

# FAQ集から自動で構築可能なQA対話システム

## Spoken Dialogue System for Question Answering based on Frequently Asked Questions

吉田尚水<sup>1\*</sup> 岩田憲治<sup>1</sup> 渡辺奈夕子<sup>1</sup> 小林優佳<sup>1</sup> 藤村浩司<sup>1</sup>  
Takami Yoshida<sup>1</sup>, Kenji Iwata<sup>1</sup>, Nayuko Watanabe<sup>1</sup>, Yuka Kobayashi<sup>1</sup>, Hiroshi Fujimura<sup>1</sup>

<sup>1</sup> 株式会社東芝 研究開発センター

<sup>1</sup> Corporate Research & Development Center, Toshiba Corporation

**Abstract:** This paper describes a conversational frequently asked question (FAQ) retrieval system that is easily applicable to new FAQs. The system searches appropriate FAQs from user's queries through multiple turns of conversation. To realize such functionality without FAQ-specific scenarios, we introduce system response selection based on the FAQ search results.

## 1 はじめに

人の代わりに問い合わせ対応を行う対話システムやチャットボットは、対応スタッフの負荷やコストを削減できるため注目されている(例えば[1, 2, 3])。特に「よくある質問(FAQ)」のように似たような質問を頻繁に受ける場合には有効に機能しやすい。

FAQ集に基づきQA対話を行うものには、一問一答で終わるものと複数ターンの絞り込みが可能なものがある。一般に、一問一答のものはシステムの構築が容易であるものの精度が低い。逆に複数ターンの絞り込みが可能なものは精度が高いものの手作業によるシナリオ作成が必要であるなど構築にかかるコストが高い。

そこで、我々は複数ターンの絞り込みが可能であり、かつ構築コストが低い対話システムの実現を目指している。今回、シナリオを作成することなくFAQ集から構築可能で複数ターンに渡る絞り込みが行える対話システムを試作した。以降では、QA対話システムの概要について紹介する。

## 2 QA対話システム

試作したQA対話システムは、ユーザがFAQ集から知りたい情報を検索する際に支援するシステムである。ユーザが困っている内容や知りたい項目について伝えると、該当するFAQを検索して提示する。その際、ユーザの入力した情報が曖昧であったり候補を



図 1: QA 対話システムの動作画面

つに絞り込むことができない場合には、複数ターンの対話を通じて絞り込みを行う。

図 1 に試作した QA 対話システムの動作画面を示す。図の左側にはユーザの入力したクエリとそれに対するシステム応答を表示している。右側には抽出した回答候補を表示している。クエリの入力にはキーボードを使ったテキスト入力と音声認識を利用した音声入力が可能である。ユーザのクエリから抽出されたキーワードのうち、FAQ集の質問文にも含まれているキーワードとシステムからの問い返しに使用されたキーワードについては強調表示している。

この対話システムは新しいFAQ集が与えられた際に、キーワードに対するラベル付けやシナリオの作成などの手作業をせずにQA対話が可能になる。これにより、例えばFAQ集にFAQを追加・削除したり、新しいFAQ集に対してQA対話システムを構築することが容易になる。

\*連絡先: (株) 東芝 研究開発センター メディア AI ラボラトリー  
〒 212-8582 神奈川県川崎市幸区小向東芝町 1  
E-mail: takami.yoshida@toshiba.co.jp

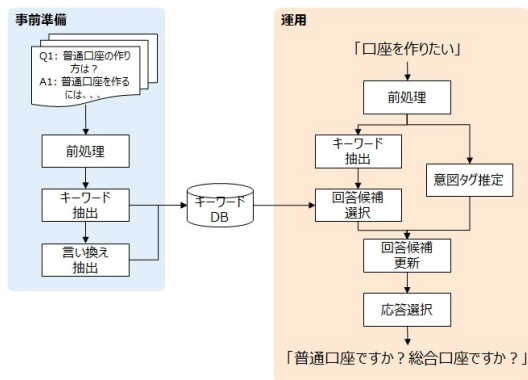


図 2: 対話処理の概要

## 2.1 処理の概要

図 2 に対話システムの処理の概要を示す。左側は事前準備として新しい FAQ 集へ対話システムを適用させるための処理を、右側は実際の運用時にユーザーのクエリへ対応するための処理を表す。事前準備では、各 FAQ に含まれるキーワードの抽出とそれらのスコアの計算を行い、FAQ 集と合わせて DB として保存しておく。運用時は、ユーザーからのクエリを解析し、事前準備で作成した DB を参照しながら回答候補の検索を行う。

