

参照構造を利用した質問応答システムの実装と評価

Implementation and Evaluation of Question and Answer System based on Reference Structure

岩間雄太 *1 伊藤孝行 *2 佐藤元紀 *1 森頭之 *1 鳥居義高 *1
Yuta IWAMA Takayuki ITO Motoki SATO Akiyuki MORI Yoshitaka TORI

*1 名古屋工業大学 情報工学科
Nagoya Institute of Technology

*2 名古屋工業大学 大学院 産業戦略工学専攻
Master Course of Techno-Business Administration, Nagoya Institute of Technology

As the number of services increases, questions from users also increase. In many industrial companies, The responses required by support centers are handled by hand. However, responding by hand to many user questions is very time consuming. It is also difficult for support centers to quickly and exactly respond to those user questions. Therefore, we implemented a question answering system that can retrieves documents that correspond to the user queries. Our system also scores answers and provides them to users. Scoring functions are implemented by a reference-based ranking algorithm. Finally, we evaluated the effectiveness of our prototype system.

1. はじめに

企業にとって、サポートセンターの重要性が高まっている。サポートセンターとはユーザの問題に対して、適切な解決策を提案し、問題を解決に導きユーザ満足度を高めるものである。したがって、企業ではサポートセンターを用意し、商品やサービスの問い合わせに対するサポートを充実させユーザの満足度を高めようとしている。また、近年ではIT技術の発達により企業は多種多様な商品、サービスを提供するようになった。サービスの増加に伴い、サポートセンターへの問い合わせが増加し、内容も多岐に渡るようになってきている。上記の対応には、サポートセンターの担当者はすべて人手で対応してきたが、サービスの多様化に伴う問題によって、サポートセンターへの問い合わせが増加し、人手と対応にかかる時間が今までより求められるようになった。

サポートセンターでは、ユーザの問い合わせがあると過去のユーザとのやりとりやWebサイトのサポートページに書いてある「よくある質問」を参考に解決案を提示している。また、サポートセンターへの問い合わせの大半が「よくある質問」などで対応可能な質問である。過去の問い合わせ履歴のからの検索、または「よくある質問」などで対応可能な質問に対して多くの人手と時間をさくことは企業にとって負担になる。

また、ユーザについてもサービスの多様化にともない、サービスが複雑化するといった問題も存在する。サービスの多様化によりWebサイトのサポートページが膨大になり、ユーザが必要としている情報の場所がわかりづらいという点である。Webサイトがわかりづらくユーザがサービスに不満を持ったままではユーザの満足度は下がり、ユーザがサービス、商品を使わなくなるという問題点がある。

上記のようなサービスの複雑化や多様化といった問題に対応するために、サポートセンターが手動で行っていた過去の問い合わせ履歴のからの検索、「よくある質問」で対応可能な問題の回答などを自動でおこない、ユーザの必要な情報が一目でわかるインターフェイスを持ったアプリケーションが求められている。

そこで、本論文ではある固有のドメインに関するユーザからの質問に対して過去のユーザ対応や「よくある質問」から

連絡先: 岩間雄太, 名古屋工業大学, 愛知県名古屋市昭和区御器所町, 052-732-2111, iwama.yuta@itolab.nitech.ac.jp

JimdoSupportCenter

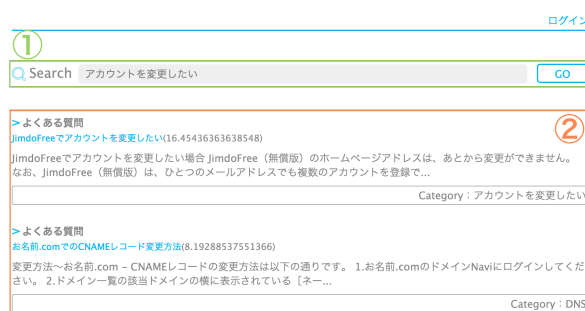


図 1: システム検索画面インターフェイス

ユーザが求めている解決案の候補とドキュメントを探し出し、探し出したドキュメントに対して参照を用いたランキングアルゴリズムを使用してスコアリングを行い、スコアリング結果に基づいて結果をユーザに提案するアプリケーションの試作について述べる。

以下に本論文の構成を述べる。1章で研究背景と目的を述べる。2章で関連研究を述べる。3章でシステムの実装、機能について述べる。4章では検索手法、およびスコアリングアルゴリズムについてを述べる。5章ではシステムの性能評価をし、既存システムに比べた有用性を述べる。6章では本論文のまとめと今後の課題を述べる。

