

# Q&A サイトの回答から取り出した質問に書くべき情報の特殊さの推定

谷口 祐亮 小島 正裕 西村 涼 渡辺 靖彦 岡田 至弘

龍谷大学 理工学部 情報メディア学科

{t070437,t10m101}@mail.ryukoku.ac.jp, r\_nishimura@afc.ryukoku.ac.jp,  
{watanabe,okada}@rins.ryukoku.ac.jp

## 1 はじめに

質問を投稿しておく他ユーザが答えてくれる、Yahoo! 知恵袋<sup>\*1</sup>、教えて! goo<sup>\*2</sup>、人力検索はてな<sup>\*3</sup>、OK-Wave<sup>\*4</sup>などのコミュニティベース質問応答サービス (以下 Q&A サイト) が、さかんに利用されている。大量の質問が投稿されている Q&A サイトでよい回答を得るためには、回答するのに十分な情報が質問に書かれていることが重要である。例えば、以下の (質問 1) は、回答するのに十分な情報が書かれていない質問の例である。

(質問 1) PC が青い画面になって起動しません。どうしたらいいでしょうか？

(回答 1) OS は何を使っていますか？ OS が分からないと答えられません。

もし、質問に書くべき情報、すなわち、OS が何であるかという情報が重要だという情報が与えられていれば、この質問を投稿したユーザはその情報を質問に記述していただく。そこでわれわれは、ユーザが質問を作成するのを支援するため、質問に書くべき情報を Q&A サイトに投稿された回答から取り出した [1] [2]。しかし、取り出した質問に書くべき情報はさまざまで数多くあり、どんな順序や形式でユーザに示すのが問題となった。

これまで、Web 検索における情報の順位づけでは、PageRank [3] [4] や、マルコフ決定過程によるランキング手法 [5]、blog 情報を用いたランキング手法 [6]、blog 内のキーワードの出現頻度を用いたブログランキング手法 [7] などが行なわれている。また、効率よく言葉の意味を理解するための学習順序についても研究が行なわれている [8]。しかし、質問作成に役立つ情報の順序づけについてはまだあまり研究されていない。そこでわれわれは、Q&A サイトに投稿された回答から取り出した質問に書くべき情報を

特定の質問でだけ書くべき特殊な情報  
さまざまな質問で書くべき情報

に区別できれば、ユーザにとって利用しやすい順序や形式で情報を提供できると考えた。例えば、以下の (質問 2) は (質問 1) とは異なる内容について質問しているが、この質問でも OS が何かという情報が書かれていない。

(質問 2) スクリーンセーバーはどこのフォルダにあるのでしょうか？

(回答 2) WindowsXP でしたら以下の場所にあります。  
C:\WINDOWS\system32

このように、OS が何かという情報はさまざまな質問で書くべき情報である。一方、以下の (質問 3) で不足している書くべき情報は特殊な情報である。

(質問 3) 128MB のフラッシュメモリには、何枚の写真が入りますか？ デジカメは CANON を使っています。

(回答 3) デジカメの画質モードはどうなっていますか？

(質問 3) で不足している情報は、画質モードがどうなっているかという情報である。この情報は、デジカメによる撮影についての質問でだけ求められる特殊な情報である。

## 2 回答するために十分な情報が書かれていない質問と質問に書くべき情報

本研究では、Q&A サイトの例として Yahoo! 知恵袋を取り上げる。Yahoo! 知恵袋は、登録されたユーザが質問と回答のメッセージをやりとりすることができる Q&A サイトである。この Yahoo! 知恵袋に 2004 年 4 月から 2005 年 10 月までに投稿された質問約 311 万件、回答約 1347 万件のメッセージが国立情報学研究所から公開されている<sup>\*5</sup>。この Yahoo! 知恵袋のデータの「パソコン、周辺機器」のカテゴリに投稿された質問と回答を用いて調査を行なった。以下に、回答で質問に書くべき情報を指摘されている質問とその回答の例を示す。

