

述語項関係に基づく How 型質問応答の高度化

佐々木 智^{1,a)} 藤井 敦^{2,b)}

受付日 2013年6月26日, 採録日 2014年1月8日

概要: 質問応答 (QA) とは, 自然文で入力された質問に対して適切な回答を提示する処理である. 本論文は, 行動や手順を問う質問に回答する How 型 QA を探求する. How 型 QA に関する既存の手法は, 入力された質問を用いて文書集合を検索し, 検索された文書群から述語項関係を行動表現として抽出し, 適切な行動表現や回答候補の文章が上位になるようスコアを計算する. 本論文は既存手法のスコア計算における問題を解決するため, 情報検索の手法を応用する. 具体的には, 行動表現に加えて単語や複合語を索引語に組み合わせる点, 情報検索の重み付け手法である IDF を行動表現のスコア計算に応用する点, 回答候補となる文章のスコア計算に擬似適合性フィードバックを応用する点の 3 つを提案する. 評価実験によって提案手法の有効性を示す.

キーワード: 質問応答, 情報検索, 情報抽出, Web マイニング

Enhancing How-type Question Answering Based on Predicate-Argument Relations

SATORU SASAKI^{1,a)} ATSUSHI FUJII^{2,b)}

Received: June 26, 2013, Accepted: January 8, 2014

Abstract: Question answering (QA) is the process for automatically presenting an appropriate answer related to a natural language question. In this paper, we explore How-type QA associated with actions and procedures. An existing method for How-type QA searches a document collection for documents related to a question, extracts predicate-argument relations as answer candidates, computes the score for each answer candidate, and sorts the candidates according to the score in descending order. To resolve problems in the existing QA method, we apply techniques for information retrieval. In practice, we propose the following three methods: combining words and compound words with predicate-argument relations, applying the inverse document frequency to weighting predicate-argument relations, and applying the pseudo-relevance feedback to reranking answer candidates. We show the effectiveness of each proposed method experimentally.

Keywords: question answering, information retrieval, information extraction, web mining

1. はじめに

インターネットの普及にともない, 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という問題や, 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る. これらの問題点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. 人工知能における QA 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」であり, 自然言語処理における QA 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 は情報の組織化がコストが高く, そのため拡張性に乏しく回答できる分

¹ 筑波大学大学院図書館情報メディア研究科
Graduate School of Library, Information and Media Studies,
University of Tsukuba, Tsukuba, Ibaraki 305-8550, Japan

² 東京工業大学大学院情報理工学研究科
Graduate School of Information Science and Engineering,
Tokyo Institute of Technology, Meguro, Tokyo 152-8552,
Japan

a) ssatoru@slis.tsukuba.ac.jp

b) fujii@cs.titech.ac.jp

野が限定される。そのため、近年は抽出型 QA に関する研究がさかんである。

抽出型 QA の手法は、対象とする質問の種類によって分類することができる。具体的には、名称、日付、数値などの客観的事実を問う質問に回答する「Factoid 型」と、それ以外の質問に回答する「Non-factoid 型」に分類される。Non-factoid 型は、行動や手順を問う質問に回答する「How 型」、原因や根拠を問う質問に回答する「Why 型」、物事の定義を問う「Definition 型」をはじめ、Factoid 型に属さない多種多様な型に細分化される。

本論文は、How 型 QA に焦点を当てて探求する。これは、How 型 QA のみでも 1 つの研究課題として成立する難しさがあるためである。具体的には、述語と項の対で構成される述語項関係を利用した How 型 QA に注目し、その手法の改善を目的とする。述語項関係は行動を表す傾向があり、How 型の質問に対して適切な回答を探す手がかりとなる。たとえば、「蜂に刺されたらどうすればいい？」という質問には、「患部を洗う」や「病院に行く」などの述語項関係が適切な行動を表す。そして、「患部を流水で洗い、毒を洗い流してください。それでも、気分が悪ければ、病院に行ってください。」のように、上記の述語項関係を含んだ文章を回答として出力する。

以下、2 章で先行研究における本研究の位置付けを明確にする。3 章で三原らの QA 手法における問題点と本論文で提案する手法について説明する。4 章で提案手法の評価実験と考察について説明する。

2. How 型 QA の先行研究における本研究の位置付け

How 型 QA に関する先行研究の分類を図 1 に示す。図 1 では、回答を抽出する文書の種類と、その抽出手法によって先行研究を分類している。

構造化文書から回答を抽出する QA 手法には、Takechi ら [10]、麻野間ら [2]、松井ら [6]、渡辺ら [11] の研究がある。Takechi ら [10] と麻野間ら [2] は HTML 文書を回答抽出の対象とした。文書構造を定義する HTML タグのうち、

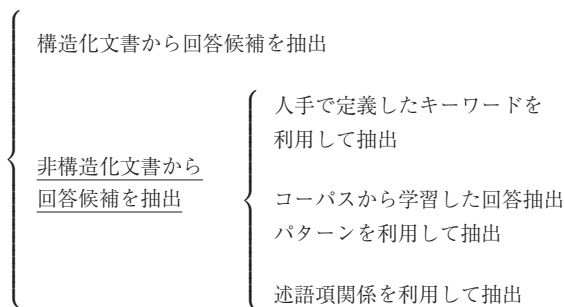


図 1 How 型 QA に関する先行研究の分類 (下線部が本研究の焦点)
 Fig. 1 Taxonomy of existing how-type QA methods (the method we focus on is underlined).

<title> や <h1> などの見出しを示すタグに囲まれた文章に質問文中のキーワードが含まれている場合は、その文書中に回答候補が含まれる可能性が高いと判定した。そして、箇条書きを示すタグである や に囲まれた文章は手順や行動を示す記述が書かれやすいという特徴に基づいて、回答候補が含まれる可能性が高いと判定された文書から、 や に囲まれた文章を回答候補として抽出した。松井ら [6] は、質問の文章と回答の文章の対応が構造化されている FAQ コーパスを利用した。ユーザが入力した質問と類似する質問を FAQ コーパスから検索し、検索された質問に対応付けられた文章を回答候補とした。渡辺ら [11] はメーリングリストに投稿されたメールを回答抽出の対象とした。メーリングリストは、あるメールに対して返信メールを送ることができる仕組みから、質問のメールとそれに対して返信された回答のメールを文書構造に基づいて特定することができる。渡辺らは、メールには最初の数行に重要な内容が書かれやすいなどの特徴に基づいて、質問のメールとそれに対する回答のメールからそれぞれ重要文を抽出することで、How 型 QA の回答抽出を実現した。

構造化文書から回答を抽出する手法には、非構造化文書から回答を抽出できないという問題点がある。そこで、回答に出現しやすい語や品詞などの言語的な特徴を分析することで、非構造化文書から回答を抽出する手法が提案されている。以下、図 1 に示すように 3 種類の手法に細分化して説明する。

人手で定義したキーワードを利用する手法として、Fukumoto [3] は「手順」や「方法」などの行動を記述している文章に出現しやすいキーワードを含むか否かを回答抽出パターンとして、回答を抽出した。Yin [16] は、「手順」や「方法」などのキーワードを含む文書を教師データとして用いることで、任意の文書を手順や行動について記述している内容かどうか分類した。しかし、これらの手法では、「手順」や「方法」などの手順や行動を示すキーワードを含んでいない文書から回答を抽出することができない。

コーパスから学習した回答抽出パターンを利用する手法として、Mori ら [8] は FAQ コーパスから回答抽出パターンを学習して回答を抽出した。Harada ら [4] は、修辞構造を利用して回答を抽出した。具体的には、単語間に意味的な関係がある場合、その意味に応じたラベルを付けた。そして付けられたラベルのうち、「implement」、「sequence」、「condition」、「manner」、「scene」という手段や方法を表現するラベルを手がかりとして回答を抽出した。しかし、Mori らと Harada らの手法はどちらも回答抽出パターンを得るために学習データを必要とし、コストが高い。また、学習データに存在しないパターンをとまなう回答は抽出できないという問題がある。

