

# 既存のFAQ集から低コストで構築できる 質問応答対話システム

Construction of Low-Cost Question-and-Answer Dialog System Using Existing FAQs

## FAQ集から手軽に構築できる対話システムで、複雑なシナリオの作成を不要にする

ユーザーからの問い合わせを受ける窓口で、スタッフに代わってFAQ (Frequently Asked Questions) に答える対話システム(チャットボット)が使われるようになってきました。対話システムを使うことで、ユーザーはいつ問い合わせてもすぐに対応してもらえますし、窓口スタッフは負担を軽減できます。しかし、従来は、対話の進め方を表すシナリオを、FAQ集に基づいて手作業で作る必要があり、対話システムを簡単に構築することができませんでした。

そこで東芝は、FAQ集に依存しない共通シナリオと、FAQ集を共通シナリオに合わせて解析して検索用データベース(DB)を作成する手法により、FAQ集から自動で構築できる質問応答対話システムを開発しました。

### 背景

窓口スタッフに代わって問い合わせに対応する対話システムは、休業日や深夜でもすぐに対応できるので、ユーザーの利便性を向上させます。また、同じような問い合わせを頻繁に受ける場合には、FAQ集に基づく対話システムが問い合わせ対応の一部を肩代わりすることで、窓口スタッフの負担を軽減できます。

FAQ集に基づく対話システムは、対話の進め方を表すシナリオに従って応答を決めます。ユーザーとの対話を通してユーザーが知りたい内容に最も近いFAQを検索し、それに対応する回答をユーザーに示します。検索の結果、複数のFAQがヒットした場合は、絞り込むための追加情報をユーザーに求めます。これにより、FAQの件数が多くても、適切な回答を探し出せます。

### 従来のFAQ集に基づく対話システムの問題

従来は、例えば、「ログインできない」と言われたら、Webサイトからスマートフォンのアプリケーションかを聞き返す”のようなルールを複数組み合わせ、手作業でシナリオを作っていました。このため、FAQ集ごとに専用のシナリオが必要であり、また、様々な場合を想定して多くのシナリオを作るため、時間とコストが掛かっていました(図1(a))。また、

一度シナリオを作った後も、FAQ集を変更する度にシナリオも修正しなければなりませんでした。

### 質問応答対話システム

今回開発した手法は、それぞれのFAQ集に専用のシナリオを作らなくても、対話システムを構築できます(図1(b))。複数のFAQ集に共通に使えるシナリオを準備して、そのシナリオに合わせて、FAQ集の内容を解析して検索用DBを事前に作成しておきます。

まず、キーワードとそのキーワードを含むFAQの関係を調べます(図2の左側)。各FAQに対して、全角・半角の統一などをはじめとするテキストの前処理、単語分割、及び品詞推定を行い、名詞と動詞をキーワードとして抽出します。全てのFAQについてキーワードを抽出し、各キーワードが何件のFAQに含まれているかに応じたスコアを計算します。

次に、「申し込み」と「申込」などの表記揺れや、「申し込む」と「申請する」などの言い換え表現による取りこぼしを防ぐために、キーワードの類義語を抽出します。キーワードの類似性を判定するには、単語を数百次元のベクトルで表す、埋め込み表現と呼ばれる手法を用います。各単語に対応する埋め込み表現は、あらかじめ大量のテキストデータを用いて、学習しておきます。この埋め込み表現を用いて、二つの単語の類似度を対応するベクトル間の距離から計算し、キーワー



図1. 従来手法と開発した手法の比較

従来は、FAQ集ごとに専用のシナリオを手作業で作成するため、時間とコストが掛かっていました。開発した手法は、共通シナリオに合わせて、FAQ集から検索用DBを作成することで、対話システムを容易に構築できます。

