

機関番号：34316

研究種目：基盤研究(C)

研究期間：2008～2010

課題番号：20500106

研究課題名(和文) 心豊かなコミュニケーションを促進する質問作成支援システムの作成

研究課題名(英文) A study of a question making support system for promoting heartwarming communications

研究代表者

渡辺 靖彦 (WATANABE YASUHIKO)

龍谷大学・理工学部・講師

研究者番号：10288665

研究成果の概要(和文): まだ十分に整理できずモヤモヤした状態のユーザの質問に対し、欠落している情報を指摘することで質問を具体化するのを支援するシステムを実現するため、欠落している情報の指摘による質問作成支援の有効性の調査、欠落している情報を指摘するための知識の抽出、欠落している情報の特殊さの推定、表記選択支援のための優勢表記辞書の作成、匿名性の悪用が疑われるユーザの検出などについて研究を行った。

研究成果の概要(英文): In order to develop a support system for making good questions by indicating unclear points in questions, we studied effectiveness of indicating unclear points in questions, a method of acquiring knowledge for indicating unclear points in questions, a method of estimating specialty of indications in answers, a method of detecting unsuitable notational variants, and a method of detecting users suspected of abusing anonymity.

交付決定額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|--------|-----------|---------|-----------|
| 2008年度 | 1,000,000 | 300,000 | 1,300,000 |
| 2009年度 | 600,000 | 180,000 | 780,000 |
| 2010年度 | 1,300,000 | 390,000 | 1,690,000 |
| 年度 | | | |
| 年度 | | | |
| 総計 | 2,900,000 | 870,000 | 3,770,000 |

研究分野：メディア工学

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：Web コンテンツ

1. 研究開始当初の背景

高度情報化・グローバル社会の到来にともない、IT技術を積極的に利用したコミュニケーションの実現がますます重要になり、初心者から専門家までさまざまな人々が心豊かに関わりあえる環境を整備する必要性が高まっていた。例えば、Yahoo! 知恵袋などの質問応答(Q&A)サイトでは、「教えてほしい」「助けてほしい」と考えている人と「教えてあげたい」「だれかのためになれたら」と考

えている人が助けあって問題を解決する心豊かなコミュニケーションが行われている。こうしたIT技術を利用した助けあいを一層促進し、心を豊かにさせるコミュニケーションの場を形成するのに役立つ技術が求められていた。

2. 研究の目的

質問応答サイトでよい回答を得るためには、回答するのに必要な情報が質問で適切に

表現されていることが重要である。そこで、まだ十分に整理できずモヤモヤした状態のユーザの質問を具体化・詳細化するのを支援するために、回答するのに必要な情報が欠落していることをユーザに示す方法について研究を行う。そして、「教えてほしい」「助けてほしい」と考えている人と「教えてあげたい」「だれかのためになれたら」と考えている人をむすびつけ、心豊かなコミュニケーションを促進することをめざす。

3. 研究の方法

まだ十分に整理できずモヤモヤした状態のユーザの質問を具体化・詳細化するのを支援する方法として、回答するのに必要な情報が欠落していることをユーザに指摘する方法について以下の研究を行った。

- (1) 欠落している情報を指摘することが質問作成の支援に有効であるかどうかの調査
- (2) 欠落している情報を指摘するための知識の抽出
- (3) 欠落している情報の特殊さの推定
- (4) 表記選択支援のための優勢表記辞書の作成
- (5) 匿名性の悪用が疑われるユーザの検出

4. 研究成果

- (1) 欠落している情報を指摘することが質問作成の支援に有効であるかどうかの調査

質問で欠落している情報を指摘することが質問作成の支援に有効であるかどうかを検討するため、Yahoo! 知恵袋の「パソコン、周辺機器」カテゴリを対象に、回答するのに必要な情報が欠落していることを指摘されたユーザがその後質問を再投稿したケースについて