(質問 4) もともと中に入っていたウイルス対策ソフトをインストールしたのですが、他のソフトに変えるためにアンインストールをしたいです。

どのような手順で行えばよいのか教えてください。

(回答 4) ソフト名を書いていたかたかったですね。

(質問 5) ついに新しいパソコンを買いましたがウィンドウズ 98 だったので XP の画面や新しい機能に慣れず戸惑っています。

98 では会計ソフトを使っていたのですが、それを新しい PC に取り込む方法が分かりません。

(回答 5) ソフト名をお書き下さい。

(質問 4) では第 1 文が質問の中心となる文であり、(質問 5) では第 2 文が質問の中心となる文である。本研究では、

<sup>\*1</sup> <http://chiebukuro.yahoo.co.jp/>

<sup>\*2</sup> <http://oshiete.goo.ne.jp/>

<sup>\*3</sup> <http://q.hatena.ne.jp/>

<sup>\*4</sup> <http://okwave.jp/>

<sup>\*5</sup> <http://research.nii.ac.jp/tdc/chiebukuro.html>

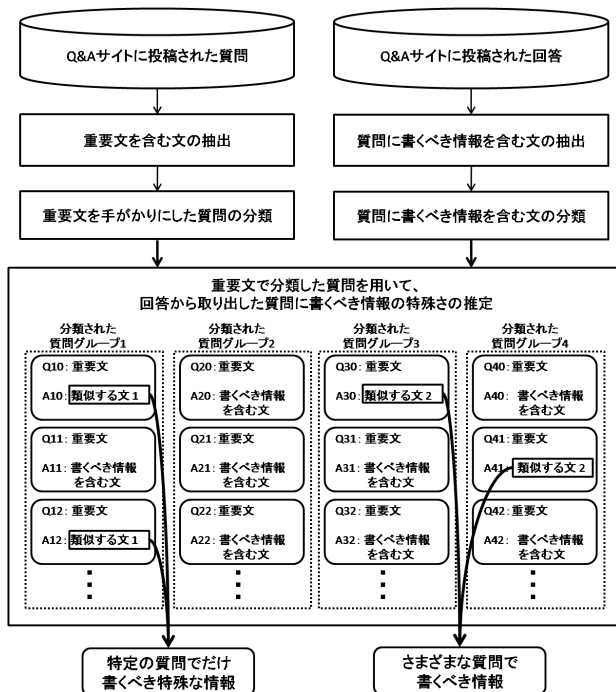


図 1 質問に書くべき情報の特殊さの推定手法の概要

質問の中心となる文を重要文とする。一方、回答の(回答 4)の「ソフト名を書いていただきかったですね。」と(回答 5)の「ソフト名をお書き下さい。」は、いずれも質問に書くべき情報である。このソフト名についての質問に書くべき情報が

特定の質問でだけ書くべき特殊な情報  
さまざまな質問で書くべき情報

のいずれかであるかは、その情報がさまざまな質問に対する回答で指摘されているか、すなわち、それらの情報を含む回答が内容の異なる質問に投稿されたかどうかで判定することができる。内容の異なる質問であるかは質問の重要文を用いて判定する。これは、質問の重要文が、質問内容の要約であることが多いからである。具体的にはまず、内容の異なる重要文を含む質問を内容の異なる質問であると判定し、それらの質問に類似する内容の質問に書くべき情報を含む回答が投稿されていれば、その情報はさまざまな質問で書くべき情報であると考えられる。そこで、

重要文を手がかりにした質問の分類  
質問に書くべき情報を含む文を手がかりにした回答の分類

を行えば、質問に書くべき情報の特殊さを推定することができる考えた。

### 3 Q&A サイトに投稿する質問に書くべき情報の特殊さの推定手法

Q&A サイトに投稿された回答から、質問に書くべき情報を含む文を取り出し、その情報が

定型表現	定型表現が使われている重要文の例
方法がありますか	電池を抜かないでもマウスの電源をオフにする方法がありますか？
方法はないですか	パソコンの電源を外すしか、長持ちさせる方法はないですか？
方法がわかりません	光ネットを PC 起動時に接続したいのですが、方法がわかりません。
方がいいですか	パソコンはパソコン専用の台にのせた方がいいですか？
どうしたらいいですか	デスクトップにあるアイコンを削除するにはどうしたらいいですか？
どうすればいいですか	パソコンの右下の音量調節が無くなって、音が出なくなりましたが、どうすればいいですか？
事はできますか	OutlookExpress2003 のデータ OutlookExpress に移す事はできますか？
困っています	パソコンの画面表示が大きくなって困っています。

