用可能にする PF である。宅内の IoT デバイスの管理には DeviceConnect を用いている。DeviceConnect は標準化団体 Open Mobile Alliance にて標準化された仕様 GotAPI に基づき開発されたソフトウェアである。種々のデバイスを REST ベースの API を通してコントロールすることができる。

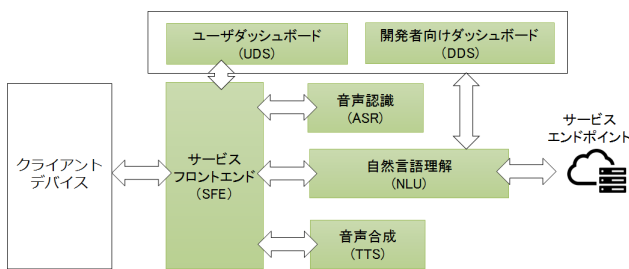


図 2 Multi-purpose dialogue platform.

DeviceConnect では、プラグインと呼ばれるデバイスの仕様を上述の REST 型の API に翻訳するためのソフトウェアを構築することで、各種デバイスを取り込むことができる。IoT 機器の標準化仕様は他にも幾つか存在しているが、それらの仕様を DeviceConnect の仕様に翻訳するプラグインを作成することで、これらの規格を取り込むこともできる。そのため対応デバイスが多い点が特徴である。IoT アクセス制御 PF は、宅内の DeviceConnect マネージャーと通信することで、インターネット網を介した屋外からの IoT 機器制御を可能にしたものである。

多目的対話 PF、先読み PF、IoT アクセス制御 PF は独立な PF であるため個別に利用することもできるが、互いに連携させることも可能である。多目的対話 PF が先読み PF にアクセスすることで、会話にユーザのタイムリーな情報を反映させることができる。また、多目的対話 PF から IoT アクセス制御 PF にアクセスすることで音声コマンドで IoT 機器を操作することができる。

3. 多目的対話プラットフォーム

3.1 全体構成

多目的対話 PF の構成を図 2 に示す。多目的対話 PF は音声区間検出 (VAD)、音声認識 (ASR)、自然言語理解 (NLU)、音声合成 (TTS) を一括で提供する。音声 IN-音声 OUT の対話だけでなく、テキスト IN-テキスト OUT も可能であり、片方をテキスト、他方を音声にすることも可能である。どのような入出力が可能かはクライアント側のデバイス (以降、デバイス) に依存するので、PF としては柔軟に切り替えられる設計になっている。

ここでは、音声 IN-音声 OUT の対話を例に PF の動作を順に説明する。まず対話を開始する前に、デバイスとユーザアカウントの紐づけが必要である。紐づけを事前に行うことで、AI エージェントはどのユーザと対話しているかを把握できる。これにはユーザダッシュボード (UDS)[8] を利用する。UDS にログインし、デバイスの固有の ID を登録する。UDS は Web 上のホームページであるが、UDS 上の各機能は API 化されており、デバイス設定アプリを作成

して所定の API を順に叩くことで紐づけ作業を行うこともできる。

対話開始時、デバイスはサービスフロントエンド (SFE) に接続する。この際、デバイスに対する認可処理が行われ、UDS と連携してデバイスとユーザの紐づけ確認を行う。デバイスからマイク信号が送られると、SFE は VAD を実施する。発話を検知すると ASR に音声データを転送し、結果が返ると、NLU にそれを通知し応答文を得る。そして TTS に合成音声の生成を依頼し、最終的にデバイスに音を届ける。VAD, ASR, NLU の結果はデバイスにも都度返却する仕様となっており、処理の経過を取得できる。

NLU は、TTS サーバーとそのモデルを指定することができる。つまり声を変えることができる。これはエージェントごとに異なる声を生成するのに利用される。特にスマートスピーカーのようなディスプレイを持たないデバイス上では、声が変わることでエージェントの切り替わりが明確になる。

一般的な音声対話の実装方法では、デバイスは ASR, NLU, TTS の各サーバーに個別に接続する。多目的対話 PF では SFE がデバイスとの通信の窓口をひとえに担い、原則 1 本のコネクションで音声 IN-音声 OUT の一連の機能を提供する。この方式には幾つかの利点がある。1つは音声対話のレスポンス性の向上である。原則上は音声を送る 1 回だけであり、Internet 網の通信の影響を最低限に抑えることができる。2つ目はデバイス側のプログラミング実装の簡潔化である。音声特有の難しい処理がサーバー側に集約されているため、プログラマーの力量がレスポンス性や安定性に及ぼす影響を抑えることができる。3つ目に将来的な機能拡張性がある。音声対話に新しい機能を搭載する場合にも、サーバー側の変更だけで済む。