指摘を反映させた質問を再投稿して、よい回答は得られたのか

回答するのに必要な情報が欠落していることを回答者はどのように指摘したのか

回答するのに必要な情報が欠落していることを指摘した回答に質問者は高い評価を与えたのか

欠落している情報についての指摘を質問者はどのように利用して質問を再投稿したのかについて調べた。

再投稿された質問のうち、欠落している情報についての指摘が反映されているものは80%で、そのうち99%がよい回答を得ていた。一方、欠落している情報についての指摘が無視されて再投稿された質問では、よい回答が得られたのは62%であった。この結果から、指摘を反映させて質問を再投稿するとよい回答が得られやすいことがわかった。

回答での指摘のうち、質問を再投稿するのに最もよく利用されたのは表1のA1-5(解決

方法の提案)であり、次いでA1-2(状況の確認)とA1-3(属性の確認)が多かった。

質問での説明が不十分で欠落している情報があることを指摘した回答の84%は質問者からベストアンサーを与えられていた。「パソコン、周辺機器」カテゴリに投稿された回答のうちベストアンサーに選ばれたのは36%であるので、質問での説明が不十分で欠落している情報があることを指摘した回答に質問者は高い評価(ベストアンサー)を与えることが多いことがわかった。

欠落している情報についての指摘の利用の仕方でも多かったのは、表2のQ2-1(指摘された情報を質問に追加)の41%で、次いでQ2-5(指摘を無視して質問を再投稿)の20%であった。欠落している情報があるとの指摘だけで、質問者がその説明を自力で直して質問を再投稿できたケースは68%あった。この結果から、質問での説明が不十分なことを指摘するだけでも質問作成支援には有効であることがわかった。

表1: 必要な情報が欠落していることについての回答での指摘の仕方のタイプ

| TYPE | 説明 |
|------|---------------------------|
| A1-1 | 質問者が問題解決のために行ったことを確認 |
| A1-2 | 問題の状況について確認 |
| A1-3 | 対象の属性について確認 |
| A1-4 | A1-1~3以外について確認 |
| A1-5 | 解決方法を提案(ただし、その方法では解決できない) |

表2: 回答で指摘された情報についての再投稿した質問での利用の仕方のタイプ

| TYPE | 説明 |
|------|-------------------------|
| Q2-1 | 指摘された情報を質問に追加 |
| Q2-2 | 指摘された情報以外の情報を質問に追加 |
| Q2-3 | 指摘された情報では問題を解決できないことを説明 |
| Q2-4 | 指摘された情報そのものについて質問 |
| Q2-5 | 指摘された情報には触れず、同じ質問を再投稿 |

- (2) 欠落している情報を指摘するための知識の抽出

質問応答サイトでよい回答を得るためには、回答するのに必要な情報が適切に表現されている質問を投稿することが重要である。研究成果(1)で示したように、欠落している情報を指摘することはよい質問を作成するのを支援するのに有効である。質問で欠落している情報を指摘する質問作成支援システムを実現するためには、質問で書くべきなの

に欠落している情報を指摘するための知識が必要である。そこで Yahoo! 知恵袋の「パソコン、周辺機器」のカテゴリに投稿された質問とその回答からその質問で書くべき情報を含む文、すなわち

知っている情報をどこまで質問で述べた方がいいのか、初心者が判断する手がかりになる情報を含む文

知らないことであっても、初心者でも簡単に確認できる情報を含む文

を取り出す方法について検討を行った。サポートベクトルマシンを用いた実験の結果、この文をそれぞれ 0.4773 と 0.6503 の精度 (F 値) で取り出せた。なお、この研究では抽出した情報の信頼性については取り扱っていない。投稿された情報の信頼性については研究成果(5)で述べる。

