

Title	Q&Aサイトにおける社会調査型質問への回答に対する良否判断支援システム
Author(s)	王, 曦虹; 小林, 智也; 小倉, 加奈代; 西本, 一志
Citation	インタラクシオン2012論文集 (情報処理学会シンポジウムシリーズ), 2012(3): 481-486
Issue Date	2012-03-16
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/10639
Rights	<p>社団法人 情報処理学会, 王 曦虹, 小林 智也, 小倉 加奈代, 西本 一志, インタラクシオン2012論文集 (情報処理学会シンポジウムシリーズ), 2012(3), 2012, 481-486. ここに掲載した著作物の利用に関する注意: 本著作物の著作権は(社)情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 Notice for the use of this material: The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright © Information Processing Society of Japan.</p>
Description	

Q&A サイトにおける社会調査型質問への回答に対する 良否判断支援システム

王 曦虹[†] 小林 智也[†] 小倉 加奈代[†] 西本 一志[‡]

近年 Q&A サイトの利用者が増加している。Q&A サイトの質問は情報検索型と社会調査型に大別される。このうち社会調査型質問には客観的な正解はなく、特定の個人あるいは集団に対してアンケート調査を行うことで回答を得る。このため、質問者は回答の信頼性を判断しにくいという問題点がある。本稿では、質問者が回答の信頼性を判断するための補助材料として、回答者の質問・回答履歴から、今回回答しようとしている質問と類似した質問とそれへの回答を抽出し、これを質問者に提示する。これによって、質問者が回答者の態度や価値観を推測可能とする手法を提案する。Yahoo! 知恵袋のデータを用いてシステムの初期的評価を行い、一定の有用性を確認した。

An assistant system for judging quality of responses to social survey questions in a Q&A website

XIHONG WANG[†] TOMOYA KOBAYASHI[†] KANAYO OGURA[†] KAZUSHI NISHIMOTO[‡]

Recently, Q&A sites have been widely used. Questions of the Q&A sites can be classified into two types: an information-seeking type and a social-survey type. Since there are no objectively correct answers for the social-survey type questions, questionnaires are usually used to obtain answers from responders. However, it is difficult for a questioner to judge credibility of the answers. For supporting the judgment, we extract questions similar to the questioner's question from the responders' histories of questions and answers as well as answers to the extracted questions. By providing them to the questioner, it is expected that the questioner becomes able to infer each responder's attitude and sense of values. We conducted pilot studies using Yahoo! Chiebukuro data and confirmed basic efficiency of the proposed method.

1. はじめに

近年、Q&A サイトと呼ばれる質問回答サイトの利用者が増加している。Q&A サイトというのは会員同士が、お互いの質問に答え、疑問を解決するウェブサイトのことである[1]。日本で代表的な Q&A サイトとしては、登録者数第 1 位の Yahoo! 知恵袋や第 2 位の OKWave などが挙げられる。栗山らは、Q&A サイトにおける質問には情報検索型質問と社会調査型質問の 2 種類があることを指摘している[2]。情報検索型質問とは、サーチエンジンや図書館レファレンスサービスを利用して回答を探すことが可能な内容に関する質問である。社会調査型質問とは、客観的な唯一の正解が存在せず、特定の個人あるいは集団に対してアンケート調査を行うことで各回答者の主観に基づく回答を得

るような質問である。

本稿では、社会調査型質問を対象とし、得られた個々の回答の良否を判断する際の助けとなる情報を提供するシステムを提案する。社会調査型質問に対しては、回答者それぞれが大きく異なった内容の回答を寄せる。各回答の正誤を判断するための客観的な基準が存在しないため、質問者は、どの回答を良い回答として受け入れるかを自分の主観に基づき判断せざるを得ない。また、多くの Q&A サイトでは、寄せられた回答を質問者が評価し、ベストアンサーを選んで優れた回答者を報賞することが求められている。社会調査型質問の場合は、ベストアンサーの選定も主観的に行わざるを得ない。

