

しゃべってコンシェルと言語処理

吉村 健[†]

しゃべってコンシェルでは自然な発話により、携帯端末の基本機能（電話、メール、スケジューラ、等）を利用したり、さまざまな情報（地域情報、リアルタイム情報、デジタルコンテンツ、等）を検索したりすることが簡単にできる。そのようなサービスを実現するにあたり、ユーザの発話の内容を解析し、ユーザの求める最適なアプリケーションや各種専門検索エンジンに処理を橋渡しする意図解釈技術と、発話内容が質問である場合には、質問に対する回答そのものを提示する知識検索技術が適用されている。本稿では、しゃべってコンシェルのサービス概要と、そこで活用されている言語処理技術について述べる。

Shabette-Concier Service realized by Natural Language Processing

TAKESHI YOSHIMURA[†]

Shabette-Concier is a voice agent application, where users can easily utilize basic functions on a mobile phone (telephone, mail, scheduler, etc.) and can access variety of information (local area information, realtime information, digital contents, etc.) by simply and naturally speaking to their mobile phones. To realize this service, technologies based on natural language processing (NLP) are applied. The technologies include “intention understanding technology”, which understands a user’s intention and bridges the intention to an appropriate function on a mobile phone or a search engine, and “question answering technology”, which estimates and presents direct answers to a user’s question. This paper describes Shabette-Concier service and NLP technologies that realize the service.

1. はじめに

フィーチャーフォンからスマートフォンへの移行が進む中で、携帯端末上で利用するアプリケーション数も増加し、各ユーザの利用するネットワークサービスも多様化している。端末アプリケーションやネットワークサービスが乱立する状況において、ユーザが利用したい機能やサービスを簡単に呼び出せること、すなわち誰にでも簡単に使えるサービス導線の提供がますます重要となってきた。

簡単に使えるサービス導線を実現する一手段として、NTT ドコモでは、2012年3月に“しゃべってコンシェル”[1]という音声エージェントサービスを開始した。ユーザが携帯端末に話しかけることで、携帯端末の各種機能（電話、メール、スケジューラ等）を活用したり、様々な専門情報（地域情報、リアルタイム情報、デジタルコンテンツ等）を検索したりすることができるアプリケーションである。単語を入力するだけの従来の検索行為と異なり、ユーザの自然な発話により、ユーザが所望する機能や情報を自由自在に提供することを実現している。しゃべってコンシェル以外にも、AppleのSiri[2]やYahoo!の音声アシスト[3]など、音声エージェントアプリケーションの実用化が各方面から始まっている。

しゃべってコンシェルのサービス設計にあたって、1)自然言語ユーザインタフェース、及び2)豊富なデータベースの2点を重視し、実用化に向けた開発を進めた。1点目の自然言語ユーザインタフェースは、ユーザの自然な発話か

らユーザの意図を汲み取り、適切なアプリケーションの起動や、情報検索を実行する部分である。ユーザは日常の会話と同様に携帯端末に話しかけることで、所望の機能を実行することが可能である。そして、2点目の豊富なデータベースにより、ユーザの求めている情報を適切に返すことができるようになる。データベースとしては、飲食店・病院・観光情報といった地域情報や、ニュースやツイート等のリアルタイムな時事情報、さらには携帯端末向けの音楽や映像、電子書籍といったデジタルコンテンツを揃えており、それらの情報・コンテンツへのアクセスを容易にしている。これらのデータベースに加え、特徴的な機能として、ユーザの質問に対し直接回答候補を提示する“知識Q&A”機能を2012年6月より開始した。本機能ではユーザの質問内容を解析し、データベース内もしくはインターネット上の情報から、もっとも適切と思われる回答を推定し、提示している。

上記の機能を実現するにあたり、しゃべってコンシェルではクラウド上で様々な言語解析技術を活用している。本稿では、しゃべってコンシェルのサービス概要及び仕組みについて紹介するとともに、しゃべってコンシェルの根幹を支える言語処理技術を解説する。具体的には、ユーザの発話内容を解析し、最適な端末機能や各種専門検索エンジンに処理を橋渡しする意図解釈技術（意図解釈エンジン）と、発話内容が質問である場合には、質問に対する回答そのものを推定し提示する知識検索技術（知識検索エンジン）について説明する。

[†] NTT ドコモ サービス&ソリューション開発部
Service & Solution Development Department, NTT DOCOMO, Inc.