述語項関係を利用する手法は、特定のキーワードやコー

パスに依存することなく、回答を抽出することができる。小澤ら [9] と山本ら [12] は、「カーテンを閉める」のようにモノとその使われ方を表している述語項関係を回答候補として抽出した。しかし、回答抽出のみを研究対象としており、抽出された述語項関係が How 型質問に対して適切な回答か否かを評価する回答評価の手法を提案していない。旭ら [1] は、ある話題や対象に対してユーザがとる行動の連鎖を抽出する手法を提案しており、How 型 QA に応用できる。しかし、回答評価においてユーザとの対話的な要素を含んでおり、ユーザが適切な回答候補を選ぶ手間がかかる。三原ら [7] は、ユーザとの対話を必要とせず、「～すべき」というフレーズがともなうか否かなど、適切な行動に特有の特徴を回答評価に利用した How 型 QA を提案した。

以上より、述語項関係は特定のキーワードやコーパスに依存せずに回答を抽出できる点で優れており、その中でユーザとの対話を必要とせず機械的な回答評価が唯一可能である三原らの手法 [7] を本論文の比較対象とする。当該手法は、回答評価において3つの解決すべき点がある。1つ目に、どの質問にも共通して出現しやすい不適切な述語項関係を回答としてしまうという問題点がある。2つ目に、正解の記述が行動を表していない場合や、係り受け解析の誤りにより述語項関係が述語と項で分かれてしまう場合があり、述語項関係以外の手がかりが回答評価に必要となる。しかし、既存手法は述語項関係以外の手がかりを使わないため、これらの場合に正解の文書を低く評価してしまうという問題点がある。3つ目に、既存の情報検索手法を応用することで性能を改善できる可能性がある。本論文は、以上の3点についてそれぞれ解決する手法を提案する。

3. 本論文で提案する QA システム

3.1 概要

本論文で提案する How 型 QA システムの構成を図 2 に示す。構成要素として、図 2 の左側は既存の QA システム [7] に基づいている。ただし、回答評価における一部を本論文で修正している。また、図 2 の右側は既存の情報検索手法である擬似適合性フィードバック (Pseudo-Relevance Feedback: PRF) であり、QA システムと PRF の連携を本論文で提案する。PRF の詳細は 3.4.2 項で説明する。図 2 に示す QA システムの構成に基づいてシステムの動作を説明する。ユーザが「～したら (どうすればいい)?」という形式で行動を問う質問を入力すると、「文書検索」が入力された質問と関連のある文書を「文書集合」から収集する。「行動表現抽出」は収集された文書から「針を抜く」や「患部を洗う」などの述語項関係を行動表現として抽出し、各行動表現に対して「回答評価」が回答としての適切さに基づいてスコアを計算する。

「回答評価」では、行動表現の詳細を記述した文章にもスコアを計算する。具体的には、その行動表現が抽出された

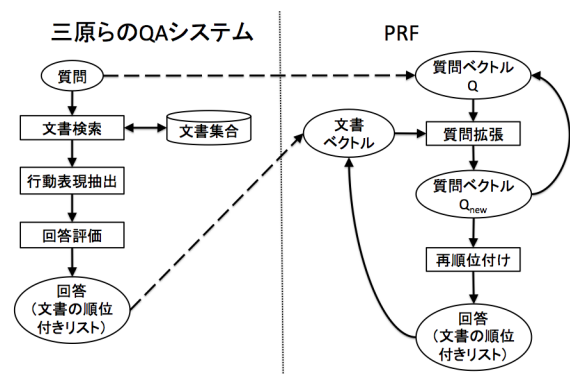


図 2 本論文で提案する How 型 QA システム
Fig. 2 Overview of our How-type QA system.

文書や文書中の段落を回答候補となる文章として扱い、文章中の行動表現に付けられたスコアに基づいて文章のスコアを計算する。最終的に、端的な回答である行動表現と、記述的な回答である文章をそれぞれスコアに基づいて順位付けし、当該システムの出力とする。実際に当該システムが出力した回答の具体例を以下に示す。

質問：やけどをしたらどうすればいい?
 回答候補の文章*1：1、ヤケドをしたら、すぐ水をかけること。服の上から熱湯をかぶった時は、服を脱がさずにそのまま水をかけること。2、塩水を作って、タオルを浸し、それで患部を冷やすのも良い。これだと傷みがひきます。手足のヤケドの場合、塩水の中へつつこんで。

本論文は、「回答評価」における行動表現と文章のスコア計算を改善することによって、回答の順位付けを高精度化することを目的とする。3.2 節～3.4 節で図 2 中の各構成要素について詳しく説明する。本論文の提案手法は 3.4.2 項で示す。

3.2 文書検索

本論文では、文書検索に Yahoo! JAPAN *2を用いる。Yahoo! JAPAN は入力された質問を解析し、その質問に含まれる単語やフレーズを含んでいる文書を検索結果として返す。

3.3 行動表現抽出

述語項関係 (述語と項の対) は必ずしも、行動を表す表現であるとは限らない。たとえば、「患部を洗う」は行動を表す一方、「患部が痛む」は行動を表していない。このように行動を表す述語項関係を、本論文では「行動表現」と定義する。ただし、次のいずれかの条件を満たす述語項関係は、行動を表していても行動表現として抽出しない。

*1 <http://www.n-bandai.com/yakedo.htm> (2011 年 3 月 3 日参照, 原文をそのまま掲載しているため, 不自然な空白などを含む)
 *2 <http://developer.yahoo.co.jp/webapi/search/>

まず、「気」、「ある」、「する」、「なる」、「やる」という適切な行動表現には使われにくい単語を持つ述語項関係は行動表現としない。次に、代名詞や「の」や「くださる」といった非自立的な名詞や動詞など、適切な行動表現を構成しにくい品詞を持つ述語項関係は行動表現としない。

行動表現抽出では、文書検索で収集された各文書に対して CaboCha^{*3}を用いて係り受け解析を行い、係り受け関係にある後置詞句と動詞の組を抽出対象となる述語項関係とする。サ変接続の名詞も動詞と見なす。たとえば、「患部を流水で洗う」という記述には「患部を洗う」と「流水で洗う」という2つの述語項関係が含まれる。

抽出された行動表現は、「患部を洗う」を「患部：洗う」のように助詞を省いた形式にする。助詞を省くのは「蜂に刺される」と「蜂が刺す」のように、格交代により助詞が変化しても同じ行動表現と見なすためである。また、「top：戻る」と「TOP：戻る」のように、大文字と小文字や全角と半角の違いがあっても同じ行動表現と見なすため、行動表現に英数字や記号が含まれる場合は、全角の大文字に統一する。なお、既存の述語項関係を抽出するツールとして SynCha^{*4}がある。しかし、ガ格、ヲ格、ニ格をともなう述語項関係しか抽出できず、本 QA システムでは「流水で洗う」のようにその他の格助詞をともなう述語項関係も正解となる場合があるため、使用しなかった。

3.4 回答評価

3.4.1 既存の手法

回答評価では、行動表現とその行動表現が含まれている文章に対してスコアを計算する。特にスコアの高い行動表現および文章が適切な回答候補として出力される。具体的には、以下に示す (a)~(g) の基準に該当する行動表現ほど大きなスコアを与える。ここで、(a)~(e) は三原ら [7] が提案した基準である。(f) は文献 [13]、(g) は文献 [14] において提案された Factoid 型 QA に関する基準である。これらも既存の手法に含め、以降では三原らの手法の一部として扱う。

(a) 後置詞句と動詞の距離が短い。

後置詞句(名詞+助詞)と動詞の距離とは、両者の間にある形態素数である。この距離が短いほど、その後置詞句と動詞の関連が強いと考える。

(b) 推奨表現や禁止表現である。

「~すること」や「~しましょう」のような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。「~してはいけない」のような禁止表現は行ってはならない対処法に用いられ、推奨表現と同様に有用である。具体的には、正規表現を用いて動詞の直後に以下のパターンが出現する行動表現を推奨表現または禁止