社会調査型質問においては、ある回答の良否の判断を行う際に、その回答の内容だけに基いて判断することには問題があると筆者らは考える。たとえば、ある回答者があちこちで内容の一貫性がない回答をしていた場合、その回答者による回答が、内容だけを見れば優れていると感じられたとしても、それを良い回答であると判断することには疑問が残る。あるいは、非常に似た回答が 2 つ寄せられ、いずれかをベストアン

[†] 北陸先端科学技術大学院大学 知識科学研究科
School of Knowledge Science, Japan Advanced Institute of Science and Technology

[‡] 北陸先端科学技術大学院大学 ライフスタイルデザイン研究センター
Research Center for Innovative Lifestyle Design, Japan Advanced Institute of Science and Technology

サーとして選定する場合、質問者は自分と近い価値観を持つ回答者による回答をベストアンサーとしたいと考えることも想定される。このように、社会調査型質問に対する回答の良否を判断する際には、個々の回答の内容だけではなく、その回答を行った回答者の特性に関する情報も加味する必要があることが多いと思われる。

本稿では、社会調査型質問における各回答の良否判断を支援するために、各回答者の質問履歴および回答履歴を用いて、質問者が各回答者の特性について判断するための参考情報を提供するシステムを提案する。Yahoo! 知恵袋のデータを利用して、提案システムを使用した場合と使用しなかった場合とを比較することにより、提案手法の有効性を評価する。

以下、第2章ではQ&Aサイトについての関連研究を述べる。第3章では、本研究で提案する手法及び理由について説明する。第4章では、提案システムの概要について述べる。第5章では、提案したシステムにおける類似度判断の妥当性を評価する予備実験とその結果について述べる。第6章では、提案手法の有効性の評価実験について述べる。第7章はまとめである。

2. 関連研究

高田ら[3]は、Web情報を用いてコンテンツを補完する手法を提案した。情報検索型質問の場合は、回答の信憑性を判断するための情報をWebから収集し補完する。社会調査型質問の場合は、今の回答において不足している回答情報をほかの類似Q&Aコンテンツから収集し補完する。この手法は、得られた回答に情報を追加して回答の内容を改変することによって回答の質を向上させることを狙っているものであり、元の回答自体の良否判定を支援するものではない。

瀧ら[4]は、Q&Aコミュニティ内で正しい回答を知らない質問者が正しい回答を判断しようとすることを支援する信頼性指標を提案した。質問者が正しい回答かどうかを判断する際に参考にする情報を増やすために、回答内容だけでなく、回答の外部参照、質問のタイミングなどの付随情報をも手がかりとして利用していることを指摘した。この研究では、元の回答自体の良否判定を支援している点で、我々の研究と目的を同じくしている。ただし、そのために利用する情報として、我々の研究では回答者の質問・回答履歴を用いている点が異なっている。

佐々木ら[5]は、回答の根拠が提示されないことが、回答として提示された行動が適切かどうかの判断を困

難にしていることを指摘し、原因や理由を回答するwhy型Q&Aを応用し、取るべき行動と理由を合わせて提示するQ&Aを提案した。これにより、ただ取るべき行動のみを示されるよりは、回答の良否判定のための判断材料が増えることは間違いない。しかしながら、特に社会調査型質問の場合、行動と根拠の両方が常に提示されたとしても、その回答の良否を最終的に判断するためには、その回答者の特性に関する情報が必要であると考えられる。

3. 提案手法

本研究の基本的な発想は、社会調査型質問の場合、質問者の価値観と似た価値観を持つ回答者ほど、質問者が良いと判定する回答をする可能性が高いのではないかとするものである。ゆえに、各回答者が持つ価値観を質問者が把握することを支援できれば、各回答の良否判定の一助となることが期待される。価値観とは、そのものにどのような価値(意義)を認めるかについての、それぞれの人の考え方である[6]。価値観はきわめて個人差が大きく多様であり、しかも一個人がもつ価値観に関しても、しばしば非論理的であったり、一貫性がなかったり、曖昧であったりする。このため、価値観を機械処理によって抽出することは現実的では無いと思われる。

そこで本研究では、質問者(すなわち提案システムのユーザ)に対して、回答者の価値観そのものを提示するのではなく、価値観を推定するための素材となる情報を提示し、価値観の推定はユーザにまかせる手段をとる。価値観推定のための情報源として、回答者の過去の質問履歴と回答履歴を利用する。具体的には、以下の2つの情報を抽出し、質問者に提供する：