2. 質問応答システムの実装

2.1 システム機能

本システムの検索画面インターフェイスを図1に示す。図1の①にある検索フォームを使用しユーザが質問を入力する。検索フォームの下②に、ユーザからの質問にあった解決案をシステムで設定したスコアの高い順に提案する。システムにはユーザとサポートセンター担当者の両方が質問をし解決案を得るために以下のような機能が実装した。実装した機能は検索機能、ログイン機能、文章変更機能、および変更履歴確認機能である。以下では質問応答システムのメイン機能である検索機能について詳述する。

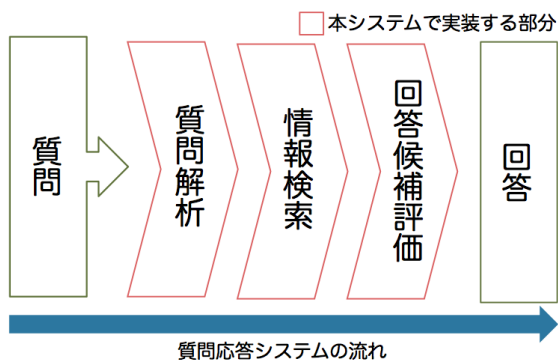


図 2: 検索機能の概要

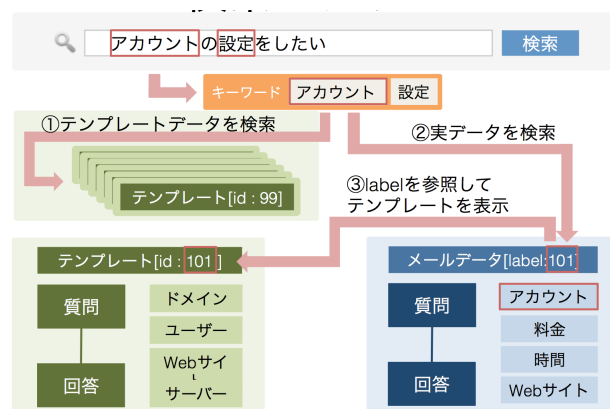


図 3: 検索方法の概要

【検索機能】

検索機能の概要図を図2に示した。ユーザが検索フォームを使用し質問を入力する(図2「質問」)。質問を解析し、システムで使用できるフォーマットに変換する(図2「質問解析」)。解析結果から、解決案の候補となるデータを取得する(図2「情報検索」)。取得したデータをランキングアルゴリズムを使用しスコアリングし評価する(図2「回答候補評価」)。最後にスコアの高い順にユーザに提案する(図2「回答」)。

質問解析では、ユーザの質問を形態素解析し、名詞のみを切り出し使用している。ユーザの質問を形態素解析する際にオープンソース形態素解析エンジンであるMeCabを用いた。名詞のみを使用した理由として、索引語の良さの指標については「文書の特徴づける語であること」と「それらをもれなく抽出すること」という2つの指標が考えられているからである[6]。

情報検索で使用するデータは2種類ある。1つ目はテンプレートデータという「過去のやりとり」や「よくある質問」のFAQデータである。2つ目はメールデータという実際のサポートセンターとユーザのやりとりのデータである。メールデータにはラベルという属性を人手により付与した。ラベルとはメールデータを作成する際に参考にした、テンプレートデータのidである。

回答候補評価では以下の5つのランキングアルゴリズムを使用した。単語の頻度、文書中で単語の出現位置、カテゴリとの一致度、テンプレートデータの参照重み、およびクリック重みである。本論文ではテンプレートデータの参照重みを利用することにより、スコアリング精度を高めている。

2.2 検索方法

本システムの検索方法について述べる。本研究では検索方法は以下3つのアプローチが存在する(図3)。1つ目の検索方法はテンプレートデータを検索する方法である(図3の①)。2つ目の検索方法はメールデータを検索する方法である(図3の②)。3つ目の検索方法はメールデータのラベルを用いて検索する方法である(図3の③)。検索方法3について以下に詳述する。メールデータのラベルとは、メールデータにはサポートセンターがユーザの質問を回答する際に参考にしたテンプレートデータを保存する属性のことである。本検索方法はメールデータが参照したテンプレートデータを検索結果とする検索方法である。本検索方法を使用することにより、1つ目の検索方法(図3の①)では検索結果になりえなかったテンプレートデータを検索結果とすることができる。つまり、より多くの質問に対する有用なデータを取得することができる。