**事前準備** はじめに前処理として、表記揺れによる取りこぼしを防ぐためにテキストを正規化し、単語毎に分割し品詞情報を付与する。次に、各単語の品詞情報に基づいて検索に用いるキーワードを抽出する。各キーワードへは、その TF-IDF に基づいたスコアを付与しておく。複数の単語からなるキーワードについては、それら全体と個々の単語の両方をキーワードとして抽出する。

言い換え抽出では、編集距離に基づいた方法、単語の分散表現に基づいた方法、手作業で作成した辞書(同義語辞書)に基づいた方法を使って言い換え表現の抽出を行う。これらにより、例えば「申し込み」と「申込」のような簡単な言い換え、「退会」と「解約」のような使われる意味は近いが表記が異なる言い換え、「テレビ」と「TV-123」のような型番や固有名詞などタスク固有の言い換えなどへ対応できるようになる。言い換えとして抽出されたキーワードのスコアは、元となった FAQ で使われていたキーワードのスコアを類似の度合いに応じて割り引いた値とした。

最後に上記の処理で抽出したキーワード、スコアと元の FAQ 集を DB として保存して事前準備は終了する。

**運用** ここでは事前準備と運用で異なる部分のみ説明する。意図タグ推定では、検索条件の伝達(「口座の作り方を教えて」など)の他に、システムからの問い返

しに対する肯定や否定(「はい」、「いいえ」など)、システムへの指示(「最初からやり直し」など)、挨拶などの意図を推定する。この意図推定は、単語の分散表現を入力とした RNN を用いる。

回答候補選択では、ユーザーのクエリから抽出したキーワードをキーワード DB と照合して回答候補を選択する。選択した回答候補は、その質問文に含まれるキーワードのスコアに基づきソートされる。次に、選択した回答候補と推定した意図を、直前のターンでの回答候補と統合して回答候補を更新する。推定した意図が「検索条件の伝達」の場合は回答候補の絞り込みを行い、推定した意図が「検索条件の伝達」以外の場合は、例えばセッションを打ち切って初期状態へ戻る、挨拶を返すなど意図に応じて実装されたルールに基づき対応する。

最後に、推定した意図と回答候補の件数に基づき事前に用意した応答の中から一つを選び、システム応答としてユーザーへ返す。

## 3 おわりに

FAQ 集から構築可能な QA 対話システムの概要を紹介した。このシステムは FAQ 集に依存せず、複数ターンの対話による絞り込みが可能でありながら構築コストも抑えられている。

今後は基本性能の改善に加え、より複雑なユーザー発話を理解できるよう改良を進めていく。また、このシステムはこれまでに開発した検索対話システム [4] と共通のインターフェースとなっている。これらのシステムと統合し、複数のタスクを行うことが出来る対話システムの実現へ向け取り組んでいく。

## 参考文献

- [1] 藤本ら: FAQ チャットボットの自動構築技術の開発, NTT DOCOMO テクニカル・ジャーナル, Vol. 26, No. 2 (2018).
- [2] 稲田ら: 対話型 FAQ システム「Desse (デッセ)」の利用シーンに着目した活用方法例, 第 7 回対話システムシンポジウム (2016).
- [3] 石川ら: ソフトバンクが提供する対話型 FAQ システム: APTWARE, 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会, Vol. 11, No. 4, pp. 20-24, 2015.
- [4] 岩田ら: 統計的対話技術を活用したショッピングセンター案内システム, 第 8 回対話システムシンポジウム (2017).