1. 質問履歴からの情報(図1)

ユーザによる質問と、ある回答者が過去に行ったすべての質問とを比較し、一定以上の類似度(類似度の算出方法は後述)を持つ類似質問を抽出する。ついで、抽出された類似質問に対して寄せられた回答のうち、その回答者がベストアンサーと評価した回答を抽出する。このベストアンサーの選定には、その回答者の価値観が反映されているはずである。そこで、こうして得られた類似質問とベストアンサーの組を、その回答者の価値観判断のための素材情報としてユーザに提示する。

2. 回答履歴からの情報(図2)

ユーザによる質問と、ある回答者が過去に回答したすべての質問とを比較し、一定以上の類似度

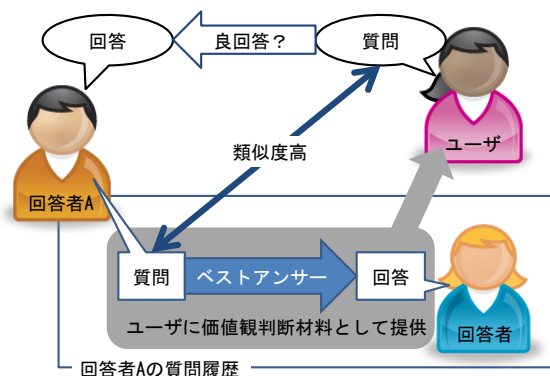


図1 回答者の質問履歴からの情報提供手法

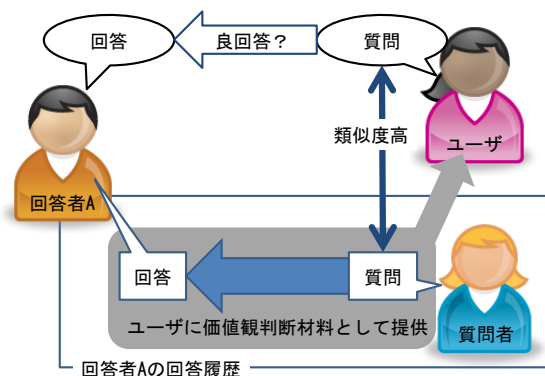


図2 回答者の回答履歴からの情報提供手法

(類似度の算出方法は後述)を持つ類似質問を、その回答者による回答と合わせて抽出する。この回答には、その回答者の価値観が反映されているはずである。そこで、こうして得られた類似質問と回答の組を、その回答者の価値観判断のための素材情報としてユーザに提示する。

以上により、ユーザによる質問に関する分野に対して、各回答者がどのような価値観を持っているかを推定するための有益な情報を提供することができると考えられる。

4. システム概要

本システムは、ウェブブラウザ上で動作するシステムとして構築し、Mozilla Firefox のアドオン機能として動作するように開発した。クライアント側システムは Greasemonkey 上で動作する JavaScript、サーバ側システムは Apache ウェブサーバ上で動作する PHP で実装した。システム構成の概要を図3に示す。質問者が Firefox で Q&A コンテンツを閲覧すると、

Greasemonkey が各回答の回答者のすぐ下にシステムを呼び出すためのボタンを追加する。追加されたボタンを押すと質問文と回答者リストを取得して、サーバ側に送信する。サーバ側では Yahoo! テキスト解析 API を用い、受信した質問文のキーワードを抽出する。また Yahoo! 知恵袋の情報をダウンロードし、回答者の回答履歴および質問履歴中の質問文をダウンロードする。ダウンロードした質問文も同様にキーワードを抽出する。質問文のキーワードと履歴中質問文のキーワードの類似度を計算し、類似質問を抽出して HTML ページを生成し Firefox に参考情報として表示する。

キーワードの抽出方法は、質問文を形態素解析し、名詞と動詞をキーワードとして抽出することに加え、カテゴリ情報も抽出する。類似度の計算方法にはベクトル空間法を用いた。キーワード毎に1次元とし、キーワードの出現回数をベクトルの長さとする質問文キーワードベクトルを作り、以下の式によって2つの質問の質問文キーワードベクトル \vec{Q}_1, \vec{Q}_2 の成す角の余弦