(3) 質問で書くべき情報の特殊さの推定

ユーザがよい質問を作成するのを支援するため、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研究成果(2)で述べた方法で取り出した。しかし、取り出した質問に書くべき情報はさまざまでも数多くあり、それらをどのようにユーザに示すのが問題となった。そこで 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一般的な情報に区別してユーザに示す方法について検討を行った。図1に提案する手法の概要を示す。まず、内容の類似性にもとづいて質問をグループ分けする。質問の内容が類似しているかどうかは、質問から取り出した重要文(質問の中心になる文)が類似しているかどうかで判定する。次に、異なるグループに分類された質問に、質問に書くべき情報を指摘する回答で内容が類似しているものが投稿されていないか調べる。もし投稿されていれば、その質問に書くべき情報はさまざまな質問で書くべき一般的な情報と判定する。それ以外の情報は、特定の質問でだけ書くべき特殊な情報と判定する。

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一般的な情報であるか、それとも特殊な情報であるかを判定する実験を行った。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それぞれ研究成果(2)で述べた方法で取り出した。提案した方法でさまざまな質問で書くべき一般的な情報を 297 件、特定の質問でだけ書くべき特殊な情報は 83405 件取り出した。なお、表記のゆれ(例:「メーカー」と「メーラ」)があるため、さまざまな

質問で書くべき一般的な情報であると判定できなかった例が複数見つかった。このため、表記のゆれについて対応することができれば、より精度の高い推定ができると考えられる。

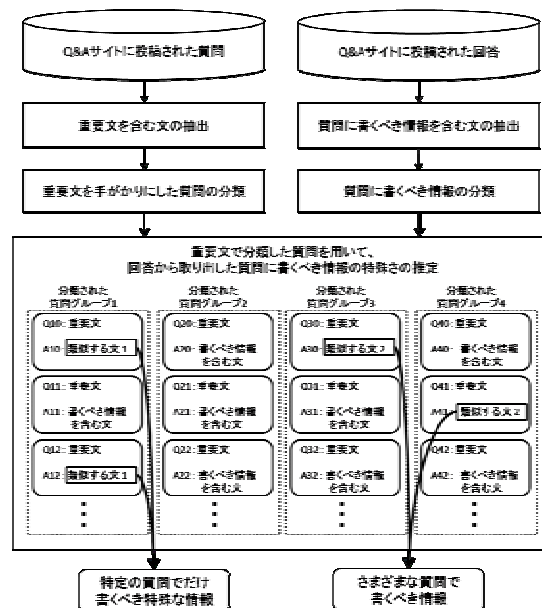


図 1: 質問に書くべき情報の特殊さの推定手法の概要

(4) 表記選択支援のための優勢表記辞書の作成

日本語の文書では、1つの単語に2つ以上の表記が存在する表記のゆれがよく見られる。表記のゆれは、情報検索や形態素解析の問題としてよく研究されているが、作文支援における問題としては、同一文書内での表記の一貫性の保持以外、これまであまり取り上げられていない。しかし、表記のゆれがある場合、どの表記がその文書において優勢に用いられているのかを書き手が判断するのを支援することは重要である。そこで、どの表記が優勢な表記であるか判断するのを支援するために、新聞記事と専門文書を用いて優勢表記辞書を作成した。優勢表記辞書の作成では、最も優勢な表記がどれくらい優勢に用いられているのかを示す優勢率をまず定義し、表記のゆれがあると判定された単語について二項検定によって優勢率の片側 95%信頼区間の下限が 0.5 よりも大きい単語の優勢表記を信頼できる優勢表記として取り出して優勢表記辞書に登録した。実験では、半年分の新聞記事からは 20929 単語の優勢表記を、160 個の学術論文からは 2840 単語の優勢表記を取り出し、優勢表記辞書を作成した。図 2 に新聞記事で表記のゆれが検出された単語と、それらのうち二項検定で信頼できる優勢表記があると判定されて優勢表記辞書に登

録された単語についての優勢率のヒストグラムを示す。この優勢表記辞書を利用した表記選択実験では、被験者に優勢な表記を選択させること、あるいは優勢な表記を選択していることを確認させることに優勢表記辞書が有効であることが確認できた。