表現と見なす。以下は推奨表現を識別するための正規表現である。

```
(ください|ましよう|必要(だ|です|がある|あります)|  
(ておく)? ことが(大切|大事){0,8}|(て|で)おこう|  
(する|させる)(の|方|こと)(が|も)(よい|いい|良い){0,8}[。]*$
```

なお、正規表現内で任意の文字列長を 0~8 文字の間で定義している箇所がある。これは三原ら [7] が定義した表現をそのまま利用している。また、以下は禁止表現を識別するための正規表現である。

```
((てはいけ(ない|ません)|ない[^、。]*)[。]*$)|な
```

(c) 検索結果における順位が高い文書から抽出されている。文書検索において順位が高かった文書は、その検索手法において質問と関連性が強いと判断された文書であり、正解の回答が含まれる可能性が高い。

(d) 質問に含まれる行動表現との距離が近い。

ここでいう距離とは、検索された文書において、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。この距離が短いほど、その行動表現は質問に対して関連が強いと考える。

(e) ガ格の後置詞句が係らない。

手続き型の説明においては、有意味動作の述語に対する動作主が二人称になる。How 型 QA においては、回答としてこのような文章を解釈すると、手続きの動作主が質問者になる。しかし、日本語においては二人称代名詞が明示されることはほとんどない。一方で、ガ格は多くの場合、深層において動作主に対応するため、ここが他の名詞で専有されている場合には、手続きの一部を表さない。

(f) 質問に含まれる述語項関係や名詞をともなわない。

質問に含まれる述語項関係や名詞をともなう行動表現は、新情報が含まれず、正解と見なすには情報に具体性が欠ける可能性が高い。たとえば、「やけどをしたら」という質問では、「やけどをする」という述語項関係が含まれる。しかし、「やけどをする」は新情報が含まれず、回答として意味をなさない。一方、質問に含まれる名詞をともなう行動表現として「やけどを治す」や「やけどを処置する」などが得られる。しかし、やけどを治すための行動として具体性に欠ける。

(g) 不要表現辞書に登録された表現ではない。

「トップページに戻る」などの Web に頻出する表現は、特定の質問に対する回答であることが少ない。そこで、Web に頻出する表現のリストを手手で作り、不要表現辞書とする。また、4章の評価実験で用いた質問とは別に用意した70件のHow型質問と、その質問に対する適否が判定された行動表現の集合を用いて、

^{*3} <http://chasen.org/~taku/software/cabocha/>

^{*4} <https://www.cl.cs.titech.ac.jp/~ryu-i/syncha/>

表 1 不要表現の例

Table 1 Examples of our stopwords.

不要表現辞書に登録された表現	元の行動表現
JAVASCRIPT:タイオウ	JavaScript に対応する
RSS リーダー:トウロク	RSS リーダーに登録する
TOP:モドル	TOP に戻る
オキニイリ:ツイカ	お気に入りに追加する
ケイタイ:アクセス	携帯にアクセスする
コメント:トウコウ	コメントを投稿する
ブックマーク:ツイカ	ブックマークに追加する

不適切な行動表現の構成要素としてのみ出現した名詞や動詞を持つ述語項関係を不要表現辞書に登録する。不要表現辞書には、行動表現を構成する名詞と動詞の読みを登録し、表記どおりではなく読みで照合する。不要表現の例を表 1 に示す。

以上の基準 (a)~(g) に対応するパラメータ $a(x_i) \sim g(x_i)$ を式 (1) によって統合し、行動表現 x のスコア $s(x)$ を計算する。検索された文書群において行動表現 x は繰り返し出現することがあるため、 i 番目に出てきた x を x_i と表記する。式 (1) では i 番目の行動表現 x_i ごとにスコアを求め、それらの総和を x のスコアとする。

$$s(x) = \sum_i \left(\frac{1}{a(x_i)} + b(x_i) + c(x_i) + \frac{1}{d(x_i)} \right) \cdot e(x_i) \cdot f(x_i) \cdot g(x_i) \quad (1)$$

式 (1) の各パラメータと値は、以下のとおりである。

- $a(x_i)$: x_i を構成する後置詞句と動詞の間にある形態素数
- $b(x_i)$: x_i が推奨・禁止表現をとらなければ 1, それ以外は 0
- $c(x_i)$: $\frac{(\text{検索された文書数}) - (\text{抽出元文書の順位})}{\text{検索された文書数}}$
- $d(x_i)$: x_i に含まれる動詞と質問に含まれる動詞の間にある形態素数
- $e(x_i)$: x_i にガ格の後置詞句が係らなければ 1, それ以外は 0
- $f(x_i)$: x_i の名詞が質問に含まれなければ 1, それ以外は 0
- $g(x_i)$: x_i が不要表現辞書に登録されていなければ 1, それ以外は 0

$a(x_i)$ と $d(x_i)$ は、値が小さいほど正解として高いスコアを与えるため、式 (1) では $a(x_i)$ と $d(x_i)$ の逆数を用いる。また、文章 p のスコア $r(p)$ を求める手法として、三原ら [7] は式 (2) のように文章 p に含まれる行動表現のスコアを総和をとる。

$$r(p) = \sum_{x \in p} s(x) \quad (2)$$

3.4.2 本論文で提案する手法

2章で説明したとおり、既存の回答評価手法は、3つの解決すべき点がある。1つ目に、どの質問にも共通して出現しやすい不適切な述語項関係を回答としてしまうという問題点がある。2つ目に、述語項関係以外の手がかりが回答評価に必要となる場合に正解の文書を低く評価してしまうという問題点がある。3つ目に、既存の情報検索手法を応用することで性能を改善できる可能性がある。本論文は、以上の3点についてそれぞれ解決する手法①~③を提案する。なお、①~③は単独で適用可能なだけでなく、任意の組合せによる適用も可能である。

① 行動表現のスコアを IDF で修正する。

どの質問にも共通して出現しやすい不適切な述語項関係を回答としてしまうという問題点への対策として、情報検索の重み付け手法である IDF (Inverse Document Frequency) を行動表現のスコア計算に組み合わせる。IDF は、「情報を集める」や「他人に聞く」など、多くの質問に共通して出現しやすく、特定の質問に対する回答であることが少ない行動表現に対してスコアを下げる効果がある。具体的には、式 (3) により、式 (1) において得られた行動表現のスコアを IDF で修正する。修正後の $s'(x)$ は式 (2) において $s(x)$ の代替として使用する。

$$s'(x) = s(x) \cdot IDF(x) \quad (3)$$

IDF は、式 (4) で計算する。

$$IDF(x) = \log \left(\frac{N}{DF(x)} + 1 \right) \quad (4)$$

式 (4) において、 N は文書集合中の総文書数である。ただし、Web 全体の文書数は分からないので、 N として十分に大きな定数を与える。本論文の評価実験では、 $N = 500$ 億とした。 $DF(x)$ は x をクエリとして Yahoo! JAPAN ウェブ検索 API^{*5}により得られたヒット件数とした。

② 行動表現に加えて、単語と複合語も回答評価に利用する。

正解の記述が行動を表していない場合や、係り受け解析の誤りに起因して述語項関係が抽出されない場合があるため、述語項関係以外の手がかりが回答評価に必要となる。しかし、既存手法は述語項関係以外の手がかりを使わないため、これらの場合に正解の文書を低く評価してしまうという問題点がある。たとえば、以下の正解事例においては下線で示す正解の記述が行動を表していない。

*5 <http://developer.yahoo.co.jp/webapi/search/> (2013年3月21日に検索)

質問	怪我をしたら
スニペット*6	<p>先ず化膿はしないし、すでに何日か経って化膿している傷でも、早急に膿は消え、治って行きます。とに角、傷痕が残らないのが特徴で、女性や子供が顔に怪我をしたら、馬油以上の妙薬は無いですよ。手術傷。馬油を二十年以上も愛用して ...</p>