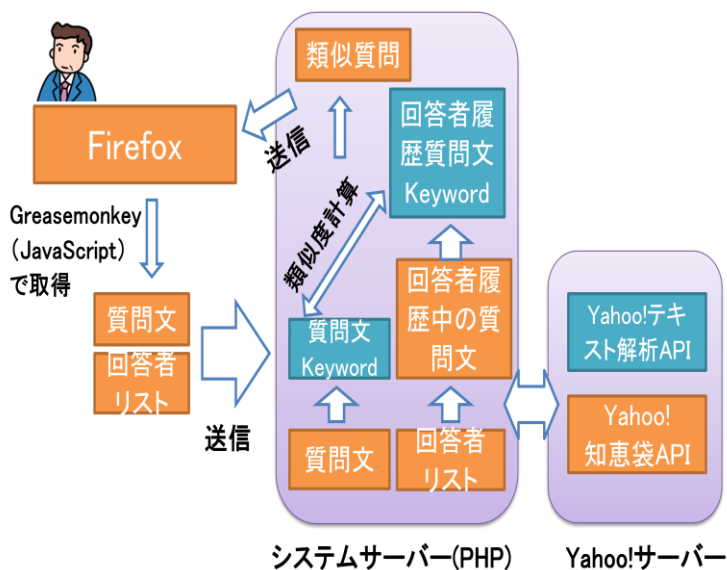


図3 システム概要

値 $\cos\theta$ を計算し類似度 $\text{sim}(\vec{Q}_1, \vec{Q}_2)$ とする.

$$\text{sim}(\vec{Q}_1, \vec{Q}_2) = \cos\theta = \frac{\vec{Q}_1 \cdot \vec{Q}_2}{\|\vec{Q}_1\| \|\vec{Q}_2\|}$$

システムのユーザ・インターフェースを図4に示す. 回答者アイコンの下には「参考情報」というボタンが表示される. 質問者がボタンを押すと, 別のページで類似質問の参考情報が提示される. 質問履歴を用いた情報提供の場合は, この回答者が以前質問した履歴の中で, 今回回答しようとしている質問との類似度がゼロでない質問に対して, この回答者が選んだベストアンサーを, 質問の類似度の高い順に提示する. 回答履歴を用いた情報提供の場合は, この回答者が以前回答した履歴の中で, 今回回答しようとしている質問との類似度がゼロでない質問に対するこの回答者の回答を, 類似度の高い順に提示する.

5. 予備実験

5.1 手順

4章で説明した手法によって求めた質問の類似度が, 実際の間人が評価する類似度と一致しているかどうかを検証するため, アンケートを実施した. 被験者は, 著者が在籍する大学院の学生5名(うち男性4人,

女性1人)である.

実験に使用した Q&A コンテンツには, Yahoo! 知恵袋に投稿された表1の5つのカテゴリから15個の Q&A コンテンツを選んだ. これら15個の各質問(以下, 「元質問」とする)に対して回答した回答者が, 過去に回答した別の質問および自らが質問者となって行った質問(以下, これらをまとめて「過去質問」とする)を集める. こうして得られた, 元質問と過去質問の組, 合計15組を被験者に提示し, 各組に関して, その組に含まれる元質問と各過去質問とを比較し, 1: 非常に似ていない 2: 似ていない 3: どちらかというように似ていない 4: どちらかというように似ている 5: 似ている 6: 非常に似ている 0: 分からない, の7段階評価で評価するアンケートに回答してもらった. 一方, 4章で述べた方法で, 各組に関して, その組に含まれる元質問と各過去質問との類似度を計算する. これを, 上述の被験者による評価結果と比較し, 両者の相関関係を分析する. なお, 15個の Q&A コンテンツに対する回答者の総数は27人, 過去質問の総数は616個であった.

