

How 型 Q&A システムの構築

佐藤 瞬 浦谷 則好

東京工芸大学工学部コンピュータ応用学科

1. はじめに

現在ではインターネット環境の急速な発展と通信技術の発達によって、Web には多種多様かつ膨大な情報が存在する。この膨大な情報の中からユーザが必要な情報を取り出すことは容易なことではない。情報検索では必要な情報が記述されているページを教えてくれるだけである。

Web 中の大量の情報の中から、ユーザの質問に対し最適な回答を直接に提示してくれるシステムが望まれている。本研究では非 Factoid 型であり、行動を問う質問に対応する How 型質問応答システムの構築を目指す。

2. 既存の研究

How 型質問応答システムの例として、佐々木ら[1]が提案する質問に対して行動表現と理由を回答するシステムがある。この研究では How 質問応答システムで出力された行動表現では、行動を取るべき疑問が解消されないという問題点を Why 型質問応答システムの利用によって解決している。諸岡ら[2]が提案する質問応答システムでは、与えられた質問文の解析を行い、質問文の表現から質問タイプの決定と検索のためのキーワード抽出を行う。質問文から抽出されたキーワードを検索に用いる手法である。また、質問のメインキーワードと動詞キーワードの抽出を行い、パターン化によって回答候補を導き出している。

3. 提案手法

How 型質問応答システムの全体の概要を記す。本システムは、質問応答システムを構築するときに代表的な「質問文解析」、情報検索・情報抽出、「回答候補の選択」、「回答候補ランキング」の、4 つのモジュールから構成する。

3.1 理想的な回答

How 型質問応答システムに限定して説明する。How 型質問は回答が文である必要がある。それは、How 型質問は行動や対処法などの説明を求める文を回答とするためである。

また、ユーザが必要としている情報を簡潔に表示する必要がある。さらに、ユーザの質問の意図を読み取る必要があり、柔軟に対応できるシステムが望ましい。しかし、質問文のみではユーザの意図を正確に読み取ることが難しい。そのため、本システムでは回答候補ランキング上位 3 位までの回答を提示する。これにより、回答が複数存在した場合にも対応できる。以下に How 型質問に対する回答例を挙げる。

質問：地下鉄にはどうやって電車を入れたのですか？

回答 A：相互乗り入れをしている場合は、そのまま線路が直通しているので、乗り入れ先から車両を入れれば良い。

回答 B：車両基地からトレーラーなどで運びクレーンで線路に据え付けます。

回答 C：トンネルの上に穴を開けて、クレーン車で一両ずつおろして入れます

以上のように、一つの質問に対して回答が複数存在することがある。今回例に挙げた回答はすべて正答とみなせる。本システムでは例のように、説明がなされている文を回答として提示することが理想である。

また、How 型質問応答システムでは物の作り方についての質問も考えられる。これは手順を問う質問であるため、回答文では行動する順番を考えて提示する必要がある。

3.2 質問文解析

質問文解析では、与えられた質問文に対し、形態素解析を行う。解析結果から情報検索で用いる検索キーワードと回答文のタイプを決定する。

質問の表層パターンから質問のタイプを判別する。How 型のパターン判定に利用する語句を以下に記載する。

「どういう」、「どうしたら」、「どうする」、「どうすれば」、「どうやったら」

また、単語一つで質問のタイプを判断できる要素として「方法」、「手法」、「対処法」、「手段」があげられる。これらの語句が含まれている質問文を How 型の質問であると推定する。

検索キーワードとしては名詞、動詞に注目する。情報検索の精度向上のため、回答タイプが判断できる要素を検索キーワードとして利用する。すなわち抽出した動詞の係り受け関係や受身などの条件を考慮してあらかじめ用意した検索キーワードを追加

する。

例えば、質問が「山で熊に襲われたらどうすればいいですか？」の場合、検索キーワードとして用いられるのが、「山」、「熊」、「襲われる」となる。この時に動詞は「襲われる」となり、受身の状態となるため、「対処法」を検索キーワードとして追加する。

3.3 情報検索・抽出

質問文解析の結果から得られた検索キーワードをもとに、Yahoo API を使用し、情報源の WEB から回答が含まれている可能性の高い WEB サイトを検索する。検索結果から回答候補の選択部へ渡すための情報（検索キーワードを含むホームページ情報）を抽出し、回答候補とする。取得件数は 300 件とする。

3.4 回答候補の選択

回答選択方法として以下の 3 つの手法を利用する。3 つの手法を単体で利用する場合と、結合して利用する場合がある。

3.4.1 名詞、動詞の出現頻度を利用した回答選定

抽出された情報内での名詞、動詞の出現頻度を求める。該当文中に高頻度で出現する動詞と名詞を多く含む文を回答の候補とする。ただし、How 型の回答の要素として考える単語として、“方法”、“手段”、“対処法”、“手法”を含む文も回答候補とする。