2. しゃべってコンシエルのサービスと仕組み

2.1 サービス概要

誰にでも簡単に使えるサービス導線の提供に向けて、しゃべってコンシエルでは、携帯端末上の主要な8つのアプリケーションと、NTTドコモのポータルサービス“dメニュー”4上で提供されている11の検索サービス、インターネット上の4つのネットワークサービス、そして知識Q&A機能との連携を実現している(表1, 2012年9月時点)。

表1 しゃべってコンシエルで連携する端末機能、ネットワークサービス(2012年9月時点)

Table 1 The terminal applications and network services provided through Shabette-Concier (as of Sept. 2012).

	提供機能・コンテンツ	提供者
端末機能	電話 メール スケジューラ アラーム カメラ タイマ メモ 音楽プレーヤー	
ネットワークサービス	天気 乗換案内 エリアガイド(地域情報) ニュース リアルタイム(ツイート) 画像コンテンツ 動画コンテンツ 音楽コンテンツ アプリ ゲーム 本/コミック	NTTドコモ (dメニュー)
	レシピ情報	クックパッド
	地図	ゼンリンデータコム
	教えてgoo	NTTレゾナント
	Wikipedia	Wikipedia
	知識Q&A	NTTドコモ

上記以外に、何か機能や情報を求めているわけではない発話に対しては、他愛もない雑談コメントを返すようになっている。これらの機能・サービスと連携することにより、下記のような発話例に対し、ユーザの求める機能・情報の提供が可能である。

- 「明日は晴れるかな？」
→ 現在地の天気情報
- 「佐藤さんにメールする」
→ 佐藤さん宛てにメール作成機能
- 「渋谷まで」
→ 最寄り駅から渋谷駅までの乗換情報
- 「トマトとキャベツのレシピは？」
→ トマトとキャベツを使うレシピ情報
- 「渋谷区の面積は？」
→ 知識Q&Aにより回答

- 「幸せだな～」
→ 雑談コメントを回答

2.2 アーキテクチャ

しゃべってコンシエルでは多くの機能をクラウド上のサーバで実行している。主要なサーバ機能は以下のとおりである。

- 音声認識エンジン
携帯端末から音声情報を伝送し、クラウド上の音声認識エンジンにより、ユーザの発話をテキストに変換する。モバイル環境での典型的な雑音環境を考慮した音響モデルと、しゃべってコンシエルでの典型的な発話内容を広くカバーした言語モデルを構築している。
- 意図解釈エンジン
しゃべってコンシエルの心臓部分に当たる最重要機能である。テキスト化されたユーザ発話から、ユーザの意図を解析する。解析結果として、処理を引き継ぐべきタスク(端末機能または専門検索エンジン)の選択結果と、選択されたタスクに入力すべきキーワードが出力される。詳細は3章にて説明する。
- 各種専門検索エンジン
天気やニュース、画像など、専門のデータベースを持った検索エンジンである。通常の実用検索エンジンと同様に単語に基づく検索機能を提供する。
- 知識検索エンジン
入力された質問文に対し、推定した回答を返す検索エンジンである。専門検索エンジンの1つではあるが、自然文による入力を前提としているところが、他の専門検索エンジンと異なる。詳細は4章にて説明する。