5.2 結果

実験期間にも質問履歴から質問を削除した回答者が存在したため, 被験者間で読むことのできる質問に

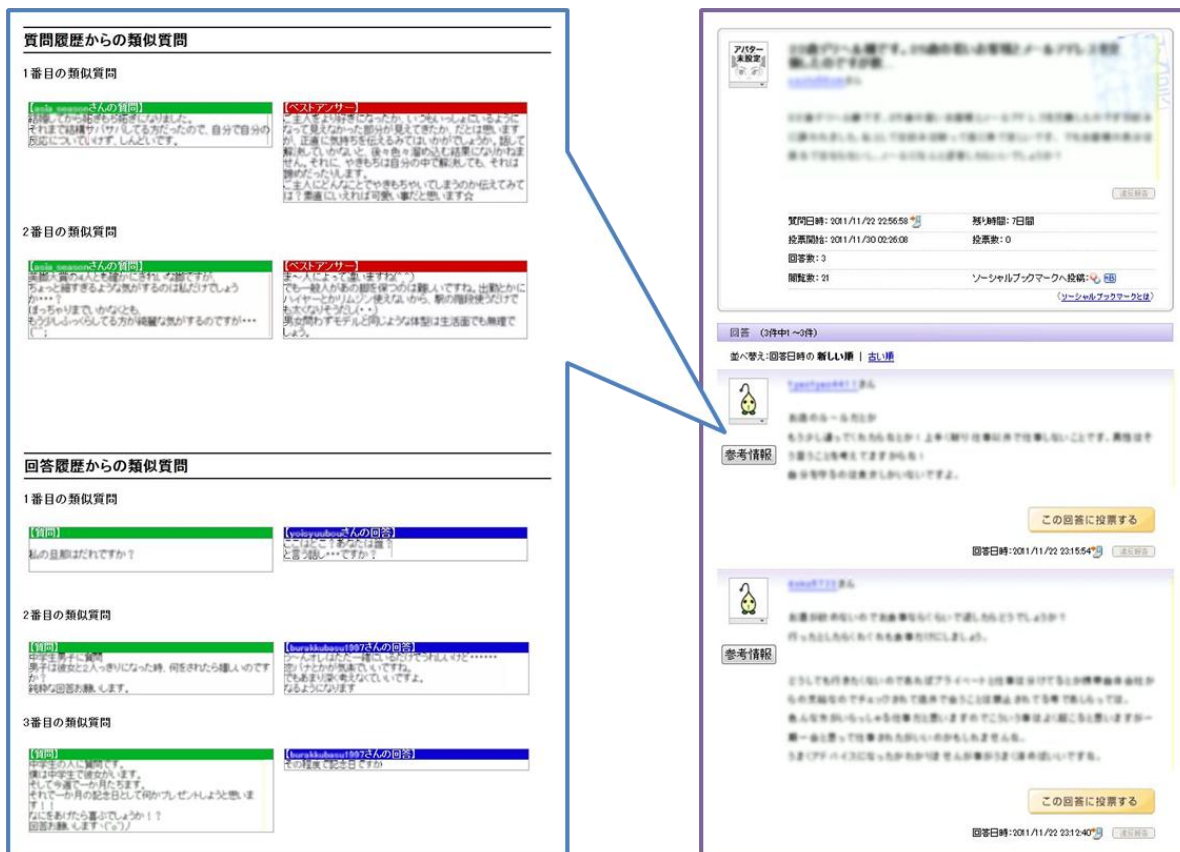


図4 システムのユーザ・インターフェース

表1 予備実験で使用した Q&A コンテンツ

1	エンターテインメントと趣味	①話題の人物, ②アニメ, ③携帯型ゲーム全般
2	暮らしと生活ガイド	④レシピ, ⑤洗濯, クリーニング, ⑥引っ越し
3	健康美容とファッション	⑦花粉症, ⑧ダイエット, ⑨ファッション
4	生き方と恋愛, 人間関係の悩み	⑩恋愛相談, ⑪友人, ⑫一人暮らし
5	子育てと学校	⑬子育ての悩み, ⑭大学院, ⑮大学の受験

差違があり, 最終的に 100 ぐらいの質問が除外された。残った 516 個の質問について, システムが算出した類似度 (0~1) と被験者の評定値 (1~6) とに関する相関係数を求めた。相関係数は 0.328 で, 弱い正の相関があり, 検定の結果 5%水準で有意であった。この結果から, 4 章で採用した質問の類似度計算手法により, 人間による評価と相関のある類似度が得られることがわかった。