メインエージェントの作成にあたっては NLU 内の対話ロジックの書き換えが許容されている。メインエージェントはデバイス側の実装との密な連携が必要であるから、ユーザ ID の取り扱い方など一部を除き、原則自由に NLU 上にロジックを組むことができる。

一方、エキスパートエージェントは様々なメインエージェントから呼び出される必要があり、呼び出しや対話の終了通知の方法など所定のルールを守る必要がある。この所定のルールを充足する必要性もあり、エキスパートエージェントを開発者が容易に実装・登録ができるように開発者向けダッシュボード (DDS) が用意されている [9]。DDS については次節で詳しく説明する。

さて、登録されたエキスパートエージェントを実際にユーザが利用する前には許諾を得る必要がある。特にエキスパートエージェントは各社が登録するエージェントであるため、対話データがその企業に開示されることについての事前承諾や、その企業のサービスを利用するための規約への同意が必要である。逆に言えば、ユーザにはエキス



図 3 Developer dashboard.

パートエージェントの選択権があり、選択はユーザの判断である。

エキスパートエージェントの選択は UDS 上で行うことができる。^{*2}自分が利用したいエキスパートエージェントの機能を、規約に承諾の上で有効にしたり、必要なユーザアカウントの認証などを行うことができる。

3.2 開発者向けダッシュボード

図 3 は DDS の画面である。エキスパートエージェントの作成に際し、まず名称、呼び出し名、呼び出し時に最初に発話されるメッセージなどの基本情報を登録する。前述の通り、音声合成の話者を選択することができ、執筆時段階で 10 種類以上の話者の中から自由に選択できる。今後更に数を増やす予定である。

対話を行う上で重要な 1 つの機能はユーザ発話の意図を分類することである。DDS では複数の意図 (intent) を登録することができる。Intent の分類は基本的にはキーワードマッチングである。ユーザの想定発話を入力しておくことで、それをもとに分類を行う。

DDS はスロットフィル型の対話を想定した設計になっており、intent 毎に複数のスロットを定義できる。スロットの設定項目に必須フラグがあり、必須スロットはその値が対話を通して取得されるまで、ユーザに聞き返しを行う。必須スロットが複数ある場合は、各スロットについて順番に聞いていく。

スロットに入る値は網羅的に登録しておく必要がある。これは単語セットを作る作業に等しいが、これは entity の登録にて行う。Entity の登録の際は、同義語・表記揺れを定義しておくことで正規化することができる。

意図およびスロットが抽出できるようになれば、それに応じた応答文を生成すればよいが、実利用上は各企業 (開発者) が保持するデータベースに問い合わせるなど、外部システムとの連携が必要になる。これは DDS 上に外部サーバーの endpoint を指定することで実現される。利用者のユーザアカウント管理が必要なサーバー向けに、認証用の URL を登録することもできる。

^{*2}執筆時段階では開発者向けにのみ公開していないため、エキスパートエージェントの公開を停止している。

```
Method: POST
Headers: "Content-type: application/json"
Request Body:
{
  "bot_id": "{ 内部的に払い出された ID }",
  "user_id": "{ ユーザ識別 ID }",
  "lang": "jaJP",
  "args": {
    "utterance": "ユーザ発話",
    "intent": "Intent 名",
    "slot 名": "slot の entity",
    ...
  }
}
```

図 4 A json example from NLU.

Endpoint に登録されるサーバーを、Context Generation Server (CGS) と呼んでいる。CGS は端的には、NLU から送られてくる json を受けて、応答を所定の json 形式で返すサーバーである。図 4 に NLU が出力する json を例示する。Intent 名や slot 名とその値が送られるので、これをもとに必要な処理を実施し、応答を返すサーバーを構築すればよい。

4. まとめ

NTT ドコモはドコモ AI エージェント・オープンパートナーイニシアティブを発表し、その中核となる AI エージェント PF を公開した。本稿では、AI エージェントの世界観を紹介し、AI エージェント PF を構成する 3 つの PF について概説した。特に対話機能を提供する多目的対話 PF について、その構成や特徴について詳しく述べた。執筆時点では、開発者向けのトライアルを開始したばかりである。今後更なる拡張や機能追加が行われていく予定である。機能の説明や使い方はホームページ上 [10] でも紹介されているので参照されたい。

参考文献

- [1] <https://www.apple.com/jp/ios/siri/>
- [2] <https://www.nttdocomo.co.jp/service/shabette-concier/>
- [3] <https://v-assist.yahoo.co.jp/>
- [4] <https://developer.amazon.com/ja/alexa-voice-service>
- [5] <https://linecorp.com/ja/pr/news/ja/2017/1667>
- [6] https://www.nttdocomo.co.jp/info/news_release/2017/06/23-00.html
- [7] <https://developer.amazon.com/ja/alexa-skills-kit>
- [8] <https://users.sebastien.ai>
- [9] <https://developers.sebastien.ai>
- [10] <http://docs.sebastien.ai>