ユーザの発話から機能・情報提供までの流れ(2012年9月時点)を図1に示す。ユーザがしゃべってコンシエルアプリ上で発話すると、端末から音声情報が音声認識エンジンに伝送される。音声認識エンジンでは、音声情報から尤もらしい認識結果を返す(1)。続いて、認識結果のテキスト文章を意図解釈エンジンに伝送し、意図解釈エンジンにて最適なタスク(端末機能または専門検索エンジン)を判定する(2)。端末機能が判定された場合には、しゃべってコンシエルアプリからAndroid OSのインテント機能[5]を活用し、該当するアプリケーションにキーワードとともに処理を引き継ぐ(3a)。一方、専門検索エンジンが判定された場合には、当該専門検索エンジンに対しキーワード検索を実行し、その検索結果をしゃべってコンシエルアプリの中で表示する(3b)。

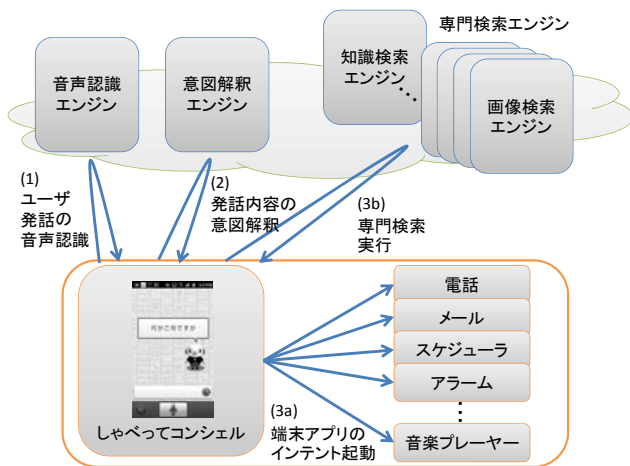


図 1 しやべってコンシェルを実現するアーキテクチャ
 Figure 1 The architecture of Shabette-Concier

3. 意図解釈エンジン

本章ではしゃべってコンシエルのコア機能である意図解釈エンジンについて述べる。意図解釈エンジンは自然言語で発話された内容を解析し、

- 1) どの端末機能・専門検索に処理を割り振るか (タスク判定)
- 2) 選択した端末機能・専門検索に対し、伝達すべき情報を抽出 (キーワード抽出)

という 2 つの処理を実行する。たとえば、「佐藤さんにメールする」という発話であれば、タスク判定処理としてメール作成機能を選択し、キーワード抽出結果は「佐藤(さん)」となり、メール作成機能に抽出したキーワードを引き渡す。「トマトとキャベツのレシピは？」という発話の場合、タスク判定結果としてはレシピ検索となり、キーワード抽出結果は「トマト」と「キャベツ」となる。

タスク判定は、ユーザの幅広い発話に対応する必要がある。たとえば「明日の天気は?」「今日は寒いかな?」「東京の降水確率は?」などの多様な言い回しに対し、天気情報を求めていることを適切に判定できなければならない。これは、いわゆる教師ありの文書分類問題[6]であり、ユーザの発話内容を入力文章に見立て、どのタスクに属する発話であるかを分類する課題である。多様な発話パターンに対応するために、相当量の発話事例を収集し、教師あり機械学習を用いてタスク判定のための学習モデルを構築している。

機械学習の特徴量には、発話内容を形態素解析した上で、各形態素の表層の unigram や bigram, 品詞, さらに各単語に付与されたカテゴリ情報等を用いる。発話内容は短いため、一般の文書分類問題と比べると情報量は非常に少ない。そのため、タスク判定において重要になるのは単語のカテゴリ情報であり、カテゴリ情報が付与された辞書を整備している。カテゴリ情報の整備にあたっては、各種専門

検索エンジンのデータベースに含まれるインデクス情報なども活用している。

キーワードの抽出は、通常は名詞を抽出することで行っている。そしてタスクごとにストップワードリストを用意し、ストップワードリストに含まれる名詞はキーワード除外する。ストップワードを除外するのは、たとえばレシピ検索に入力するキーワードとして「レシピ」は不要であるように、タスクによって不要なワードがそれぞれあるためである。また地域情報検索や乗換案内では、地名や駅名が必須であることから、現在地に基づいて地名や最寄り駅名を自動補完するなどの処理も行っている。また知識検索エンジンに対しては、キーワード抽出処理は行わず、発話文をそのまま入力文としている。