6. システムの初期的評価実験

提案手法の有効性に関する初期的な評価のために, 4 章で述べたシステムを使用した場合としない場合の比較実験を実施した。本実験で使用したデータは「Yahoo! 知恵袋」の Q&A コンテンツである。その中からもっとも社会調査型質問に支配されると考えられるカテゴリ「恋愛相談」を選び, 実験で用いた。実験には, 平均回答 10 個の Q&A コンテンツを 3 つ用いた。被験者は, 著者が在籍する大学院の 20 代学生 6 名 (うち男性 2 人, 女性 4 人) である。質問者を支援することが目的であるから, 本来であれば被験者が質問者として質問を投稿すべきである。しかし, 回答を集めるのに時間がかかったり, 回答数が少なかったりするなどの問題が生じる可能性が予想されたため, 今回の実験では, 既存の Q&A コンテンツを使用することとした。この場合, 被験者は本来は第三者としての Q&A 閲覧者であり, 質問者の立場に立つことが難しい場合もあると思うが, 本実験では質問者の立場に立って評価してもらった。

6.1 実験手順

本実験はシステム利用前とシステム利用後の 2 部分に分けて実施された。まず被験者は, システムを使用せずに回答を閲覧し, ベストアンサーを選択する。選択が終わったら, システム利用前アンケートに回答する。アンケートでは, ベストアンサーの選択理由とそ

の確信度 (5 段階評価), ベストアンサー候補の数, および候補には挙げたがベストアンサーとしなかった回答に関して, それらをベストアンサーとしなかった理由を尋ねた。

次いで, 提案システムを利用して参考情報を提示する状態で, 被験者はもう一度ベストアンサーを選択する。その後, システム利用後のアンケートを回答する。アンケートには, システム利用前と同じ質問に加えて, 提示された参考情報からどの程度回答者の考え方を理解できたか, 回答者の回答履歴にどの程度一貫性があると思ったか, ベストアンサーを決める際にどの程度参考情報を参考にしたか, ベストアンサーの回答者の考え方と自分の考え方がどの程度似ていると思ったかについて, それぞれ 5 段階 (いずれについても, 1 が最も低い評価値であり, 5 が最も高い評価値とした) で評価してもらった。

提案システムの利用後, 質問者が利用前に選んでいたベストアンサーを変更したり, ベストアンサーに対する確信度が高くなったりすれば, それは提案システムが提供した参考情報が有効に影響した結果であると見なすことができよう。

6.2 結果

今回の実験は, 6 人の被験者全員に, 同じ質問 3 個を評価してもらい, 全部で 18 件の評価データを得た。

システム利用前と利用後の変化から見れば, システム利用後ベストアンサーとして選びたい回答者の候補数がシステム利用前より減ったのは 7 件, 変わらなかったのは 10 件, 増えたのは 1 件であった。このように, ベストアンサー候補数は, システムの利用によって減少する傾向があることがわかった。この結果は, 提案システムが提供する参考情報によって, 候補選択時の迷いがある程度解消され, 結果としてベストアンサー候補数がシステム利用前より減少した可能性を示唆している。

システム利用後にベストアンサーが変わったのは 4 件, 変わらなかったのは 14 件であった。この 14 件のうち, システム利用後, ベストアンサーの確信度がシステム利用前より高くなったのは 8 件であった。このように, システム利用によってもベストアンサーを変更することはあまり多くないが, 選ばれたベストアンサーに対する確信度は高くなる傾向にあることがわかった。ベストアンサーが変わらなかった 14 件についてアンケートの自由記述回答を見ると, 回答者のことが詳しく分かったため, ベストアンサーとして選ぶことをより強く確信できたという理由が多かった。この

表 2 システム利用後アンケートの結果

被験者の評価	平均
回答者の考え方に対する理解度	4.00
ベストアンサー回答者の考え方と自分の考え方の類似度	4.11
回答者履歴の一貫性	4.24
ベストアンサーとしたユーザ	4.44
ベストアンサーではないユーザ	4.00
質問履歴がベストアンサーを決める際にどれぐらい参考になったか	2.79
回答履歴がベストアンサーを決める際にどれぐらい参考になったか	4.30