3.4.2 手がかり表現を利用した回答選定

回答を導き出す際に、特定な語の周辺に回答が集中する場合がある。手がかりとなる語句の前後合計 200 文字に回答があると

考える。手がかりとなる語句を以下に記載する。

表 1.手がかり語

以下	結論	対処法
応急処置	対処	手順
方法	処置	手法

さらに、抽出した範囲内で名詞、動詞の出現頻度を調べる。出現頻度の高い単語を多く含む文を回答として出力する。

3.4.3 ユーザに適応した動的な要約手法を利用した回答選定

回答は、ユーザが入力したクエリに即したのものになっている必要がある。塩見ら[3]が提案する手法では従来通りの文の重要度に、クエリ中の単語が文中に出現する頻度に応じたスコアを加味することで、クエリに依存した重要文抽出手法を実現している。Tombrosら[4]は、テキストのタイトル情報やテキスト中での位置情報、テキスト中の単語の出現頻度に基づいた手法を提案している。

これらの、研究成果よりクエリやテキスト中の単語が重要であることが分かる。

3.5 回答候補ランキング

回答候補の選択によって複数の回答候補が抽出される。得られた複数の回答候補をランキングするために各回答候補をスコアリングする必要がある。

大友ら[5]は、名詞の出現頻度を指標とする方法を提案している。

著者らの手法では名詞、動詞の出現頻度を利用した回答選定の場合は、高頻度の名詞と動詞が多く出現する文を回答として提示する。

手がかり表現を利用した回答選定の場合、抽出した回答候補の中で、名詞や検索キーワードとして利用した語句が多く含まれる文を回答として提示する。

ユーザに適応した動的な要約手法を利用した回答選定の場合、テキスト中の語句の出現頻度とクエリ中のキーワードの出現頻度を用いてランキング付けをする。

各手法ともランキング付けをした上位3位までを回答として提示する。

4 評価実験・考察

4.1 実験結果

How 型質問文 30 問における人手評価の結果を示す。

4.2 評価実験

表 2.出現頻度を利用した回答選定

正答数	正答率
6	20%

表 3.手がかり表現を利用した回答選定

正答数	正答率
2	7%

表 4.ユーザに適応した動的な要約手法を利用した回答選定

正答数	正答率
9	30%

表 5.出現頻度とユーザに適応した動的な要約手法の結合した回答選定

正答数	正答率
9	30%

4.3 考察

名詞・動詞の出現頻度を利用した回答選定の場合、回答が長文になる傾向が多い。これは、長文になるほど名詞・動詞を多く含むためである。そのため、回答として必要な情報以外が紛れ込むことが多く見受けられた。また、形容詞や副詞などの要素を考慮しなかったため単語(名詞・動詞)の羅列が多い文を回答として出力していた。

情報源の中に手がかりの語句が含まれないことも多い。そのため、手がかり表現を利用した回答選定の場合、回答文が情報源内にある場合でも手がかりの語句がないために回答候補として抽出されない場合が多く見受けられた。

ユーザに適応した動的な要約手法を利用した回答選定の場合、テキスト中の高頻度出現単語を利用する場合よりクエリ中の単語を利用することで正答率が向上することがわかった。回答候補文中の係り受け関係を調べることにより正答率がさらに向上することが期待できる。

回答選定方法の結合では、ユーザに適応した動的な要約手法と名詞、動詞の出現頻度を利用した。

その結果、正答率は動的な要約手法と同じになってしまった。しかし、回答候補の上位に正答が含まれる割合は大きくなった。

5 おわりに

本研究では How 型質問応答システムを構築し、検索キーワードの追加による情報検索精度の向上と 3 種類の回答選定方法を用いて回答を提示するシステムを構築した。質問文の構造解析することにより情報検索で用いることのできる要素を洗い出した。3

種類の回答選定方法を利用することにより回答が複数ある場合にも対応できることが分かった。

今後の課題として情報源であるサイトの信頼性を調べることで回答率の向上が望めると考えられる。他にも各回答候補で提示された情報から自動要約技術を用いて必要な情報のみを抽出できるようにしたいと考えている。また、同じ回答が出力された場合、複数テキストを対象とした要約手法を利用して一つにまとめることが出来る How 型質問応答システムの構築を目指していく予定である。

6 参考文献

- [1]佐々木智, 藤井敦: 取るべき行動を理由と共に答える質問応答システム how 型と why 型の統合, 言語処理学会第 15 回年次大会, pp36-39, (2009)
- [2]諸岡心, 福本淳一: 非 Factoid 型質問に対応した質問応答システム, 言語処理学会第 13 回年次大会, pp958-961, (2007)
- [3]塩見隆一, 徳田克己, 青山昇一, 柿ヶ原康二: 視点を考慮した文書要約手法の提案, 情報処理学会第 56 回全国大会講演論文集, pp104-105, (1998)
- [4]Tombros, A. and Sanderson, M.: Advantages of Query Biased Summaries in Information Retrieval. In Proc. of the 21st Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pp2-10, (1998)
- [5]大友謙一, 村上仁一, 徳久雅人, 池原悟: Why 型 QA システムにおける回答抽出方法の改良, 言語処理学会第 15 回年次大会, pp586-589, (2009)