本正解事例のスコアを高くする手法として、「馬油」や「妙薬」といった質問との関連語を利用する手法が考えられる。そこで、回答候補に含まれる単語と複合語も回答評価に利用する。具体的には、形態素解析によって回答候補の文章を単語に分割し、品詞が表 2 以外の単語をすべて抽出する。ただし、名詞が連続する場合は複合語として扱い、単語と同様にスコアを計算する。単語と複合語のスコアはその IDF とする。IDF の計算は式 (4) を利用する。行動表現のスコアは式 (3) の値を使う。ただし、比較対象として既存手法では式 (1) の値を使う。式 (2) において、行動表現だけでなく単語と複合語のスコアも $s(x)$ として適用し、文章のスコアとして総和する。なお、TF は使用しない方が有効であると実験的に確認したため、IDF のみを適用する。

③ 文書のスコア計算において PRF を使用する。

既存の情報検索手法を応用することで本 How 型 QA の性能を改善できる可能性がある。具体的には、式 (2) の代わりに PRF を使用する。PRF により、質問と関連のある文書を適合文書と見なして質問拡張に利用することで、質問に含まれていないものの質問と関連する語を含む回答を上位にすることができる。図 2 に基づいて PRF の処理を説明する。あらかじめ、本 QA システムにおいて入力された質問と回答候補として得られた各文書をそれぞれベクトルとして表現する。文書をベクトルとして表現する際に利用する索引語 (index term) として、本 QA システムで抽出された行動表現を用いる。また、提案手法 ② より単語と複合語も索引語として用いる。質問拡張において、擬似適合文書に含まれる索引語の重みは大きく、擬似不適合文書に含まれる索引語の重みは小さくなるように質問中の索引語の重みを修正する。再順位付けにおいて、修正後の質問ベクトルと文書ベクトルの類似度を計算する。そして、類似度の降順に文書を再順位付けし、最終的な回答候補の順位付きリストとする。再順位付けで得られた文書の順位付きリストを質問拡張に再利用することで、繰り返し再順位付けを行うことが可能である。しかし、本論文では一度の再順位付けにとどめる。

PRF では、平松らの手法 [5] を用いる。まず、式 (5) により質問を索引語の 2 値ベクトルとして表現する。

*6 <http://www.bahyu.com/baayu/1-20/kega.html> (2009 年 6 月 2 日参照)

表 2 回答評価において使用しない品詞

Table 2 Parts-of-speech not used for answer evaluation.

感動詞, 記号, 形容詞-非自立, 形容詞-接尾, 助詞, 助動詞, 接続詞, 接頭詞, 動詞-接尾, 動詞-非自立, フィラー, 名詞-接尾, 名詞-代名詞, 名詞-動詞非自立, 名詞-特殊, 名詞-非自立
--

*品詞名の表記は IPA 品詞体系*7に基づく。

$$Q = (q^{t_1}, q^{t_2}, \dots, q^{t_N}) \tag{5}$$

q^{t_N} は索引語 t_N が質問に含まれていれば 1 であり、含まれていなければ 0 である。また、文書のベクトル表現は式 (6) のとおりである。

$$D_j = (d_j^{t_1}, d_j^{t_2}, \dots, d_j^{t_N}) \tag{6}$$

$d_j^{t_N}$ は j 番目の文書ベクトルにおける索引語 t_N の重みであり、質問に対して正解を見つける手がかりとなる索引語ほど重みを大きくする。具体的には、索引語が行動表現である場合は、式 (3) の値を重みとする。索引語が単語や複合語である場合は、その IDF を重みとする。

質問拡張において、質問ベクトルと文書ベクトルの類似度が閾値以上となる文書を擬似適合文書とし、閾値未満の文書を擬似不適合文書とする。そして、式 (7) を用いて質問ベクトル Q を Q_{new} に修正する。

$$Q_{new} = (q^{t_1} + w^{t_1}, q^{t_2} + w^{t_2}, \dots, q^{t_N} + w^{t_N}) \tag{7}$$

式 (7) において、 w^{t_N} は索引語 t_N の修正された重みである。 w^{t_N} は式 (8) より求める。式 (8) では、質問ベクトルと文書ベクトルの類似度が閾値 α 以上の文書を擬似適合文書とする。そして、擬似適合文書に含まれる索引語は重みが大きく、擬似不適合文書に含まれる索引語は重みが小さくなるように修正値を得る。なお、類似度は内積により求める。

$$w^{t_k} = \frac{\sum_{j=1}^T \text{sortscore}(Q, D_j)}{U} \cdot \frac{Ucount(t_k)}{U} - \frac{Lcount(t_k)}{T-U} \tag{8}$$

式 (8) における、各パラメータの意味は以下のとおりである。

- $\text{sortscore}(Q, D_j)$: Q と d_j の内積が閾値 α 以上の場合はその内積とし、それ以外は 0
- $Ucount(t_k)$: Q との内積が α 以上の文書ベクトルに出てくる索引語 t_k の出現回数
- $Lcount(t_k)$: Q との内積が α 未満の文書ベクトルに出てくる索引語 t_k の出現回数
- U : Q との内積が α 以上となる文書ベクトルの数
- T : 対象の質問について、図 2 の「文書検索」で収集された文書の総数

*7 <http://www.unixuser.org/~euske/doc/postag/>

表 3 評価実験に用いた質問と各手法の比較

Table 3 Test questions and query-by-query comparison for different methods.

質問	AP			
	手法 A	手法 B	手法 C	手法 D
インフルエンザにかかったら	0.450 (↓)	0.501 (↓)	0.684 (↑)	0.550
浮気をされたら	0.652 (↓)	0.696 (↑)	0.566 (↓)	0.656
MRSA にかかったら	0.212 (↑)	0.193 (↑)	0.113 (↓)	0.189
お金に困ったら	0.783 (↓)	0.828 (↓)	0.702 (↓)	0.831
火事になったら	0.604 (↓)	0.797 (↓)	0.759 (↓)	0.828
急病にかかったら	0.612 (↓)	0.819 (↓)	0.769 (↓)	0.893
車が故障したら	0.653 (↑)	0.586 (↓)	0.643 (↓)	0.647
クレームが入ったら	0.500 (↓)	0.633 (↓)	0.653 (↓)	0.718
警察に逮捕されたら	0.476 (↓)	0.706 (↓)	0.640 (↓)	0.780
怪我をしたら	0.539 (↓)	0.725 (↓)	0.705 (↓)	0.761
子供ができたら	0.620 (↓)	0.713 (↓)	0.734 (↓)	0.756
献立に困ったら	0.928 (↓)	0.966 (↓)	0.943 (↓)	0.971
事故にあったら	0.762 (↑)	0.752 (↓)	0.860 (↑)	0.759
障害者になったら	0.603 (↓)	0.740 (↑)	0.525 (↓)	0.714
生活に困ったら	0.880 (↓)	0.959 (↑)	0.902 (↓)	0.955
セクハラを受けたら	0.586 (↓)	0.765 (↓)	0.705 (↓)	0.777
相談を受けたら	0.414 (↑)	0.350 (↑)	0.361 (↑)	0.338
盗難にあったら	0.540 (↓)	0.845 (↓)	0.786 (↓)	0.874
ニキビができたら	0.751 (↓)	0.925 (↓)	0.880 (↓)	0.933
パスポートを紛失したら	0.672 (↓)	0.765 (↓)	0.654 (↓)	0.767
パソコンが壊れたら	0.513 (↑)	0.527 (↑)	0.442 (↑)	0.502
蜂に刺されたら	0.707 (↓)	0.900 (↑)	0.772 (↓)	0.883
鼻血が出たら	0.848 (↓)	0.966 (↑)	0.922 (↓)	0.958
人が倒れていたら	0.791 (↓)	0.930 (↑)	0.876 (↓)	0.923
プレゼントをもらったら	0.275 (↓)	0.250 (↓)	0.405 (↑)	0.320
発作がおきたら	0.606 (↓)	0.648 (↓)	0.711 (↑)	0.707
迷子になったら	0.647 (↓)	0.828 (↓)	0.622 (↓)	0.830
迷惑メールが届いたら	0.730 (↑)	0.622 (↓)	0.662 (↑)	0.645
目が疲れたら	0.881 (↓)	0.937 (↓)	0.911 (↓)	0.945
やけどをしたら	0.880 (↓)	0.993 (=)	0.958 (↓)	0.993

