

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

佐々木 智 藤井 敦

筑波大学大学院図書館情報メディア研究科

1 はじめに

インターネットの普及に伴い、多種多様な情報が Web に発信されるようになった。大量の Web 文書から有益な情報を抽出する代表的な手法としてキーワード検索がある。しかし、「情報要求によってはキーワードで表現しにくい」、「検索された文書のどこに欲しい情報があるか分かりにくい」という問題があり、これらを解決する手段として質問応答 (QA) がある。

QA は人工知能と自然言語処理の分野で研究されている。前者はシステム固有の形式で組織化された情報を用いて回答を推論する「推論型」である。後者は組織化されていない文書集合から回答を抽出する「抽出型」である。推論型 QA は情報の組織化が高価であり、拡張性が乏しく回答できる分野が限定される。そのため、近年では抽出型 QA に関する研究が活発である。

抽出型 QA は、対象とする質問や回答の種類によって手法を分類することができる。名称、日付、数値など客観的事実を問う質問に回答する QA は「factoid 型」、行動、原因、定義などを問う質問に回答する QA は「non-factoid 型」と呼ばれる。non-factoid 型 QA は、質問の種類により、行動や手順を問う質問に回答する「how 型」[1, 2]、原因や根拠を問う質問に回答する「why 型」[3] などに分かれる。単一の手法で non-factoid 型 QA における全種類の質問に回答する手法が提案されている [4]。しかし、本研究は how 型 QA を対象として深く探求する。

how 型 QA の研究事例として、三原ら [1] が提案した「ヘルプデスク型 QA」がある。このシステムは、行動を問う質問に対し、名詞句と動詞からなる「行動表現」を回答として出力する。例えば、「蜂に刺されたらどうすればいい?」という質問には、「針を抜く」や「流水で洗う」などの行動表現が回答となる。

しかし、ヘルプデスク型 QA には問題点がある。それは、「なぜ蜂に刺されたら流水で洗うとよいのか?」といった、行動を取るべき疑問が解消されない点にある。そこで、提示された行動の適否を判断することがユーザにとって困難である。渡辺ら [2] や Mori ら [4] の手法も、行動を問う質問に回答できる。しかし、いずれの手法も取るべき行動を回答するだけであり、その理由も合わせて答える手法は存在しない。QA で出力された回答に対して理由を抽出する手法は CLEF で研究されているも

のの、factoid 型 QA が対象である [5]。

以上の背景を踏まえ、本研究は取るべき行動を理由と共に答える QA システムを提案する。

2 提案する QA システム

2.1 概要

具体的なシステムの構成を考える前に、人間同士の対話について考察した。以下は、「蜂に刺されたらどうすればいいですか?」という how 型質問文を質問者が発してから、取るべき行動とその理由を回答者が発するまでの過程である。

質問者：蜂に刺されたらどうすればいいですか?

回答者：患部を水で洗います。

質問者：なぜ蜂に刺されたら患部を水で洗うのですか?

回答者：蜂の毒は水溶性なので、毒を洗い流すことができます。

前半の「蜂に刺されたらどうすればいいですか?」という質問と、「患部を水で洗います。」という回答は how 型 QA である。他方において、後半の「なぜ蜂に刺されたら患部を水で洗うのですか?」という質問と、「蜂の毒は水溶性なので、毒を洗い流すことができます。」という回答は why 型 QA である。

以上より、how 型 QA と why 型 QA を組み合わせることで、取るべき行動を理由と共に答える QA のモデルを考案した。当該モデルに基づく QA システムの構成を図 1 に示す。ユーザが行動を問う質問を入力すると、すでに処理された質問であるか FAQ データベースに問い合わせる。もし処理済みであれば FAQ データベースに登録されている回答を取得し、ユーザに出力する。未処理であれば QA の処理を通じて回答を抽出する。実際に抽出された行動表現と理由の具体例を表 1 に示す。