図 2 に意図解釈エンジンでの処理フローの一例をまとめる。はじめに発話内容に対し形態素解析を行い、形態素に分割する。つづいて発話に含まれる単語に対し、カテゴリ情報を付与し、これらの情報から特徴量を抽出する。そして予め学習したモデルを適用しタスク判定を実行し、どのタスク (端末機能・専門検索エンジン) に処理を引き継ぐか決定する。最後に発話内容に含まれる名詞等の情報からキーワードを抽出し、選択したタスクに対し、当該キーワードを入力情報とする。

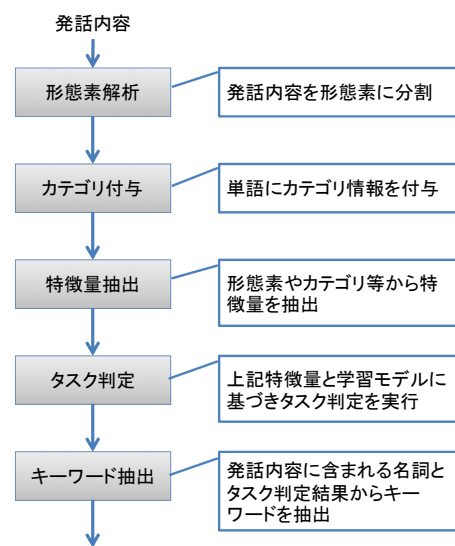


図 2 意図解釈エンジンにおける処理フロー例
 Figure 2 A process flow of intention engine.

4. 知識検索エンジン

4.1 基本構成

しゃべってコンシェルの特徴的な機能のひとつとして、知識 Q&A 機能がある。ユーザの質問に対し直接的に回答を推定し提示するため、ちょっとした調べものやエンターテイメント的な利用価値を提供している。本章では、本機能を実現する知識検索エンジンについて説明する。この知識検索エンジンは、NTT メディアインテリジェンス研究所の基盤技術に基づき [7]、NTT ドコモで開発したものである [8]。

知識検索エンジンは、2つの QA システムによって構成されている。1つは DB 型 QA システムであり、予め蓄積された知識データベースの中から回答を提示するシステムである。2つ目は検索型 QA システムであり、入力された質問文に基づき Web 検索を実行した検索結果から、回答と推定される部分を抽出し提示する。DB 型 QA システムは、質問カバー率は低いものの、精度の高い回答を返すことが可能である。一方、検索型 QA システムは、精度は DB 型 QA システムに比べて高くはないものの、質問カバー率の高い応答が可能である。2つの QA システムにより、精度とカバー率を補完しあう知識検索エンジンを構成している。

図 3 に知識検索エンジンの基本構成を示す。フロントサーバで一旦質問を受け付けると、最初に DB 型 QA システムに問い合わせる。DB 型 QA システムから回答がある場合には当該回答を提示し、回答がない場合には検索型 QA システムに問い合わせた結果を提示している。どちらの QA システムから出力された回答なのかは、表示方法の違いによってユーザにわかるようになっている。

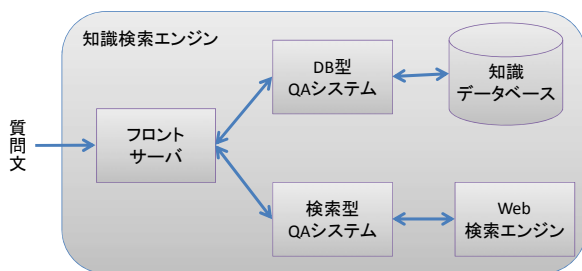


図 3 知識検索エンジンの基本構成

Figure 3 The basic system components of QA engine.