↓：手法 D より AP が低い質問，↑：手法 D より AP が高い質問，=：手法 D と AP が同じ質問

ただし、内積の閾値 α として適切な値を決めることが難しいため、本論文では具体的な閾値を設定せず、質問に含まれる索引語を 1 件以上含んでいる文書を擬似適合文書とする。

再順位付けでは、修正された質問ベクトル Q_{new} と各文書ベクトルとの内積を計算し、内積の降順にソートする。最終的に、その上位の文書を適切な回答候補とする。

4. 評価実験

4.1 実験結果

評価には、三原ら [7] が収集した 30 件の質問を用いた。当該 30 件の質問を表 3 に示す。なお、表 3 における「AP」については式 (9) で説明する。各質問を Yahoo! JAPAN に入力して Web 検索を行い、上位 100 件のスニペットを収集した。ただし、スニペットの抽出元文書がリンク切れで存在しない場合や、Flash や HTML のフレーム構造が

使われているためテキストの抽出が困難な場合は、そのスニペットを評価対象から外した。結果として、評価に使用されたスニペット数は最も少ない場合で 1 つの質問に対して 66 件となり、平均で 93 件となった。正解判定は著者以外の者が行い、スニペットに対して正解の記述を含んでいるかどうかの 2 値判定を行った。すなわち、QA を「スニペットの順位付け問題」とすることで、既存の情報検索手法との比較を可能にした。

評価尺度には、最も順位が高い正解だけを考慮する MRR (Mean Reciprocal Ranking) ではなく、本研究のように、正解が複数存在することを前提として精度と再現率を合わせて評価する MAP (Mean Average Precision) を用いた。MAP は、質問ごとに式 (9) を用いて AP (Average Precision) を計算し、30 件の質問における AP の平均により計算した。なお、式 (9) では計算式に再現率が含まれていない。しかし、各正解が得られるまでの精度 P_i を求め

表 4 MAP による各手法の比較

Table 4 Comparison of MAP for different methods.

手法	A	B	C	D
MAP	0.637	0.729	0.696	0.747

るうえで、 i より低い順位で出現する正解は回答漏れとなる。この回答漏れが少ないほど AP の値は高くなるため、再現率を AP の計算に直接利用していないものの、再現率も考慮された評価尺度となっている。

$$AP = \frac{1}{n} \sum_n P_i \quad (9)$$

- n : 正解の総数
- i : i 番目の正解
- P_i : 上位の回答候補から順に i 番目の正解まで出力された場合の精度

MAP は収集されたすべてのスニペットを対象に計算した。また、既存の手法と提案手法の MAP に有意差があるか、統計検定により確認した。比較する 2 つの手法において、どちらがより高い評価値を持つかが比較するまで不明である場合、片側 t 検定よりも両側 t 検定の方が適切である [15]。本実験では既存の手法と提案手法のどちらが高い MAP を持つかが不明であり、両側 t 検定を用いた。本実験では、以下に示す 4 つの手法 A~D をスニペットの順位付けという観点から比較した。

A: Yahoo! JAPAN

B: ベクトル空間モデルに基づく情報検索手法

C: 三原らの手法

D: 手法 C に ①~③ を組み合わせる

A は、Yahoo! JAPAN で検索した結果である。B は、A の結果を再順位付けする。具体的には、単語と複合語を索引語として、③の PRF を用いる。索引語の重みは、式 (4) を使って付ける。C は既存の How 型 QA であり、行動表現のスコア計算に式 (1)、文章のスコア計算に式 (2) を用いる。D は本論文の提案手法である。

手法 A~D を MAP で比較した結果を表 4 に示す。表 4 より、手法 D が最も高い MAP を示した。両側 t 検定において手法 D と比較した結果、手法 A~C とはすべて有意水準 1% で有意差があった。すなわち、本論文で提案した手法の組合せによって MAP が有意に向上したことが分かった。

4.2 提案手法の有効性分析

4.2.1 提案手法を個別に使用した場合の分析

提案手法 ①~③ の有効性を個別に調べるために、①~③ の組合せ方を変えて評価した結果を表 5 に示す。表 5 により、①~③ は 1 種類または 2 種類のみ適用すると手法 B や C よりも MAP が下がる場合があった。一方、①~③ をすべて組み合わせた場合に最も MAP が高くなり、手

表 5 提案手法の組合せ方による MAP の比較

Table 5 Comparison of MAP for different combinations of proposed methods.

提案手法			MAP
①	②	③	
○			0.698
	○		0.640
		○	0.682
	○	○	0.729
○		○	0.682
○	○		0.703
○	○	○	0.747

○: 組合せに用いた手法

法 B や C よりも優れていた。ゆえに、提案手法はすべてを組み合わせることで相乗効果が得られ、最も有効であることが分かった。

以下、①~③ それぞれにおいて、順位が改善された正解事例を示す。各スニペットの例において、下線が引かれた箇所は質問に対して適切な行動表現である。また、太字の箇所は正解の記述である。

質問	事故にあったら
スニペット*8	トップページ / 各課一覧 / 保険医療課トップ / 交通事故にあったら。 交通事故にあったら、 <u>すぐに警察に届ける</u> と同時に、 <u>国保担当窓口にも届出</u> （第三者行為による傷害届）をしなければなりません。届出がないまま診療を受けると、国保が使えない場合があります。...

本スニペットは手法 A~C において順位がそれぞれ 57 位、66 位、7 位であり、手法 C に ① のみを追加した場合に順位が 6 位へと改善された正解事例である。本スニペットで抽出された「国保担当窓口にも届出」のように名詞部分が複合語である行動表現は IDF が大きくなりやすく、本スニペットの順位を改善する原因となった。

質問	事故にあったら
スニペット*9	事故対応マニュアル (印刷バージョン) PDF ファイルはこちらをクリック。事故車を安全な場所にうつす。まず車を安全な場所に移し、事故が広がらないようにしてください。けが人を救護する。911 で救急車を呼ぶ。警察に連絡する。(所管警察、または 911) ...

本スニペットは手法 A~C において順位がそれぞれ 33 位、21 位、15 位であり、手法 C に ② のみを追加した場合

*8 <http://www.city.date.hokkaido.jp/shimin/hokeniryo/n96bln0000008gd6.html> (2009 年 6 月 2 日参照)

*9 <http://www.ja-syunan.or.jp/kyosai/jiko/default.htm> (2009 年 6 月 2 日参照)

表 6 質問ごとの手法 C との比較において、2 種類の提案手法を組み合わせた結果と該当する質問数の内訳

Table 6 Details of numbers of questions based on comparison with method C and each couple of proposed methods.

組合せる提案手法	①と②	①と③	②と③
どちらの提案手法も単体で AP 向上, 組み合わせると AP 向上	4	9	7
どちらの提案手法も単体で AP 向上, 組み合わせると AP 低下	2	1	1
片方の提案手法のみ単体で AP 向上, 組み合わせると AP 向上	12	5	5
片方の提案手法のみ単体で AP 向上, 組み合わせると AP 低下	6	9	3
どちらの提案手法も単体で AP 低下, 組み合わせると AP 向上	2	0	9
どちらの提案手法も単体で AP 低下, 組み合わせると AP 低下	4	6	5

に順位が 5 位へと改善された正解事例である。行動表現に単語や複合語を組み合わせて使用することで、「所管警察」や「911」といった関連語をスニベットのスコア計算に利用でき、行動表現のみを使用する場合よりも他のスニベットに比べて高順位となった。

また、単語と複合語のみを使用した場合と、行動表現に単語と複合語も組み合わせて使用した場合を比較するため、単語と複合語のみを用いた (2) によりスコアを計算する手法の MAP を求めたところ、0.633 であった。これは、表 5 において ② のみ適用した場合の MAP よりも低く、行動表現に単語と複合語を組み合わせる有効性が確認された。

質問	事故にあったら
スニベット*10	事故にはあいたくないものですが、でも、もしも事故にあってしまったら？ あわてずに次のように対応しましょう。 1. ただちに車を停止させ、人や物の被害状況を確認する。車を止めずに事故現場から離れた場合、「当て逃げ」「ひき逃げ」とみなされる場合があります。 ...