3. 参照に基づくアルゴリズム

3.1 単語の頻度

1つ目の指標である「単語の頻度」について述べる。本指標は、ユーザが入力した質問の解析結果がドキュメントデータ、つまりテンプレートデータとメールデータ中で出現するときの出現頻度を評価する指標である。本指標はテンプレートデータとメールデータをとともに使用することが可能である。

ここでidが*i*のドキュメントデータの単語の頻度のスコアを $frequency(i)$ と定義すると、idが*i*のドキュメントデータ中に出現した単語の総数が*N*であり、ユーザが入力した質問の解析結果の単語の出現数が*n*とすると、単語の頻度は式(1)のように表すことができる。このとき、*j*はユーザの質問の解析結果の*j*番目の単語とし、解析結果の単語のidを*k*とする。式(1)はユーザが入力した質問を形態素解析し、いくつかの単語に分けて、それぞれの単語について、本指標を用いていることを表している。

$$frequency(i) = \sum_{j=1}^k \frac{n_j}{N_i} \quad (1)$$

例えば、ユーザが「アカウントを変更したい」と入力したとする。まずはシステムでユーザの質問を解析し「アカウント」および「変更」という名詞を取り出す。「アカウント」および「変更」を使用し2.2の検索方法によりドキュメントデータを検索する。このときの「アカウント」に注目する。idが100のテンプレートデータが存在したとする。idが100のデータの文書中には「アカウント」が7回出現しており、総単語数は76個であった。この時の「アカウント」に対してのテンプレートデータのスコアは $frequency(100) = \frac{7}{76} = 0.0921$ となる。

3.2 文書中での単語の出現位置

2つ目の指標である「文書中での単語の出現位置」について述べる。本指標は、ユーザが入力した質問の解析結果がドキュメント中のどの位置に出現するかを解析して評価するものである。本指標はテンプレートデータとメールデータととも使用することが可能である。具体的には、質問の解析結果となった単語がドキュメント中でどの程度早く出てくるかを評価する。つまり、解析結果の単語がドキュメント中においては早く出てくれば出てくるほど本指標の評価値は高くなる。早く出現した単語のスコアが高くなる理由として、文書の主題となる単語は最初の部分に出現する可能性が高いからである。

ここで id が i のドキュメントデータの文書中での単語の出現位置のスコアを $location(i)$ と定義すると、 id が i のドキュメントデータ中に出現した単語の総数が N であり、ユーザが入力した質問の解析結果の単語の出現位置を $index$ とすると、式 (2) のように表すことができる。このとき、 j はユーザの質問の解析結果の j 番目の単語とし、解析結果の単語の id を k とする。式 (2) はユーザが入力した質問を形態素解析し、いくつかの単語に分けて、それぞれの単語について、本指標を用いていることを表している。

$$location(i) = \sum_{j=1}^k \left(1 - \frac{index_j}{N_i}\right) \quad (2)$$

例えば、ユーザが「アカウントを変更したい」と入力したとする。まずはシステムでユーザの質問を解析し「アカウント」および「変更」という名詞を取り出す。「アカウント」および「変更」を使用し 2.2 節の検索方法によりドキュメントデータを検索する。このときの「アカウント」について見ていく。 id が 100 のテンプレートデータが存在したとする。 id が 100 のデータの文書中には「アカウント」が初めて出現したのは 2 番目で、総単語数は 76 個であった。この時の「アカウント」に対するテンプレートデータの文書中での単語の出現位置のスコアは $location(100) = 1 - \frac{2}{76} = 0.973$ となる。

3.3 カテゴリとの一致度

3 つ目の指標である「カテゴリとの一致度」について述べる。本指標は、ユーザが入力した質問の解析結果の単語とカテゴリがどの程度一致しているかを評価する指標である。本指標で使われているカテゴリとはテンプレートデータに付与されているどのようなカテゴリの質問か、という属性のことである。本指標はテンプレートデータには使用することが可能だが、メールデータはカテゴリのような属性を保持していないため適用しない。

ここで id が i のドキュメントデータのカテゴリとの一致度のスコアを $category(i)$ と定義すると、式 (3) のように表すことができる。この時、 i はドキュメントデータの id を表すものとする。

$$category(i) = \begin{cases} 1 & \text{if its in the consist category} \\ 0 & \text{otherwise out of the consist category} \end{cases} \quad (3)$$

例えば、ユーザが「アカウントを変更したい」と入力したとする。まずはシステムでユーザの質問を解析し「アカウント」および「変更」という名詞を取り出す。「アカウント」および「変更」を使用し 2.2 節の検索方法によりドキュメントデータを検索する。このときの「アカウント」に注目する。 id が 100 のテンプレートデータが存在したとする。このテンプレートデータのカテゴリが「アカウント」だった場合カテゴリとの一致度のスコアは $category(100) = 1$ となる。