4.2 DB 型 QA システム

DB 型 QA システムでは、質問文を解釈し、解釈結果に基づき知識 DB を検索して、質問文に対する回答を提示する。質問文の解釈方法は様々ではあるが、その代表例として、Entity-Property タイプの質問解釈方法について説明する。

Entity-Property タイプの質問とは、たとえば「渋谷区の面積は?」「清水の舞台の高さは?」「ハリポットの監督は?」というように、ある対象物 (Entity) の属性情報

(Property) を聞くような質問のことを指している。質問タイプとして、もっとも典型的なものである。

Entity-Property タイプの質問の場合、次の処理により Entity と Property を抽出する。Entity の抽出には系列ラベリング手法である CRF (Conditional Random Field: 条件付き確率場) [9]を用いる。これにより上記の質問例では、「渋谷区」「清水の舞台」「ハリポッター」を抽出することができる。一方、Property の判定は、意図解釈エンジンにおけるタスク判定と同じように、機械学習による分類問題として判定する。機械学習を用いるのは幅広い言い回しにも Property を適切に判定するためである。

Entity と Property を抽出した後、Entity と Property をキーとして知識データベースに対し問合せを行う。Entity と Property の抽出結果は各 1 つとは限らず複数の場合もあるため、優先度をつけて問合せを行う。

知識データベースに回答がある場合は、当該回答を質問に対する回答として返答する。DB 型 QA システムの回答結果の一例を図 4 に示す。回答結果ページにおいて、回答だけでなく、質問をどのように解釈したのかも含めて提示している。一方、知識データベースに回答がない場合は、DB 型 QA システムでは回答が得られなかったものとして、検索型 QA システムに処理を移す。



図 4 DB 型 QA システムの回答結果例

Figure 4 Result pages from DB-based QA system.

4.3 検索型 QA システム

検索型 QA システムでは、質問文に基づき Web 検索を実行し、その検索結果ページから回答と推定される回答候補を抽出・評価し、回答をランキング表示する。図 5 に検索型 QA システムにおける処理の流れの一例を示す。

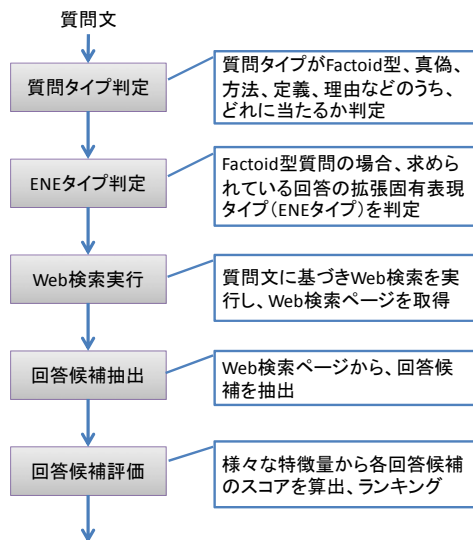


図 5 検索型 QA システムにおける処理フロー例
 Figure 5 A process flow of search-based QA system.

質問文が入力されると最初に質問タイプの判定を行う。質問タイプ例としては、単語程度の短い回答を求める質問タイプ (Factoid 型) や、Yes/No を聞く質問タイプ (真偽)、または少し長めの回答を期待する方法や定義、理由を聞く質問タイプ等に分類される。以下では、質問の中でも特に多い Factoid 型質問の処理を説明する。

Factoid 型質問と判定されると、次に人名、地名、組織名、日付、数量など、100 を超える拡張固有表現タイプ (ENE タイプ。一例として[10]など) の中から何を聞く質問なのかを判定する。例えば、「日本で一番高い山はどこですか?」という質問に対し、回答として適切な ENE タイプは山地名であると推定する。

つづいて、質問文から検索ワードを抽出し、Web 検索エンジンに検索を行い、検索結果ページを得る。検索結果ページの中には、検索ワードを含む Web サイトへのリンクと、当該 Web サイトのタイトルや概要文 (スニペット) が含まれている。この検索結果ページの中から、上記で推定した ENE タイプと一致する固有表現を回答候補として抽出する。上述の質問例の場合、山地名にあたる単語を検索結果ページの中から抽出することになる。