本スニベットは手法 A~C において順位がそれぞれ 4 位、26 位、31 位であり、手法 C に ③ のみを追加した場合に順位が 2 位へと改善された正解事例である。本スニベットは、波線が引かれた箇所質問に含まれる索引語を含んでいるため、PRF によって適合文書と判定された。結果として、PRF を応用することで、質問との関連性をスニベットのスコア計算に考慮することができ、本スニベットの順位が改善された。

4.2.2 提案手法を複数組み合わせた場合の分析

通常、改善手法を複数適用する場合に見られる傾向として、個々の手法を導入するたびに徐々に性能が向上する。しかし、本提案手法はその傾向に反し、表 5 においては 1 種類または 2 種類のみ適用すると手法 C よりも MAP が下がる場合があった。これは、本提案手法には複数を組み合わせることにより、単体で適用する場合には得られない特

別な改善効果があることを示唆している。

組合せによる改善効果について詳細を分析するため、本提案手法を単体で適用した場合と、任意の 2 種類を組み合わせで適用した場合に、手法 C に比べて各質問の AP がどのように変化するか表 6 にまとめた。たとえば、①と②をどちらを単体で適用しても AP の向上につながり、①+②として組み合わせで適用しても AP の向上につながった質問は 4 件あった。表 6 を通して、各組合せにおける特徴の違いをそれぞれ分析した。

表 6 より、①と②の組合せにおける特徴として、単体で適用した場合は片方の手法が AP の向上につながり、①+②と組み合わせで適用した場合は AP が向上した質問が最も多く 12 件であった。これは①または②のどちらか片方が AP を低下させるという問題点を、もう片方の手法が解消する効果があることを示している。①の問題点として、述語項関係が存在しないスニベットにスコアを付けられない。一方、②の問題点として、述語項関係より単語や複合語の方がスニベットに含まれる件数が多いため、スニベットのスコア計算に対する影響が述語項関係より単語や複合語の方が強く、適切にスコア計算できない可能性がある。しかし、①と②を組み合わせることで、単語や複合語をスニベットのスコア計算に利用しつつ、述語項関係のスコアをスニベットのスコア計算に適切な影響を与えられる大きさに調整できたと推測される。

質問	やけどをしたら
スニベット*11	やけどをした場合の簡単な応急処置、対処方法。 ... 危険な病気ほど潜伏期間が長く病気の進行に気がつきにくいので日頃からの注意がとても重要です。 病気・病名検索ナビ やけどをしたら。 頭部の病気。 脳・脊髄・神経の病気。 目の病気 ...

本スニベットは誤答であり、手法 C において順位が 75 位であった。手法 C に ①、② を単体で適用した場合の順

*10 <http://www.yms-s.jp/hoken/accident/index.html> (2009 年 6 月 2 日参照)

*11 <http://www.airrains.com/illness/first-aid/word1.html> (2009 年 6 月 2 日参照)

位は、それぞれ 75 位, 60 位であり, MAP の向上に貢献しなかった。しかし, ①と②を組み合わせで適用した場合, 順位は 82 位に低下し, MAP の向上に貢献した。本スニペットの特徴として, スコアが 0 を超える述語項関係が含まれていない。そのため, 手法 C および①のみを適用した場合はスニペットのスコアが 0 点であり, 順位に差異がなかった。②のみ適用した場合, 単語と複合語をスニペットのスコア計算に利用できるため, その分スニペットの点数が上がり, 手法 C より順位が向上した。

質問	やけどをしたら
スニペット*12	やけどをしたら, すぐに冷やそう! 水道の水を出しっぱなしにして, すぐに患部を冷やしましょう。(器にためた水ではダメです! 流し続けている清潔な水が必要になります。) 痛みがなくなるまで冷やす。水ぶくれをつぶさないように冷やす。...

本スニペットは正解であり, 手法 C において順位が 29 位であった。手法 C に①, ②を単体で適用した場合の順位は, それぞれ 29 位, 67 位であり, MAP の向上に貢献しなかった。しかし, ①と②を組み合わせで適用した場合, 順位は 28 位に向上し, MAP の向上に貢献した。本スニペットは, 0 を超えるスコアを持つ唯一の述語項関係として下線で示す「患部を冷やす」が得られた。②を単体で適用した場合には本ページの 2 件目に示した誤答のスニペットの方が高順位であったが, ①を適用し述語項関係のスコアを IDF で倍増させたことで順位が逆転した。すなわち, ②の問題点である, 述語項関係より単語や複合語の方がスニペットのスコア計算に与える影響が大きい点に対して, 述語項関係のスコアを IDF で倍増させることで解消されたことをこれらの事例は示している。

表 6 より, ①と③の組合せによる特徴として, 単体で適用した場合はどちらも AP の向上につながり, ①+③と組み合わせで適用した場合にも AP が向上した質問が最も多く 9 件であった。同時に, 単体で適用した場合は片方の手法が AP の向上につながったものの, ①+③と組み合わせで適用した場合は AP が低下した質問も 9 件あった。これは, ①と③には組合せによる改善効果がないことを示している。①と③はどちらも述語項関係が存在しないスニペットにスコアを付けられず, この点は①と③を組み合わせでも解消されない。また, PRF は索引語が少ない文書に対して適切な索引語が得られにくいいため, その効果が薄い。述語項関係のみを索引語とした場合, 全スニペットの平均として 1 件のスニペットに含まれる述語項関係の数は 0.82 件であるため, PRF が有効に機能しない可能性が

*12 <http://park5.wakwak.com/~kosodate/chie/mame/mame4.html> (2009 年 6 月 2 日参照)

高い。

質問	事故にあったら
スニペット*13	介護保険の基礎. 介護苦情事例集. 交通事故にあったら ... 交通事故にあったら. 国民健康保険(国保)を使って治療が受けられます。... 交通事故にあったら. 突然のことですので, 慌てずに落ち着いて適切な対応をとることが大切です。 ...

本スニペットは正解であり, 手法 C において順位が 18 位であった。手法 C に①, ③を単体で適用した場合の順位は, それぞれ 18 位, 65 位であり, MAP の向上に貢献しなかった。また, ①と③を組み合わせで適用した場合, 順位は 65 位であり, MAP の向上に貢献しなかった。本スニペットは下線で示す「治療を受ける」のみ 0 を超えるスコアが付き, 本スニペットの索引語として使用された。また, 本スニペットは PRF において不適合文書と見なされた。質問に含まれる述語項関係が「事故にあう」であるのに対して, スニペット中は名詞部分が複合語となり「交通事故にあう」という述語項関係として抽出されたためである。

表 6 より, ②と③の組合せによる特徴として, 単体で適用した場合はどちらも AP が低下したにもかかわらず, ②+③と組み合わせで適用した場合に AP が向上した質問が最も多く 9 件であった。これは②および③がそれぞれ持つ問題点を, 組み合わせることでお互いに解消し合う効果があることを示している。

質問	火事になったら
スニペット*14	HOME 救急と安全 消防・救急 もし火事になったら. 消防・救急. 普段の備え. もし火事になったら. 休日・夜間の. 救急医療情報. 防災. 災害対策について ... もし火事になったら. 消防車を呼ぶときは 119 番. もし火災が起きたら...