図 1 に基づいてシステムの動作について説明する。ユーザは、「蜂に刺されたらどうすればいい?」といった行動を問う質問文を入力する。「how 型 QA」は質問に対する回答として、「患部を洗う」や「アウトドアに行く」といった行動表現を出力する。前者は正解であり、後者は誤答である。「理由を問う質問文の生成」は、how 型 QA に入力された質問と出力された回答を用いて、「なぜ蜂に刺されたら患部を洗うのか?」といった理由を問

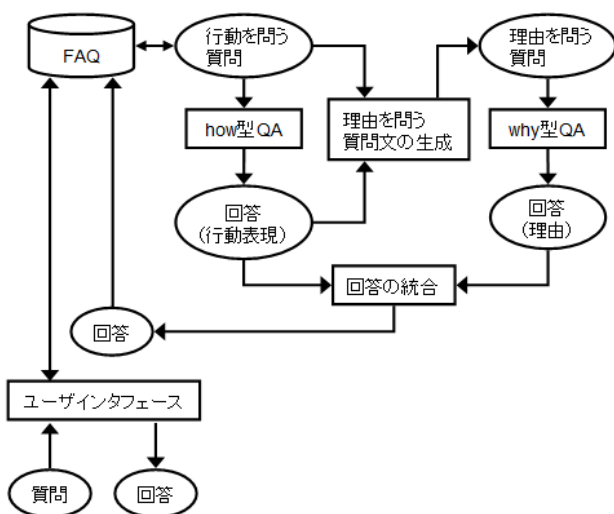


図 1: 本研究で提案する質問応答システムの構成

表 1: 抽出された行動表現と理由の具体例

質問	蜂に刺されたらどうすればいい?
行動表現	水で洗う
理由	毒が水溶性なので水で洗う。

う why 型質問文を生成する。この質問を「why 型 QA」の入力とし、「患部を洗う」という行動を取るべき理由を回答する。「回答の統合」は how 型 QA で得られた回答と why 型 QA の回答を組み合わせ、一つの回答にする。ここで、回答のスコアを再計算し、理由が抽出されなかった行動表現は順位を下げる。

先ほどの例では、「患部を洗う」という行動表現は蜂に刺された時の対処法として正しいため、why 型 QA で理由が抽出される可能性が高い。しかし、「アウトドアに行く」は、蜂に刺された時の対処法として不適切であるため、why 型 QA で理由が得られない。そこで、理由が得られている「患部を洗う」という行動表現を、理由が得られなかった「アウトドアに行く」よりも下位にする。その結果、how 型 QA 単体よりも、why 型 QA を組み合わせることでシステムの精度を高めることができる。

2.2～2.5 節で、how 型 QA、理由を問う質問文の生成、why 型 QA、回答の統合についてそれぞれ説明する。

2.2 how 型 QA

how 型 QA は、ヘルプデスク型 QA [1] を拡張して用いる。具体的には、以下の手順で回答抽出を行う。

1. 入力された質問文をクエリとして Web を検索する。
2. 検索された Web 文書から、「名詞+助詞+動詞」という係り受けを全て抽出する。ここで、「名詞+助詞」は複数回繰り返してよい。

3. 抽出した行動表現にスコアを計算し、順位を付ける。行動表現のスコア付けでは、以下に示す a～i を用いる。なお、a から e までは三原ら [1] が提案した基準であり、f から i までは本研究で新たに提案する基準である。

- a 名詞句 (名詞+助詞) と動詞の係り受け距離が近い。係り受けの距離とは、名詞句と動詞の間にある形態素数である。
- b 「～すること」や「～しましょう」のような推奨表現または「～してはいけない」のような禁止表現を伴う。推奨表現は問題解決に有効な対処法を述べる時に用いられる。禁止表現は行ってはならない対処法を述べる時に用いられ、推奨表現と同様に有用である。
- c 抽出元ページの検索結果における順位が高い。抽出元ページとは、行動表現が抽出されたページである。抽出元ページに対するスコア付けは、式 (1) を用いて行う。