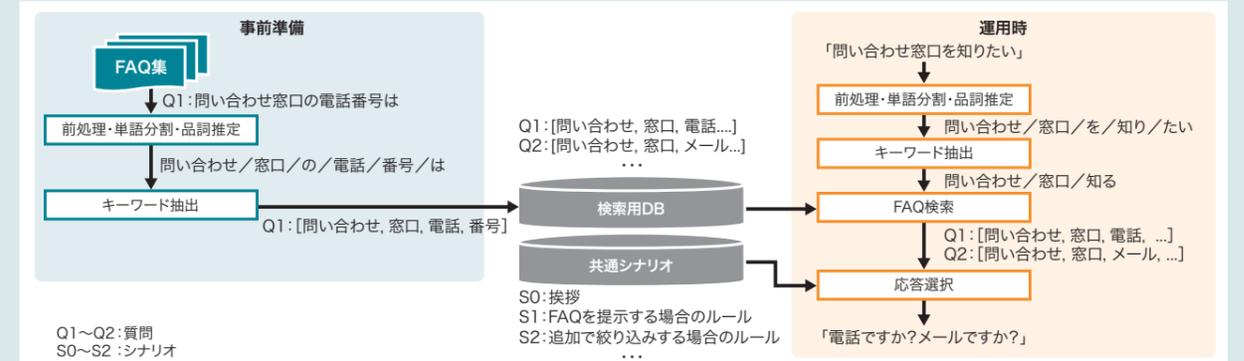


図2. 事前準備と運用時の処理

事前準備では、各FAQに含まれるキーワードを抽出しておきます。運用時は、ユーザーの質問からキーワードを抽出し、適切なFAQを検索して、共通シナリオに従って対応します。

ドとの類似度が高い単語を類義語として抽出します。

このようにして得たキーワードごとのスコアと類義語を、そのキーワードを含むFAQとひも付けて保存する検索用DBを作成するまでが、事前準備です。

運用時は、ユーザーからの問い合わせを受け取ると、事前準備と同じように、テキストの前処理、単語分割、品詞推定を行い、名詞と動詞をキーワードとして抽出します(図2の右側)。そして、抽出したキーワードを検索用DBで検索し、事前準備で計算したキーワードのスコアに基づいて、検索したFAQの順番を決めて表示します。ユーザーがキーワードの類義語を使用している場合は、その類義語と元のキーワードとの類似度を考慮して、並び順を修正します。

対話システムは、検索結果の件数によって、共通シナリオで規定された応答を選択します。共通シナリオは、例えば、「検索結果が2件以上」の場合は「追加のキーワードを問い返す」のような、特定のFAQ集に依存しないルールから構成されます。追加のキーワードを問い返すときは、FAQを効率良く絞り込むために必要なキーワードを自動で抽出して、幾つかの選択肢を提示します。

質問応答対話システムの動作検証では、約250件のFAQ集を使った場合、1分程度で事前解析が終わり、対話システムを運用できる状態になりました。処理に使うPC(パソコン)の性能やFAQの量によって処理時間は変化します

が、手作業で専用シナリオを作る場合に比べ、システム構築の時間とコストを低減できます。

### 今後の展望

FAQ集から容易に構築できる質問応答対話システムを開発しました。現在、このシステムは、東芝デジタルソリューションズ(株)から、コメンドリ™という名称の対話システムとして提供されています<sup>(1)</sup>。

ここでは、FAQ集に基づく質問応答対話システムについて述べましたが、今後は、特定のタスクを行う対話システム<sup>(2)</sup>と組み合わせて、様々な問い合わせに柔軟に対応できる対話システムの実現に取り組みます。

### 文献

- 東芝デジタルソリューションズ, “シナリオレス型AIチャットボットサービスコメンドリ™”, <<https://www.global.toshiba/jp/products-solutions/ai-iot/taiwa-faq/>>, (参照 2021-03-25).
- 岩田憲治, 吉田尚水, 開発コストやカスタマイズコストを削減する統計的手法を用いた音声対話システム, 東芝レビュー, 2018, 73, 5, p.58-62, <[https://www.global.toshiba/content/dam/toshiba/migration/corp/techReviewAssets/tech/review/2018/05/73\\_05pdf/b08.pdf](https://www.global.toshiba/content/dam/toshiba/migration/corp/techReviewAssets/tech/review/2018/05/73_05pdf/b08.pdf)>, (参照 2021-03-25).

### 吉田 尚水

研究開発センター 知能化システム研究所 メディアAIラボラトリー 博士(工学)