図 2 質問の重要文から取り除く定型表現の例

特定の質問でだけ書くべき特殊な情報  
さまざまな質問で書くべき情報

のどちらであるかを推定する方法を以下に述べる。また、推定する方法の概要を図 1 に示す。

- step 1 [形態素解析および係り受け解析]  
Yahoo! 知恵袋データの質問と回答を対象に、形態素解析および係り受け解析を行う。形態素解析には JUMAN [9]、係り受け解析には KNP [10] を用いた。
- step 2 [重要文と質問に書くべき情報を含む文の抽出]  
質問からは重要文、回答からは質問に書くべき情報を含む文を機貝らの方法によって取り出す [1] [2]。質問に書くべき情報を含む文を取り出せた回答とその質問だけを以後の処理 (step 3 ~ 5) で用いる。
- step 3 [質問の重要文を手がかりにした質問とその回答の分類]  
step 2 で質問から取り出した重要文を手がかりにして、質問とその回答を以下の手順で質問グループに分類する。まず、重要文の文末に定型表現がある場合は、それらを取り除く。図 2 に取り除く定型表現とその例を示す。次に、重要文の文末の用言によって質問とその回答を分類する。したがって、文末の用言が同じ重要文を含む質問とその回答は同じ質問グループに分類される。
- step 4 [質問に書くべき情報を含む文を手がかりにした質問とその回答の分類]  
step 2 で回答から取り出した質問に書くべき情報を含む文を手がかりにして、質問とその回答を回答グループに分類する。同じ回答グループに分類されるのは、回答から取り出した質問に書くべき情報を含む文について、文末の用言とその直前格が一致して

表 1 質問に書くべき情報の特殊さの推定結果, ( ) 内は情報の種類の数

質問に書くべき情報を含む文の数	さまざまな質問に書くべき情報を含む文の数	特定の質問でだけ書くべき特殊な情報を含む文の数
120460 (83702)	1719 (297)	118741 (83405)

いて、それら以外の文節がその文にある場合はさらに、以下のいずれかの条件が満されるものである。

- － 文末の用言とその直前格以外の文節が 3 個以上ある場合、それらのうち 2 個以上が一致する
- － それ以外の場合、それらのうち 1 個以上が一致する

#### step 5 [質問に書くべき情報の特殊さの推定]

step 4 で 1 つの回答グループに分類された回答が、step 3 の複数の質問グループに属していれば、その回答グループに分類された回答から取り出した質問に書くべき情報はさまざまな質問で書くべき情報であると推定する。そうでなければ、特定の質問でだけ書くべき特殊な情報であると推定する。

## 4 実験結果と検討

提案手法の有効性を評価するために、Yahoo! 知恵袋の「パソコン、周辺機器」カテゴリを対象に、回答から取り出した質問に書くべき情報が以下のどちらであるのか推定する実験を行った。

特定の質問でだけ書くべき特殊な情報  
さまざまな質問で書くべき情報

実験データには、Yahoo! 知恵袋データの「パソコン、周辺機器」カテゴリの質問と回答から抽出した、質問の重要文と回答の質問に書くべき情報を含む文を用いた。Yahoo! 知恵袋データの「パソコン、周辺機器」カテゴリに投稿された質問は 171848 件であり、回答は 474687 件である。これらの回答から質問に書くべき情報を含む文を 120460 文取り出した [1] [2]。3 節で述べた方法により特殊さの推定した結果を表 1 に示す。さまざまな質問で書くべき情報については 4.1 節で述べ、特定の質問でだけ書くべき特殊な情報については 4.2 節で述べる。

### 4.1 さまざまな質問で書くべき情報

さまざまな質問で書くべき情報であると 3 節で述べた方法で推定されたものは 297 種類あり、それらは 1719 文の質問で書くべき情報を含む文から得られた。以下に例を示す。