$$\frac{(\text{検索ページ数} - \text{抽出元ページの順位})}{\text{検索ページ数}} \quad (1)$$

- d 質問に含まれる行動表現との距離が近い。距離とは、具体的には行動表現中の動詞と質問中の動詞の間にある形態素数である。
- e 抽出元の文でガ格の名詞句が係らない。例えば「蜂が巣を守る。」という文から「巣を守る」という行動表現が得られる。しかし、抽出元の文で「蜂が」というガ格の名詞句が係り、動作主は「蜂」と分かる。ゆえに、質問者がすべき行動として不適切である。
- f 行動表現に含まれる名詞が 3 文字以上である。名詞が 2 文字以下の行動表現は「他人に聞く」や「死に至る」など情報が少ない傾向がある。
- g Web 上に頻出する行動表現でない。「トップページに戻る」など、Web に頻出する表現は誤答であるにも拘らず抽出されやすい。そこで、Web に頻出する表現のリストを人手で作し、リストに登録されている表現はスコアを下げる。
- h 助詞が「が」や「は」でない。助詞が「が」や「は」である行動表現は名詞が主語となるため、行動の記述でなく主語に対する性質を述べている可能性が高い。例えば、「景気を回復するにはどうすればいい?」という質問に対して「物が売れる」などの表現が当てはまる。
- i 質問文に含まれる名詞を伴わない。「風邪をひいたらどうすればいい?」という質問に対して「風邪を治す」のように、質問文に含まれる名詞を含んだ行動表現はユーザの要求を満たすには不十分な内容である傾向がある。

基準 a、b、c、d、f を満たさない行動表現は誤答と判断せず、満たさない割合に応じてスコアを低くする。それに対して、基準 e、g、h、i のいずれかを満たさない行動表現は誤答と判断し、スコアを 0 にする。ゆえに、

基準 a~i を式 (2) によって統合し、行動表現 x のスコア $S(x)$ を計算する。

$$S(x) = \sum \left(\frac{1}{a_x} + b_x + c_x + \frac{1}{d_x} + f_x \right) \cdot e_x \cdot g_x \cdot h_x \cdot i_x \quad (2)$$

ここで、 x は行動表現である。 a_x は係り受けの距離、 d_x は質問との距離である。 c_x は式 (1) で計算される値である。残りの変数は 0 か 1 の値を取る。 b_x は x が推奨・禁止表現を伴えば 1、 e_x は x がガ格を伴えば 0、 f_x は名詞が 3 文字以上であれば 1、 g_x は x が Web 上に頻出する行動表現のリストに登録されていれば 0、 h_x は助詞が「が」か「は」であれば 0、 i_x は名詞が質問中に含まれる名詞であれば 0 である。

2.3 理由を問う質問文の生成

how 型 QA で得られた行動表現に対し、その行動を取るべき理由を問う why 型質問文を生成する。生成は以下の手順に沿って行う。

1. 行動を問う質問文から「動詞句+たら」または「動詞句+には」という記述を抽出する。例えば、「蜂に刺されたらどうすればいい?」という質問文からは、「蜂に刺されたら」という記述が抽出される。
2. 1. で抽出された記述と how 型 QA で回答として得られた行動表現を連結する。例えば、「蜂に刺されたら」と「患部を洗う」からは、「蜂に刺されたら患部を洗う」が得られる。
3. 2. で得られた記述に「なぜ」と「のか?」を付けて質問文とする。例えば、「蜂に刺されたら患部を洗う」という記述からは、「なぜ蜂に刺されたら患部を洗うのか?」という質問文が生成される。

2.4 why 型 QA

why 型 QA は、洪沢ら [3] の手法を拡張して用いる。具体的には、以下の手順で回答抽出を行う。

1. 入力された質問文をクエリとして Web を検索する。
2. 検索された Web 文書から、行動表現が出現する文を探す。この文を「質問相当文」と呼ぶ。
3. 特徴表現として、「理由語」、「前方指示語」、「後方指示語」、「疑問語」の 4 種類を手で定義する。特徴表現の一例を表 2 に示す。質問相当文に出現する特徴表現の種類によって、表 3 のように回答文の位置を特定する。回答文の探索は質問相当文の前後各 L 文までとする。本研究では、 $L = 5$ とした。
4. 回答の候補にスコアを計算し、順位を付ける。