回答候補を抽出した後、各回答候補の評価とランキングを行う。回答候補の評価は、様々な特徴量に基づく機械学習によりスコアを計算する。特徴量の例としては、検索結果ページに含まれる回答候補の出現数 (多いほど評価が良い) や、出現するスニペットの検索ランキング順位 (高いほど良い)、検索ワードとの近さ (近いほど良い) などを用いる。こうして回答候補ごとに得られるスコアをもとに、最終的な回答ランキングを生成し出力する。

図 6 に検索型 QA システムの回答結果の一例を示す。インターネット上の情報から推定した結果であることがわか

る表現とともに、Web サイトへのリンクを提示することで、ユーザを Web サイトに誘導するページ構成にしている。



図 6 検索型 QA システムの回答結果例
 Figure 6 Result pages from search-based QA system.

5. おわりに

本稿ではしゃべってコンシェルサービスの概要と、そのサービスの根幹を支える言語処理技術として、意図解釈エンジンと知識検索エンジンについて解説した。しゃべってコンシェルはサービス提供後も随時機能更新や機能拡張を行っており、本稿の説明内容はすべて 2012 年 9 月時点のものであることを注意された。

しゃべってコンシェルは、実用的な要素だけでなく、エンターテインメント的な要素も多分にあり、多種多様な楽しみ方をさせていただきたいと思っている。今後は、より多くの端末機能やネットワークサービスとの連携を推進することで実用性を向上させるとともに、エンターテインメント性についても追求していく予定である。

またしゃべってコンシェルに活用されている要素技術・基盤エンジンを多方面に展開することも計画している。その一例として、NTT ドコモのスマートフォン向けナビゲーションサービスである「ドコモドライブネット」[11]としゃべってコンシェルの音声認識及び意図解釈エンジンを連携させ、自然な音声発話により簡単にナビゲーション機能进行操作できる音声操作機能を提供する予定である。さらには、2012 年の CEATEC JAPAN にて展示した「しゃべってロボ」[12]のように、携帯端末に限らず様々なデバイスとしゃべってコンシェルの基盤エンジンを接続し、多様なコミュニケーションサービスをマルチデバイスに展開していくことも検討中である。

参考文献

- 1) しゃべってコンシェル | サービス・機能 | NTT ドコモ
http://www.nttdocomo.co.jp/service/information/shabette_concier/
- 2) アップル - iOS 6 - あなたの声を使って Siri にもっといろいろ
たのみましょう.
<http://www.apple.com/jp/ios/siri/>
- 3) 音声アシスト for Android - Yahoo! JAPAN
<http://v-assist.yahoo-labs.jp/>
- 4) dメニュー <http://smt.docomo.ne.jp/>
- 5) Intent | Android Developers
<http://developer.android.com/intl/ja/reference/android/content/Intent.html>
- 6) 高村大也 (監修: 奥村学), 言語処理のための機械学習入門,
コロナ社, 2010.
- 7) 東中竜一郎 他, “しゃべってコンシェルにおける質問応答技
術”, NTT 技術ジャーナル 2013 年 2 月号 (掲載予定).
- 8) 内田渉 他, “自然文質問への直接回答を実現する知識 Q&A シ
ステム”, NTT ドコモテクニカルジャーナル 2013 年 1 月号 (掲載
予定).
- 9) John Lafferty, Andrew McCallum, Fernando Pereira, "Conditional
Random Fields: Probabilistic Models for Segmenting and Labeling
Sequence Data" ICML2001 282-289.
- 10) 関根の拡張固有表現階層
<https://sites.google.com/site/extendednamedentityhierarchy/>
- 11) ドコモ ドライブネット | サービス・機能 | NTT ドコモ
http://www.nttdocomo.co.jp/service/information/drive_net/
- 12) 報道発表資料: 「CEATEC JAPAN 2012」への
出展について | お知らせ | NTT ドコモ
http://www.nttdocomo.co.jp/info/news_release/2012/09/19_00.html