(質問 6) エクセルのシートを追加したくて、[挿入] - [ワークシート] を選びましたが、'ワークシート.xls' が見つかりません。  
仕方なく別のシートをコピーして内容を削除の方法

で新しいシートを作りましたが、なぜ上の方法でできないのでしょうか？

(回答 6) エクセルの何かが無くなっているのでは？  
再インストールすればよいのでは？

(質問 7) PC を使っていると突然電源がおちてしまいます。

修理したほうがいいのでしょうか？

(回答 7) 再インストールしました？  
あと熱暴走も考えられるので PC の蓋を開けて埃が溜まってないか確かめましょう。溜まってたら掃除は慎重にね。

重要文として、(質問 6) からは第 1 文 (「エクセルのシートを追加したくて、[挿入] - [ワークシート] を選びましたが、'ワークシート.xls' が見つかりません。」)、(質問 7) からは第 1 文 (「PC を使っていると突然電源がおちてしまいます。」) が取り出された。一方、質問に書くべき情報を含む文として、(回答 6) からは第 2 文 (「再インストールすればよいのでは？」)、(回答 7) からは第 1 文 (「再インストールしました？」) が取り出された。(質問 6) と (質問 7) から取り出した重要文によって、(質問 6) と (回答 6) は (質問 7) と (回答 7) とは異なる質問グループに分類されている。一方、(回答 6) と (回答 7) から取り出した質問に書くべき情報を含む文によって、これら 2 組の質問とその回答は同じ回答グループに分類されている。このため、(回答 6) と (回答 7) から取り出した質問に書くべき情報はさまざまな質問で書くべき情報であると推定できる。

また、以下の質問に書くべき情報は、複数の質問グループに分類された回答で見つかり、さまざまな質問で書くべき情報であると推定された。しかし、それらの回答はすべて 1 人の回答者によって投稿されていた。

PC の質問をされるなら、使用中のパソコンの OS のバージョン、PC のスペックを書いてください。

この質問に書くべき情報では、パソコンの OS について指摘している。これは 1 節の (回答 1) と (回答 2) でも指摘しているように、さまざまな質問に書くべき情報であると推定できる。ただし、同じ回答者による指摘の場合は、その回答者のみが重要と考える情報なのか、それとも、さまざまな回答者が重要と考える情報なのか明らかではない。このため、同じ回答者のみが指摘している情報は、本当にさまざまな質問で書くべき情報かどうか検討する必要があると考えられる。

### 4.2 特定の質問でだけ書くべき特殊な情報

特定の質問でだけ書くべき特殊な情報であると 3 節で述べた方法で推定されたものは 83405 種類あり、それらは 118741 文の質問で書くべき情報を含む文から得られた。以下にその例を示す。

(質問 8) 大事なメールが来るのでパソコンにきたメールを携帯に転送したいのですが、送信が出来ません。  
ソフトは Outlook Express6 です。

(回答 8) プロバイダはどこですか？

プロバイダによって、転送できるものとできないものがあります。また転送の設定も違います。

(質問 9) outlook express でエラーが発生し、その後の送信が出来なくなりました。どうすればいいのでしょうか？

(回答 9) プロバイダはどこですか？

その電話サポートに連絡して、アドバイスを受けることが一番早いと思いますよ。

重要文として、(質問 8) からは第 1 文(「大事なメールが来るのでパソコンにきたメールを携帯に転送したいのですが、送信が出来ません。」)、(質問 9) からは第 1 文(「outlook express でエラーが発生し、その後の送信が出来なくなりました。」)が取り出された。一方、質問に書くべき情報として、(回答 8) と (回答 9) それぞれから第 1 文(「プロバイダはどこですか？」)が取り出された。(質問 8) と (質問 9) から取り出した重要文は類似している。同様に、(回答 8) と (回答 9) から取り出した質問に書くべき情報を含む文も類似している。実際、これらの質問と回答は 3 節で述べた方法で同じ質問グループと回答グループに分類されている。このため、(回答 8) と (回答 9) で表現されている質問に書くべき情報は特定の質問でだけ書くべき特殊な情報であると推定できる。