式 (3) を用いて回答候補 y にスコアを計算する。

$$S(y) = r_y \times (L - l_y) \times c_y \times e_y \times q_y \quad (3)$$

r_y は y に出現する理由語が持つ重みの総和である。理由語の重みは表 2 における括弧内の数字である。抽出元ページにおいて i 番目の文が質問相当文で j 番目の文が回答文である場合、 $l_y = |j - i|$ である。 c_y は抽出元ペー

表 2: 特徴表現の例

特徴表現の種類	特徴表現の例
理由語	理由 (5), 為 (3), 原因 (5), 故 (1)
前方指示語	それが, これは, 以上が, 以上は
後方指示語	次に, 以下
疑問語	なぜ, 何故, どうして

表 3: 特徴表現の出現パターンと回答位置の特定処理

(理: 理由語、前: 前方指示語、後: 後方指示語、疑: 疑問語)

理	前	後	疑	回答位置の特定処理
	-	-		質問相当文を回答文とする
	×	×	×	質問相当文の前方から、理由語を含む文を探索し回答文とする
		×	×	質問相当文の後方から、理由語を含む文を探索し回答文とする
×	-	-	×	質問相当文の後方から、理由語を含む文を探索し回答文とする
×	×	×	×	

ジのスコアで式 (1) を用いて計算する。 e_y と q_y は 0 か 1 の値を取る。 e_y は質問相当文に含まれる行動表現にガ格が伴えば 0、 q_y は y または y の前後文に how 型 QA で入力された質問に含まれる単語が出現すれば 1 である。 r_y と l_y は洪沢らが提案するスコア付けの基準に基づいており、 c_y 、 e_y 、 q_y は本研究で提案した。

2.5 回答の統合

how 型 QA で得られた行動表現と、各行動表現に対して why 型 QA で得られた理由を一つの回答として統合する。行動表現に与えられたスコアと理由に与えられたスコアも統合する。その結果、理由が伴わない不適切な行動表現は順位が下がる。

式 (4) を用いて行動表現と理由のスコアを統合する。 $S(x)$ は行動表現 x のスコア、 $S(y)$ は行動表現 x に対して得られた複数の理由のうち、もっとも順位が高い理由 y のスコアである。

$$S(x, y) = S(x) \times S(y) \quad (4)$$

ここで、式 (4) を計算する前に、式 (5) を用いて x と y のスコアを 0.5 以上 1 未満の範囲に正規化する。

$$S_{\text{正規化後}} = \frac{1}{1 + e^{-1 * S_{\text{正規化前}}}} \quad (5)$$

3 評価実験

3.1 実験方法

「2.2 節で提案した基準 f~i」と「how 型 QA と why 型 QA の統合」によって QA の精度がどの程度向上する

かを評価した。評価は、式 (2) または式 (4) で順位付けされた行動表現に対して行い、why 型 QA で出力された理由の適否は評価しなかった。行動表現は、三原ら [1] と同じ基準で評価した。

A : 回答として適切

B : 単体では回答として不適切であり、A との組み合わせによっては有用

C : 回答として不適切 (質問に関連する表現)

D : 回答として不適切 (質問に関連しない表現)

評価値 A と B が与えられた行動表現を正解とし、以下に示す精度と再現率を用いて F 値を計算した。

$$\text{精度} = \frac{\text{上位 } N \text{ 件の正解数}}{N}$$

$$\text{再現率} = \frac{\text{上位 } N \text{ 件の正解数}}{\text{全正解数}}$$

$$F \text{ 値} = \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}}$$

以下に示す 5 種類の質問に対して F 値を計算し、それらの値を平均した。

「パスワードを忘れたらどうすればいい?」「急病にかかったらどうすればいい?」「蜂に刺されたらどうすればいい?」「ニキビができたらかうすればいい?」「やけどをしたらかうすればいい?」