3.4 テンプレートデータの参照重み

4 つ目の指標である「テンプレートデータの参照重み」について述べる。2.1 節で述べたように、メールデータはテンプレートデータを参考に回答しているものがあり、参考にした場合はメールデータに新たにラベルという属性が付与されている。本指標はテンプレートデータをスコアリングする際に当該テンプレートデータがどのくらいメールデータに参照されているかを評価する指標である。つまり、2.2 節で述べた検索方法

3 によって検索結果となったテンプレートデータが存在した際に、検索結果となったテンプレートデータを参考に回答を作成したメールデータが 1 つ以上存在するので、参考にしたメールデータのスコアを足しあわせ、合計スコアを参照数で割るということである。本指標は参照されているテンプレートデータにのみ使用することができる。

ここで id が i のドキュメントデータのテンプレートデータの参照重みを $label(i)$ と定義する。 $MailScore_j$ は参照しているメールデータを表し、テンプレートデータを参照しているメールデータの数を $RefN$ とすると、 $label(i)$ は式 (4) のように表すことができる。

$$label(i) = \frac{\sum_{j=1}^{RefN} MailScore_j}{RefN} \quad (4)$$

例えば、ユーザが「アカウントを変更したい」と入力したとする。まずはシステムでユーザの質問を解析し「アカウント」および「変更」という名詞を取り出す。「アカウント」および「変更」を使用し 2.2 節の検索方法によりドキュメントデータを検索する。このときの「アカウント」に注目する。 id が 100 のテンプレートデータが存在し、このテンプレートデータは 2 つのメールデータから参照されているとする。また、2 つのメールデータのスコアをそれぞれ、2.78 と 3.21 だとすると、テンプレートの参照重みは $label(100) = \frac{2.78+3.21}{2}$ となる。

4. 実験評価

実験の設定を述べる。サポートセンターとユーザとの実際のメールのやり取りの中から無作為に 30 個の質問を選んだ。また、クリック重みの学習のために 550 個のデータを使用した。質問を本システムと現行のシステムに入力しそれぞれの評価を行った。既存システムとは共同研究先の企業の web サイトで使われている単純な単語一致の検索システムである。本研究ではシステムの評価指標として次の 3 つの指標を用いる。1 つ目はランクの逆数の平均である。2 つ目は 5 位正解率である。3 つ目は上位 5 位における正解データの占有率である。

ランクの逆数の平均はシステムが検索結果を出力した時に、検索結果中にある正解の回答のランクの逆数の平均のことである。ランクの逆数の平均が大きいということは、回答の候補を提案する時の順序が適切であると言える。

5 位正解率はシステムが上位 5 個の検索結果を出力したときに、検索結果中に正解がある問題の割合である。質問の回答の候補となりえる回答が検索結果のより上位になるようにスコアリング関数が作用しているためである。

上位 5 位における正解データの占有率とはシステムが検索結果が出力した時に、検索結果の上位 5 位を中に正解の文書がある割合である。本指標が高ければ、ユーザからの質問に対する解決案がランク 1 以外でも正解になる可能性が高くなる。

評価結果として、全ての指標で現行システムに比べ優れた評価を得た (表 1)。

表 1: 評価結果

評価指標	現行システム	本システム
ランクの逆数の平均	0.531	0.721
5 位正解率	80.0%	89.6%
正解データの占有率	36.6%	48.2%

本結果には3つの要因がある。1つ目の要因は既存システムに比べスコアリング関数が優れていることがあげられる。本システムでは単語の頻度、文書中で単語の出現位置、カテゴリとの一致度、およびテンプレートデータの参照重みの4つの関数を使用してスコアリングを行った。特にテンプレートデータの参照重みを用いたことが、既存システムとことなり参照度の高い重要な文書のスコアを大きな数値にすることを可能にした。

2つ目の要因は既存システムに比べ検索方法が優れていることがあげられる。本システムでは、テンプレートから検索、メールアドレスから検索、およびメールアドレスのラベルを用いた検索といった3つの検索方法を利用している。特にメールアドレスを用いた検索を利用したことにより、既存システムでは検索結果として表示されないテンプレートデータも検索結果に含めることが可能となった。