本スニペットは正解であり, 手法 C において順位が 59 位であった。手法 C に②, ③を単体で適用した場合の順位は, それぞれ 63 位, 59 位であった。②と③を組み合わせで適用した場合, 順位は 8 位に向上した。

*13 http://www.kokuho-kumamoto.or.jp/contents_dbpac/detail/style/detail.asp?T.GRP=2&KEY=35&PAGE_ID=7 (2009 年 6 月 2 日参照)

*14 <http://www.town.kawagoe.mie.jp/home/safety/kyukyuu/index2.html> (2009 年 6 月 2 日参照)

質問	火事になったら
スニペット*15	もし、みんなが家や学校が火事になったらどうしますか？ 火事になってもあわてずに落ち着いて行動しましょう。 それでは順番に説明していきます。 大声で「火事だー」とさけんで周りの人たちに知らせる。 近くの電話から 119 番にかけて消防署に知らせる。 ...

本スニペットは正解であり、手法 C において順位が 31 位であった。手法 C に②、③を単体で適用した場合の順位は、それぞれ 56 位、47 位であった。②と③を組み合わせて適用した場合、順位は 2 位に向上した。

上記の 2 つのスニペットには、「火事」や「119 番」という共通の単語が含まれている。式 (8) より、適合文書と見なされたスニペットに類出する索引語ほどスコアが高くなるため、これらの単語は高いスコアに修正された。結果として、②と③の組合せにより、これらのスニペットは順位が向上した。

以上より、①と②、②と③の組合せにおいてそれぞれ組合せによる特別な改善効果があった。全部を組み合わせた場合、①と②、②と③の改善効果が合わせて働き、最も高い MAP につながった。

4.3 誤り分析

表 3 に、質問ごとの AP を示す。各質問において最も高い AP を太字で示す。手法 D と手法 A を比較した場合、手法 D より AP が高い質問が 6 件あった。同様に、手法 B に対しては 10 件、手法 C に対しては 6 件あった。これらの質問を対象に、手法 D の AP が下がってしまった原因について考察する。表 3 において手法 D を手法 A~C のそれぞれと比較した結果、手法 A~C の方が高い AP となってしまった質問におけるスニペットの内訳を表 7、表 8、表 9 に示す。表 7 において、各欄の数字を合計すると 561 件である。残り 39 件のスニペットは抽出元文書が正しく取得できなかったために評価対象から外されたスニペットである。同様に、表 8 と表 9 ではそれぞれ 48 件と 21 件のスニペットが評価対象から外された。

表 7~表 9 において、手法 A~C の方が高順位となった正解のスニペットは、手法 D の MAP を低下させた。ゆえに、これらのスニペットにおいて、手法 A~C の方が高順位となった原因を分析した。分析によって得られた原因と、手法 A~C において該当するスニペット数を表 10 に示す。

以下、表 10 の原因ごとに例を交えて説明する。各スニペットの例において、下線が引かれた箇所は回答誤りの原

表 7 手法 D より手法 A の AP が高い質問に関するスニペット数の内訳

Table 7 Details of numbers of snippets related to the questions for which AP of method A was greater than that of method D.

	手法 D の方が高順位	手法 A の方が高順位	順位差なし
正解のスニペット	134	119	1
誤答のスニペット	168	131	8

表 8 手法 D より手法 B の AP が高い質問に関するスニペット数の内訳

Table 8 Details of numbers of snippets related to the questions for which AP of method B was greater than that of method D.

	手法 D の方が高順位	手法 B の方が高順位	順位差なし
正解のスニペット	197	333	41
誤答のスニペット	146	176	59

表 9 手法 D より手法 C の AP が高い質問に関するスニペット数の内訳

Table 9 Details of numbers of snippets related to the questions for which AP of method C was greater than that of method D.

	手法 D の方が高順位	手法 C の方が高順位	順位差なし
正解のスニペット	124	149	10
誤答のスニペット	161	130	5

因であり、詳細は例の直後で説明する。また、太字の箇所は正解の記述である。

原因ア

質問	迷惑メールが届いたら
スニペット*16	迷惑メール報告完了のメッセージが表示されます。... <u>迷惑メールフィルターが設定されているかどうかを確認する</u> 。1. ... しかし、不審なメールが届いたら、開かないことが何よりも安全です。「迷惑メール報告機能」を使って ...

「メッセージが表示」、「迷惑メールフィルターが設定」、「メールが届く」という 3 つの述語項関係が含まれる。これらはすべて 3.4.1 項の基準 (e) により、ガ格の後置詞句をとまなうため、スコアが 0 の行動表現として扱われた。その結果、本スニペットの順位が下がった。基準 (e) は、動作主が明示されている場合に、その述語項関係は手続きを表さないため排除する基準である。しかし、行動表現を構成する動詞が受動態の用法をとまなう場合にはガ格がともなう場合であっても適切な述語項関係となる場合が見られた。

*15 <http://www.nagasaki-kenoukumiai.jp/syoubou/kitu/119suruto/119suruto.html> (2009 年 6 月 2 日参照)

*16 <http://antispam.yahoo.co.jp/tools.html> (2009 年 6 月 2 日参照)

表 10 手法 D が正解のスニペットを低く順位付けた誤りの原因とその該当件数

Table 10 Reasons of errors for low-ranking correct answers in method D, and the numbers of errors.

	誤りの原因	手法		
		A	B	C
ア	「ガ格の後置詞句をとともう行動表現のスコアは 0 にする」により、質問に対して適切な行動表現のスコアが 0 となった	9	16	8
イ	「質問に含まれる名詞をとともう行動表現のスコアは 0 にする」により、質問に対して適切な行動表現のスコアが 0 となった	9	10	10
ウ	不要表現辞書の作成元となった 70 件の How 型質問と行動表現の集合が原因で、質問に対する適切な行動表現が不要表現辞書に登録されていたために、スコア計算に利用されなかった	56	168	71
エ	適切な行動表現の構成要素に使われない種類の名詞や動詞（代名詞や非自立的な名詞や動詞）を持つと判断され、質問に対して適切であるにもかかわらず抽出されない行動表現があった	58	140	61
オ	形態素解析および係り受け解析の失敗により、質問に対して適切であるにもかかわらず抽出されない行動表現があった	41	103	54
カ	質問に対して適切な行動を表す単語や複合語があった	25	109	24
キ	原因不明	7	13	13

原因イ

質問	迷惑メールが届いたら
スニペット*17	迷惑メールの対策. 対策 1. <u>迷惑メールに返信したり、記載されたウェブページにアクセスするのは避け、迷惑メールが届いたら、削除するようにしましょう。...</u> 面倒な削除作業や、メールソフトへの設定作業は必要もなく、簡単に <u>迷惑メールを防ぐ</u> ことができます。...

「迷惑メールの対策」, 「迷惑メールに返信」, 「迷惑メールを防ぐ」と質問に含まれる名詞を含む述語項関係を複数含んでいる。基準 (f) を適用した結果、これらはすべてスコアが 0 の行動表現として扱われた。その結果、本スニペットの順位が下がった。

原因アとイについては、スコアが 0 となった行動表現に対して、ある程度低いスコアが付くよう手法を修正することで、当該スニペットの順位が下がらないようにする必要がある。

*17 <http://www.starcat.ne.jp/mediacat/customer/security/meiwaku.html> (2009 年 6 月 2 日参照)

原因ウ

質問	迷惑メールが届いたら
スニペット*18	最近迷惑メールが多いんです。... <u>しかしながら迷惑メールが減るわけも無く、なんとかして頂きたいものです。...</u> 迷惑メールが届いたら <u>警察の専用アドレスに転送。自動で解析・・・プロバイダ等にも協力して貰って検挙。</u> 715 ...