3.2 結果と考察

回答数 N による F 値の推移を調べるため、行動表現を段階的に増やしていき、最大で 200 件までの行動表現について F 値を計算した。結果を図 2 に示す。グラフの横軸は行動表現の件数、縦軸は F 値を表す。図 2 中の系列 X、Z、Y、W は以下の手法に対応する。

W : 式 (2) において基準 a~e のみを使用

X : 式 (2) において基準 a~i を全て使用

Y : 式 (2) において基準 a~e のみを使用し、更に式 (4) を使用

Z : 式 (2) において基準 a~i を全て使用し、更に式 (4) を使用

W と X を比較すると、全体を通して X の方が F 値が高く、本研究で提案した基準 f~i は有効であった。

W と Y を比較すると、行動表現の件数が 30~70 および 100~180 件の場合に、Y が優位であった。X と Z を比較しても、40~60 および 110~160 件の場合に、Z が優位であり、how 型 QA 単体よりも why 型 QA を統合することの効果が部分的にあった。

しかし、行動表現の件数が 10、100、200 となる付近では why 型 QA を統合すると F 値が低下したため、原因を調査し改善につなげる必要がある。

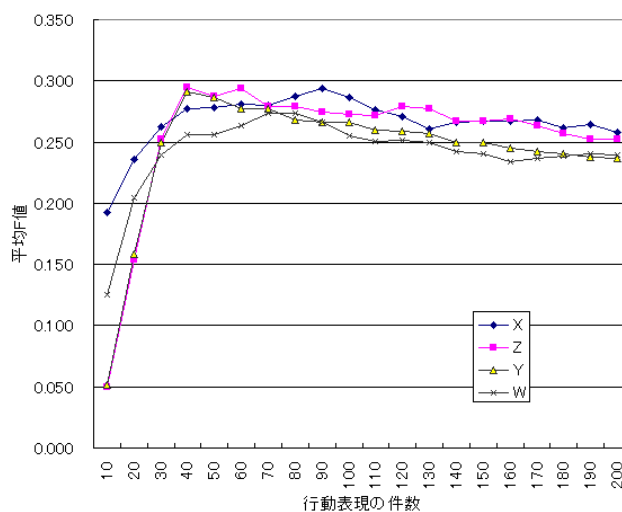


図 2: 異なるスコア付け手法の比較

4 おわりに

従来の how 型 QA には、取るべき行動を回答してもその行動を取る理由を提示しないという問題があった。この問題を解決するために、本研究は、how 型 QA と why 型 QA を統合し、取るべき行動を理由と共に答える QA システムを提案した。今後、様々な観点からシステムの評価を行う予定である。

謝辞

本研究の一部は、文部科学省科研費特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」(課題番号: 19024007) によって実施された。

参考文献

- [1] 三原英理, 藤井敦, 石川徹也. Web を用いたヘルプデスク指向の質問応答システム. 言語処理学会第 11 回年次大会発表論文集, pp. 1096–1099, 2005.
- [2] 渡辺靖彦, 園和也, 岡田至弘. メーリングリストを利用した質問応答システムのための知識獲得. 情報処理学会研究報告, 2004-NL-162, pp. 131–138, 2004.
- [3] 渋谷潮, 林貴宏, 尾内理紀夫. Why 型質問の回答文を Web ページから抽出するシステム RE:Why の試作. コンピュータソフトウェア, Vol. 24, No. 3, pp. 20–28, 2007.
- [4] Mori Tatsunori, Okubo Takuya, and Ishioroshi Madoka. A QA system that can answer any class of Japanese non-factoid questions and its application to CCLQA EN-JA task. *Proceedings of the 7th NTCIR Workshop Meeting*, pp. 41–48, 2008.
- [5] Rodrigo Alvaro, Perias Anselmo, and Verdejo Felisa. Overview of the answer validation exercise 2008. *Working notes for the CLEF 2008 Workshop*, 2008.