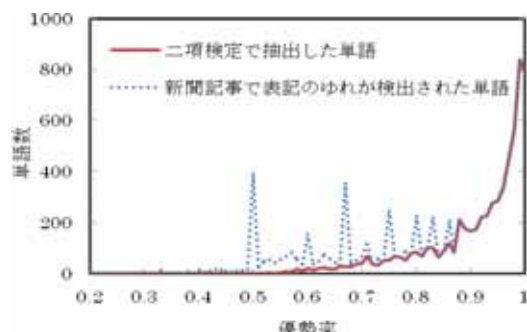


図 2：新聞記事で表記のゆれが検出された単語と、それのうち二項検定で信頼できる優勢表記があると判定されて優勢表記辞書に登録された単語の優勢率のヒストグラム

(5) 匿名性の悪用が疑われるユーザの検出

Q&A サイトでは匿名で投稿できることが重要である。匿名であることで、照れ、恥かしさ、評判などを気にしないで質問、相談、回答、議論などをすることができるからである。しかし、匿名性を悪用しようとするユーザもいる。したがって、Q&A サイトにおける心豊かなコミュニケーションを促進するためには、匿名性を悪用するユーザを検出することが重要である。そこで、特定の質問者と回答者について

その質問者の質問に対してその回答者が異常に多くの回答を投稿していないか

その質問者がその回答者から異常に多くの回答を受け取っていないか

その質問者がその回答者の回答に対して異常に頻繁に高い評価(ベストアンサー)を与えていないか

を二項検定を用いて判定し、その質問者と回答者が同一人物ではないかと疑われるケース、すなわち、複数のアカウントを利用して質問とその回答を繰り返し Q&A サイトへ投稿して回答者としての評価を操作しようとするユーザ(図 3)を検出する。Yahoo! 知恵袋(286 カテゴリ)に回答を投稿したすべてのユーザを対象に実験を行ったところ、異常な投稿を行なった質問者と回答者のユーザペアを 387 組検出した。このうち、回答者が複数のカテゴリで異常な投稿に関わっているユーザペアは 242 組であった。この 242 組のユーザペアから、「パソコン、周辺機器」と「政治、社会問題」で異常な投稿を行なったユーザペア 18 組を取り出して調査したところ、

意見や表記選択などで不自然な一致が見つかった。このため、提案した方法で質問者と回答者が同一人物であることが疑われるユーザペアを検出できることがわかった。質問者と回答者が同一人物であることが疑われるユーザペアの例は、信頼できる知識の抽出に役立つだけでなく、複数のアカウントを用いてソーシャルメディアでのコミュニケーションを操作しようとするユーザの目的や方法について調査するのに役立つと考えられる。



図 3：複数のアカウントを利用して質問とその回答を繰り返し Q&A サイトへ投稿し、回答者としての評価を操作しようとするユーザ

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 4 件)

石川尚季、梅本顕嗣、西村涼、渡辺靖彦、岡田至弘、Q&A サイトにおける自分の投稿に対する評価を操作するために複数のアカウントを利用するユーザの検出、情報社会学会論文誌、査読有、Vol.5、No.3、2011、pp.5-14

Murata,M., Uchimoto,K., Utiyama,M., Ma,Q., Nishimura,R., Watanabe,Y., Doi,K., and Torisawa,K., Category Estimation, Feature Extraction, and Error Correction, Cognitive Computation, Using the Maximum Entropy Method for Natural Language Processing、査読有、Vol.2、Issue 4、2010、pp.272-279

Nishimura,R., Watanabe,Y., Murata,M., Oota,Y., and Okada,Y., Estimation of Connectivity between Paragraphs in a Mail Text, Journal of Natural Language Processing、査読有、Vol.17、No.3、2010、pp.3-23