一方、質問に書くべき情報に表記ゆれがあるため、さまざまな質問で書くべき情報に推定されなかった例を以下に示す。

(質問 10) メールを送信すると 送信者のところに 送信者 小泉 と漢字で出ます。送信者の名前をローマ字に変えるにはどうしたら良いでしょうか？教えてください。

(回答 10) メーラは何？

OutlookExpress なら以下の操作 ツール> アカウント> 対象のアカウントを選択> プロパティ 名前欄が送信者名として表示されるものです。お好きなように変更してください。

(質問 11) 東芝のノートパソコンを使っています。まだ使い方がよくわからないのですが、電子メールが届いてるかどうかはメールの画面を出さないと確認できないのでしょうか？それとも画面のどこかに届いたらサインが出てるのでしょうか？

(回答 11) お使いのメーラーは何でしょうか？

来たら音が鳴ったり、パソコンの機種によってはランプがついたりするものもありますよ。メーラーを開いて、ツール オプションの中の「メッセージが届いたら音を鳴らす」などにチェックを入れると、音が鳴るようになります。

重要文として、(質問 10) からは第 2 文(「送信者の名前をローマ字に変えるにはどうしたら良いでしょうか？」)、(質問 11) からも第 2 文(「まだ使い方がよくわからないのですが、電子メールが届いてるかどうかはメールの画面を

出さないと確認できないのでしょうか？」)が取り出された。一方、質問に書くべき情報として、(回答 10) からは第 1 文(「メーラは何？」)、(回答 11) からも第 1 文(「お使いのメーラーは何でしょうか？」)が取り出された。(質問 10) と (質問 11) の質問内容は異なっていて、異なる質問グループに分類されていた。一方、質問で書くべき情報として (回答 10) と (回答 11) で表現されているのは、電子メールを送受信するためのアプリケーションはなにかという情報である。しかし、(回答 10) と (回答 11) で「メーラ」と「メーラー」で表記が異なるため、異なる回答グループに分類されていた。その結果、この情報はさまざまな質問で書くべき情報であると推定されなかった。このため、表記ゆれを検出して対応できれば、より精度の高い推定ができると考えられる。

謝辞 本研究の一部は、日本学術振興会科学研究費補助金(課題番号 20500106 と 22650021)の助成を受けて行われたものです。

## 参考文献

- [1] 磯貝, 西村, 渡辺, 岡田: Q&A サイトへの質問の作成を支援するための情報の複数のカテゴリからの抽出, 言語処理学会第 15 回年次大会, P1-8, pp152-155, (2009).
- [2] 磯貝, 西村, 渡辺, 岡田: Q&A サイトへの質問の作成を支援するための情報の抽出, 情報社会学会第 1 回知識共有コミュニティワークショップ, pp.19-28, (2008).
- [3] Brin and Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Network and ISDN Systems, Vol.30, Issues 1-7, pp.107-117, (1998).
- [4] Kamvar, Haveliwala, Manning, and Golub: Extrapolation methods for accelerating PageRank computations, 12th International World Wide Web Conference, pp.261-270, (2003).
- [5] 岡村, 宮内, 土肥: マルコフ決定過程による Web ページランキングアルゴリズム, 電子情報通信学会論文誌, Vol.J89-D, No.2, pp.210-219, (2006).
- [6] 中島, 竹原, 舘村, 日野, 原, 田中: blog 解析に基づく Web 情報検索の信頼性向上技術, 人工知能学会第 6 回セマンティックウェブとオントロジー研究会, SIG-SWO-A401-05, (2004).
- [7] 中島, 稲垣, 草野: ブロッガーの熟知度に基づいたブログランキング方式の提案, 電子情報通信学会第 19 回データ工学ワークショップ (DEWS2008), B7-6, (2008).
- [8] 西原, 砂山, 谷内田: Web ページの難易度と学習順序に基づく情報理解支援システム, 電子情報通信学会論文誌, Vol.J89-D, No.9, pp.1963-1975, (2006).
- [9] 黒橋, 河原: 日本語形態素解析システム JUMAN version 5.1 使用説明書, 京都大学, (2005).
- [10] 黒橋, 河原: 日本語構文解析システム KNP version 2.0 使用説明書, 京都大学, (2005).