3つ目の要因として、データ数の既存システムに比べ多いということがあげられる。既存システムではデータ数が約300件であるのに対し、本システムでは5245件のデータを使用している。データ数が多くなることで上記に述べた要因1と要因2も更に精度が上がっている。本結果により、本システムは既存システムに比べ優れており有用である。

5. 関連研究

ユーザの質問に対して自動で適切な回答をするシステムを質問応答システムといい、多くの既存研究が行われている。既存研究と本研究との差異を述べる。

名須川らは、コールセンターに寄せられた実際のデータを使用して、質問を自然言語処理やデータマイニングといった技術を用いて解析している [3]。また [3] では質問内のキーワードを頻度は考慮しておらず、質問がどのような文書なのかという情報を解析している。さらに、テキストマイニング技術を用い、顧客とのやりとりから有益な知識の獲得を目標としており本論文の目的とは大きく異なる。

片山らは質問の類似度をもちいて、質問者ではなく回答者に質問を適切に推薦することを目標とした [6]。推薦の適切さを確認するためにQAサイトにおける回答者の履歴と質問との類似度を定義し、実際に回答した質問とそれ以外の質問とで比較した。[6]では質問者ではなく回答者に質問を推薦するという点が本論文とは異なる。

車らはWhy型質問文に対して、Personalized PageRankを用いて、回答する手法を提案した [4]。Why型質問文とは、方法や理由を問う文章のことである。Personalized PageRankは文書の重要文を抽出する手法としてよく用いられている。Personalized PageRankを使用した理由として、Why型質問の回答として適切な説明文は、周辺の文で補足して説明されたり、回答候補文書中で重要な文となっていることが多いと仮定したからである。[4]は質問文の形をWhy型質問文に絞っているため本論文とは異なる。

また、Why型質問応答システムに関して、様々な手法が提案されてきた。渋沢らは「から、原因」といった原因や理由を表す表現と「それ、以下」といった指示表現を手がかりとして、人手で作成したパターン辞書を元に理由文を抽出する手法を提案した [5]。

石下らは、質問のタイプに依存しない質問応答システムを提案し、タイプごとに質問と回答の特徴をYahoo!知恵袋から学習し、回答の特徴と質問文との関連度を用いてWeb文書から回答部分を抽出した [7]。石下らのシステムでは大学入試センター試験で高得点を得ることを目標としている。そこで回答

を複数出力するのではなく、1つの回答が出力される。本論文では、複数の答えがある場合を想定し、複数の回答を出力しており石下らとの論文と異なる。

6. おわりに

本論文では、サポートセンターが抱えているユーザからの質問は増加しているがほとんどの質問に「よくある質問」を回答することができる。また、「ユーザが抱えているサービスが多様化している」、「複雑化しヘルプページから目的のデータを探し出すことができない」、といった問題点を解決するためにランキングアルゴリズムを用いた質問応答システムを試作した。また、試作したシステムの性能評価を行い、有用性を確認した。結果として、全ての指標で既存のシステムよりも良い結果を得ることを確認した。

今後の課題としては、システムのスコアリングの精度を上げることである。本論文のクリック重みの学習が簡易的なものであるためSVMを使用したクリック学習をすることである [1]。

謝辞

本研究で利用した文書データは株式会社KDDIウェブコミュニケーションズに提供していただいた。ここに感謝の意を表す。

参考文献

- [1] T.Joachims: Optimizing search engines using click-through data, *KDD '02 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining(2002)* .
- [2] Ito Takayuki,Iwama Yuta: An Implementation of Question Answering System Based on Reference-based Ranking Algorithm, *The 27th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2014)* .
- [3] 那須川 哲也: コールセンターにおけるテキストマイニング, *人工知能学会誌 16(2), 219-225, 2001-03-01*
- [4] 車 智修, 鍋島 啓太, 水野 淳太, 岡崎 直観, 乾 健太郎: 文書構造を用いた Why 型質問応答システム, 第 27 回人工知能学会全国大会 (2013) .
- [5] 渋沢 潮, 林 貴宏, 尾内 利紀夫: Why 型質問の回答文を WEB から自動抽出するシステムの開発と評価, *情報処理学会論文誌 48.3 (2007): pp. 1512-1523* .
- [6] 片山 亮, 川村 秀憲, 鈴木 恵二: QA サイトにおける質問推薦へ向けた履歴データの分析, *電子情報通信学会技術研究報告. AI, 人工知能と知識処理 109(439), pp. 11-16, 2010-02-22* .
- [7] 石下 円香, 狩野 芳信, 神門 典子: 質問応答システムでの解答に向けた大学入試問題の分析, 第 27 回人工知能学会全国大会 (2013) .