「迷惑メールが減る」や「専用アドレスに転送」は、不要表現辞書の作成元となった 70 件の How 型質問と行動表現の集合が原因で、「減る」や「転送」が行動表現の構成要素として不適切な動詞と判断され、不要表現辞書に登録されていた。その結果、スコアが付くべき行動表現を抽出できず、本スニペットの順位が下がった。

原因エ

質問	やけどをしたら
スニペット*19	... 受ける必要があります。 <u>じゃがいもをすり下ろし、その汁をガーゼに染み込ませ、患部に湿布します。1日4~5回は最低取り替えましょう。なお、やけどをしたらまずは冷やすことです。流水や氷水に患部をつけて、十分に冷やしましょう。...</u>

3.3 節より、「やけどする」のように名詞を動詞化させる「する」という動詞が使われている場合は行動表現として抽出しない。しかし、本スニペットでは調理動作を表す「する」が誤って条件に一致してしまい、「じゃがいもをする」という行動表現が抽出されなかった。また、「つける」という動詞は非自立的な動詞として形態素解析の辞書に登録されており、非自立的な動詞が使われている場合も行動表現として抽出しない条件としていたため、「患部をつける」という行動表現が抽出されなかった。

原因オ

質問	やけどをしたら
スニペット*20	3歳の息子がカップラーメンをこぼして太ももにかかり、やけどをしてしまいました。このような場合の応急処置はどうすればよいのでしょうか。 <u>やけどをしたらすぐに受傷部位を水(なるべく水道などの流水)で冷やします。患部...</u>

*18 <http://www.konna.jp/title/000472.htm> (2009 年 6 月 2 日参照)

*19 <http://www.ne.jp/asahi/tattyan/kenkoukouza/Site/Byoki/Byo-Yakedo.htm> (2009 年 6 月 2 日参照)

*20 <http://www.skin.or.jp/qa/qa14/index.html> (2009 年 6 月 2 日参照)

本スニベットには「受傷部位を冷やす」と「水で冷やす」という質問に対して適切な行動表現が含まれる。しかし、係り受け解析の結果、名詞と動詞の間に注釈が丸括弧で囲われる形で埋め込まれていたために、名詞と動詞間の係り受け関係を正しく抽出することができず、行動表現として抽出できなかった。

原因ウ～オについては、スニベットから抽出されずスコア計算の対象から外された行動表現に対して、ある程度低いスコアが付くよう手法を修正することで、当該スニベットの順位が下がらないようにする必要がある。特に、原因オについては、名詞と動詞間に係り受け関係が特定されない場合でも、近接する名詞と動詞を行動表現として抽出し、係り受け関係がある行動表現に比べて低いスコアを付ける必要がある。

原因カ

質問	やけどをしたら
スニベット*21	つかはら. 美容外科. 形成外科. クリニック. やけどをしたら. けがをしたら ... やけどをしたら. <u>すぐ冷やす</u> . どのように冷やす. どれくらい冷やす ... やけどをしたら. けがをしたら. おでき, ほくろ. きずは ...

本スニベットには「すぐ冷やす」という質問に対して適切な行動表現が含まれる。これは述語項関係でなく、副詞と動詞の単語列であったため、行動表現として抽出されなかった。

原因カの対策として、「冷やす」のような動詞単体の場合や、「すぐ冷やす」や「早く冷やす」など副詞や形容詞が動詞にともなう場合が述語項関係でない正解の記述として見られたため、これらを行動表現の抽出条件に追加する。また、「30分冷やす」のように名詞と動詞の間に助詞が省略されており述語項関係と判断されない場合もあった。ゆえに、助詞の省略も考慮して行動表現を抽出する必要がある。

5. おわりに

本論文は、述語と項の対を用いて行動を問う質問に回答する How 型 QA 手法の改善を目的とした。当該 QA 手法は、入力された質問をクエリとして Web を検索し、検索された Web 文書から述語項関係を行動表現として抽出し、適切な行動表現や回答候補の文書が上位になるようスコアを付ける回答評価の順に処理を行う。本論文では、最後の回答評価を改善した。具体的には、行動表現に加えて単語や複合語を索引語に組み合わせる点、情報検索の重み付け手法である IDF を行動表現のスコア計算に応用する点、回答候補となる文章のスコア計算に擬似適合性フィードバック

クを応用する点の3つを提案した。評価実験では、既存の手法と比較し、提案手法が有効であることを示した。また、質問ごとに回答誤りの傾向を分析し、誤りへの対処法に関する考察を行った。今後の課題は、考察によって得られた対処法を実現することである。また、基準 (f) において項だけでなく述語を考慮するなど、行動表現のスコア計算基準を見直す必要がある。

謝辞 本研究の一部は、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 21013003) ならびに科研費基盤研究 (B) (課題番号 22300050) によって実施された。

参考文献

- [1] 旭 直人, 山本岳洋, 中村聡史ほか: 行動連鎖を用いた情報検索支援と Web からの行動連鎖の抽出, 第 1 回データ工学と情報マネジメントに関するフォーラム 発表論文集 (2009).
- [2] 麻野間直樹, 古瀬 蔵, 片岡良治: How-to 型質問応答の実現に向けた質問回答文書の特徴分析, 電子情報通信学会技術研究報告, NLC2005-9, pp.55-60 (2005).
- [3] Fukumoto, J.: Question Answering System for Non-factoid Type Questions and Automatic Evaluation based on BE Method, *Proc. 6th NTCIR Workshop Meeting*, pp.441-447 (2007).
- [4] Harada, M., Kato, Y., Takehara, K., et al.: QA System Metis Based on Semantic Graph Matching, *Proc. 6th NTCIR Workshop Meeting*, pp.448-459 (2007).
- [5] 平松文嗣, 松下大輔, 平澤茂一: 初期検索結果から抽出した単語を用いた擬似フィードバック手法, FIT2005 発表論文集, pp.101-102 (2005).
- [6] 松井くにお, 田中穂積: 初期質問文から蓄積された質問応答への効果的マッチング法, 自然言語処理, Vol.10, No.5, pp.121-138 (2003).
- [7] 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).
- [8] Mori, T., Okubo, T. and Ishioroshi, M.: A QA system that can answer any class of Japanese non-factoid questions and its application to CCLQA EN-JA task, *Proc. 7th NTCIR Workshop Meeting*, pp.41-48 (2008).
- [9] 小澤俊介, 内元清貴, 松原茂樹: モノの用途表現を手がかりとした Web からのノウハウの獲得, 情報処理学会研究報告, Vol.2010-NL-196, No.6, pp.1-7 (2010).
- [10] Takechi, M., Tokunaga, T., Matsumoto, Y., et al.: Feature Selection in Categorizing Procedural Expressions, *Proc. 6th International Workshop on Information Retrieval with Asian Languages*, pp.49-56 (2003).
- [11] 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).
- [12] 山本正範, 金盛克俊, 福田雅志ほか: 行動表現の連鎖に基づく方法説明領域の抽出, 2007 年電子情報通信学会総合大会 発表論文集, p.53 (2007).
- [13] Lin, J., Fernandes, A., Katz, B., et al.: Extracting Answers from the Web Using Knowledge Annotation and Knowledge Mining Techniques, *Proc. TREC-11* (2001).
- [14] Manuel, J., Soriano, G., Buscaldi, D., et al.: QUASAR: The Question Answering System of the Universidad Politecnica de Valencia, *Proc. CLEF 2005*, LNCS 4022, pp.439-448 (2006).

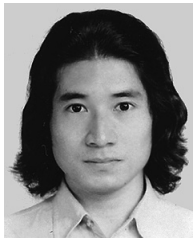
*21 <http://www.oct.zaq.ne.jp/tsukahara/BURN.HTML> (2009 年 6 月 2 日参照)

- [15] Sanderson, M.: Test Collection Based Evaluation of Information Retrieval Systems, *Foundations and Trends in Information Retrieval*, Vol.4, No.4, pp.247–375 (2010).
- [16] Yin, L.: A Two-Stage Approach to Retrieving Answers for How-To Questions, *Proc. 11th Conference of the European Chapter of the Association for Computational Linguistics*, pp.63–70 (2006).



佐々木 智

2007年3月山梨大学工学部循環システム工学科卒業。2009年3月筑波大学大学院図書館情報メディア研究科博士前期課程修了。同年同大学院博士後期課程に進学し現在まで在学中。



藤井 敦 (正会員)

1993年3月東京工業大学工学部情報工学科卒業。1998年3月同大学大学院博士課程修了。現在、東京工業大学大学院情報理工学研究科准教授。博士(工学)。自然言語処理、情報検索、音声言語処理、Webマイニングの研究に従事。電子情報通信学会、人工知能学会、言語処理学会、日本データベース学会各会員。