結果から、システムが提示する情報がベストアンサー選択に役に立ったことが分かった。

表 2 に示す通り、参考情報から被験者による回答者の考え方に対する理解度の平均値は 4.0 であった。また被験者が、ベストアンサー回答者の考え方と自分の考え方が似ていると思った評価の平均値は 4.11 であった。さらに、被験者が回答者の履歴を読んで、一貫性があると思った評価の平均値は 4.24 であった。その内ベストアンサーとしたユーザの一貫性評価の平均値は 4.44 であり、ベストアンサーではない回答者の平均値 4.0 より高かった。これらの結果から、ベストアンサーとされた回答者の考え方は被験者の考え方と似ており、より一貫性あると感じられた回答者であることが分かった。

履歴には質問履歴と回答履歴がある。質問履歴の場合は、被験者がベストアンサーを決める時に参考になったと思う評価の平均値は 2.79 であったが、回答履歴の場合は、被験者が参考になったと思う評価の平均値は 4.3 で、質問履歴の場合より評価が高かった。ただしこの結果は、質問履歴の平均数が回答履歴より少なく、0 の場合もいくつかあるなど、履歴の数が少なかったことも原因の 1 つだと思われる。

7. まとめ

本稿では、Q&A サイトにおける社会調査型質問に対する回答を評価するために、回答者の質問履歴と回答履歴を用いて、各回答者の一貫性や価値観を判断するための素材情報を提供するシステムを提案した。提案システムを用いることにより、システムを用いない場合に選択したベストアンサーが覆されることはあまり多く生じなかったが、選択したベストアンサーに対

する確信度が向上する結果が得られた。この結果から、提案手法には一定の有効性があることが明らかとなった。

4章で述べた類似度算出方法は、非常に単純であり、人による類似度判定とある程度の相関を得られるものではあったが、まだ十分高い相関を得るには至っていない。今後、より高い相関を得られる類似度計算方法を検討したいと考えている。また、今回の実験において、被験者は本来は閲覧者の役割であるが、仮想的に質問者の立場に立ってもらうことによって評価を行った。実験の結果から見ると、質問に対して自分の心の中の答えに合致する回答を探していた被験者が多かったようである。この場合被験者は、実は質問者の立場ではなく、回答者の立場に立ってしまっていたと見なされよう。現実の質問者には、心の中に回答がない場合があるはずである。本当に迷っていて、どうすればいいかわからなくて、回答者のアドバイスを求める質問者がいる。このような質問者の立場に、第三者の被験者が立つことは難しい。このため、被験者に実際に質問者の立場で質問してもらい、得られた回答を用いて評価することが必要であると考えている。また、今回の実験では質問履歴と回答履歴の数が違ったことが、被験者の評価に影響を与えた可能性がある。今後は履歴の数をバランスよく提示することを検討したいと考える。また、考え方が似ていて、一貫性がある以外のベストアンサー選択理由も検討したいと考える。

謝辞 本研究は北陸先端科学技術大学院大学ライフスタイルデザイン研究センターの支援を受けて実施された。特に示唆に富むコメントをいただいた金井秀明准教授に感謝する。

参 考 文 献

- 1) <http://ja.wikipedia.org/wiki/ナレッジコミュニティ> (2011/12/13 現在)
- 2) 栗山和子, 神門典子: Q&A サイトにおける質問と回答の分析, 情報処理学会研究報告, (2009).
- 3) 高田夏希, 山本裕補, 小山聡, 田中克己: 質問応答コンテンツに対する Web による回答補完, DEIM Forum 2009 C4-6 (2009).
- 4) 瀧寛文, 森崎修司, 大平雅雄, 松本健一: Q&A コミュニティを対象とした回答の信頼性指標構築に向けた分析, 情報社会学会誌, Vol.4, No.1, pp.49-58 (2009).
- 5) 佐々木智, 藤井敦: 取るべき行動と理由を提示するヘルプデスク指向の質問応答システム, (2010).
- 6) 新明解国語辞典第 5 版, 三省堂