磯貝直毅、西村涼、渡辺靖彦、岡田至弘：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、情報社会学会論文誌、査読有、Vol.3、No.2、2009、pp.5-14

[学会発表](計 38 件)

小島正裕、水野友之、渡辺靖彦、岡田至弘、質問で不足している情報を回答で指摘されたユーザが再投稿した質問の調査、言語処理学会第 17 回年次大会、2011 年 3 月 8 日、愛知県、豊橋技術科学大学 谷口祐亮、小島正裕、西村涼、渡辺靖彦、岡田至弘、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、言語処理学会第 17 回年次大会、2011 年 3 月 8 日、愛知県、豊橋技術科学大学 梅本顕嗣、石川尚季、西村涼、渡辺靖彦、岡田至弘、回答の投稿から質問解決までの時間に着目した Q&A サイトにおける評価を操作しようとするユーザの検出、電子情報通信学会技術研究報告、2011 年 1 月 27 日、石川県、金沢大学サテライトプラザ

石川尚季、梅本顕嗣、西村涼、渡辺靖彦、岡田至弘、Q&A サイトにおける自分の投稿に対する評価を操作するために複数のアカウントを利用するユーザの検出、第 3 回知識共有コミュニティワークショップ、2010 年 12 月 18 日、京都府、龍谷大学 深草キャンパス

Kojima,M., Murata,M., Kazama,J., Kuroad,K., Fujita,A., Aramaki,E., Tsuchida,M., Watanabe,Y., and Torisawa,K., Using Various Features in Machine Learning to Obtain High Levels of Performance for Recognition of Japanese Notational Variants、PACLIC 24、2010 年 11 月 7 日、Tohoku University, Sendai, Japan.

小島正裕、西村涼、磯貝直毅、渡辺靖彦、岡田至弘、不足している情報を指摘する質問作成支援システム、人工知能学会情報編纂研究会 第 3 回研究会、2010 年 10 月 15 日、滋賀県、立命館大学 びわこ・くさつキャンパス

Ishikawa,N., Watanabe,Y., Nishimura,R., Umemoto,K., Okada,Y., and Murata,M., Detection of users suspected of using multiple user accounts and manipulating evaluations in a community site、IEEE NLPKE 2010、2010 年 8 月 22 日、Beijing, China.

Nishikawa,A., Watanabe,Y., Nishimura,R., and Okada,Y., A Context Sensitive Variant Dictionary for Supporting Variant Selection、LREC 2010、2010 年 5 月 20 日、Mediterranean Conference Center, Valletta, Malta

Nishikawa,A., Watanabe,Y., Nishimura,R., and Okada,Y., Writing Support System Based on a Context

Sensitive Variant Dictionary、CSEDU 2010、2010 年 4 月 9 日、Valencia, Spain
西川彩、渡辺靖彦、西村涼、村田真樹、岡田至弘、表記選択支援のための優勢表記辞書の作成、電子情報通信学会技術研究報告、2009 年 7 月 22 日、北海道、北見工業大学

Isogai,N., Nishimura,R., Watanabe,Y., and Okada,Y., Information Extraction for Supporting a Learner's Efforts to Recognize What the Learner did not Understand、CSEDU 2009、2009 年 3 月 23 日、Lisboa, Portugal

6. 研究組織

(1) 研究代表者

渡辺 靖彦 (WATANABE YASUHIKO)
龍谷大学・理工学部・講師
研究者番号：10288665

(2) 研究分担者

岡田 至弘 (OKADA YOSHIHIRO)
龍谷大学・理工学部・教授
研究者番号：30127063
馬 青 (MA QING)
龍谷大学・理工学部・教授
研究者番号：30358882
吉見 毅彦 (YOSHIMI TAKEHIKO)
龍谷大学・理工学部・准教授
研究者番号：50368031
南條 浩輝 (NANJO HIROAKI)
龍谷大学・理工学部・助教
研究者番号：50388162