

QA コミュニティにおける質問者の期待に基づく質問分類に関する一検討

渡邊 直人[†] 島田 諭^{††} 関 洋平^{††} 神門 典子^{†††} 佐藤 哲司^{††}

[†] 筑波大学情報学群知識情報・図書館学類 〒305-8550 茨城県つくば市春日 1-2

^{††} 筑波大学大学院図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2

^{†††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: [†]s0711658@u.tsukuba.ac.jp, ^{††}{sat,yohei,satoh}@slis.tsukuba.ac.jp, ^{†††}kando@nii.ac.jp

あらまし ユーザの質問に対し別のユーザが回答する, Yahoo!知恵袋などの QA コミュニティでは, 質問の内容や投稿のタイミングなど様々な要因によって, 適切な回答を得られないことがある. 多くの QA コミュニティでは, カテゴリに基づく質問の分類が行われているが, 質問者の回答への期待には, カテゴリの一致だけでは充足されないようなものもあると考えられる. 本論文では, 回答者への質問推薦を目的とし, 質問に対する客観的な正解の有無や, 回答の意見性への期待の有無などを考慮した質問分類を検討する. 実際の QA コミュニティに投稿された質問文に対し, 人手で質問分類を行った結果, および, その結果を用いて機械学習による分類を試みた結果について報告する.

キーワード QA コミュニティ, 質問分類, SVM

A Study for Questions Classification based on Questioner Demands in QA Communities

Naoto WATANABE[†], Satoshi SHIMADA^{††}, Yohei SEKI

^{††}, Noriko KANDO^{†††}, and Tetsuji SATOH^{††}

[†] College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba
1-2, Kasuga, Tsukuba, Ibaraki, 305-8550 Japan

^{††} Graduate School of Library and Information Science and Media Studies, University of Tsukuba
1-2, Kasuga, Tsukuba, Ibaraki, 305-8550 Japan

^{†††} National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan

E-mail: [†]s0711658@u.tsukuba.ac.jp, ^{††}{sat,yohei,satoh}@slis.tsukuba.ac.jp, ^{†††}kando@nii.ac.jp

Abstract In QA Communities like Yahoo! Chiebukuro (Yahoo! Answers in Japan), sometimes users get no appropriate answers by various factors, e.g., the content of question, the timing of contribution, etc. Though questions are classified based on various categories in QA communities, it is thought that there is the demands of questioner that are not fulfilled only by category matching. In this paper, we discuss the question classification that considers the expectation of questioner that the answer is right objectively, and that the answer contains subjective opinion. We report on the result of manual classification for actual articles in a QA community, and try classification by machine learning using those manual-annotated data as supervisor.

Key words QA Communities, Questions Classification, SVM

1. はじめに

近年, Yahoo!知恵袋^(注1)や教えて!goo^(注2)など, QA コミュニティの利用が増加している. ここでは, 身近に感じる疑問や生活の中での困り事を質問として投稿し, それに対して別の利用者が回答している. 日本語の QA コミュニティの中で最大級の Yahoo!知恵袋では, 2010 年 12 月の時点で, 約 690 万人の利用者によって, 5,100 万件の質問と 1 億 3,000 万件の回答がなされている. 質問者は匿名 (ユーザ ID) で質問を投稿でき

(注1): <http://chiebukuro.yahoo.co.jp/>

(注2): <http://oshiete.goo.ne.jp/>

ることから、健康など人に直接には聞きづらい質問も多くなされている。また、一般の Web 検索とは異なり、ユーザは自然文で質問できること、様々な年齢、性別、職業の異なる回答者などから多様な回答を得られることが特徴である。このように初心者から専門家まで幅広いユーザの興味・関心や疑問に応えることができることから、サービスの利用満足度も高いとの報告もある^(注3)。

多くの QA コミュニティでは、質問内容に応じたカテゴリが用意されている。回答者は、最初にカテゴリを選択し、そのカテゴリ内で回答可能な質問に回答することが基本となる。しかし、選択したカテゴリ内であっても、比較的容易に誰でも回答することができる質問から、知識や経験を持つ回答者でなければ回答することができない質問までが混在している。例えば、「ニュース、政治、国際情勢」のカテゴリにおいて「みなさんは今度の選挙の投票に行きますか?」という質問は前者にあたり、「比例区って何ですか?」という質問は後者にあたる。以上のように、同一カテゴリ内で質問のタイプが異なる質問が混在することで、回答する質問を選択することが困難になっている。その結果、質問者がたくさんの回答を期待して投稿をしても、数件の回答しかつかない、専門的な知識を問う質問に対して、回答者の主観的に回答がされているといった問題が生じている。

本研究は、QA コミュニティにおいて回答者の質問の選択を容易にすることを目的とし、新たな質問タイプの分類方法を提案する。質問者が回答に期待するタイプを質問タイプと定義し、質問タイプという観点から質問を分類することで、回答者が回答できる、あるいは回答をしたいと思う質問の選択を容易にすることを目指す。質問タイプを分類するにあたり、質問タイプの特徴は質問文に現れると考え、機械学習による分類を行う。

以下、本論文の構成を述べる。2 章では、関連研究と本研究の位置付けについて述べる。3 章では機械学習を用いた質問をタイプ分類する提案法について説明する。4 章では提案する手法について評価実験を行う。5 章で評価実験の考察をする。6 章でまとめと今後の課題について述べる。

2. 関連研究

QA コミュニティにおいて、様々な視点から質問タイプの分類が提案されている。質問の意図を分類するものと、質問のコミュニケーション上の特性を分類する研究とがある。

前者として、Kim ら [1] は、質問を「Information」、「Suggestion」、「Opinion」、「Others」の 4 タイプに分類し、質問タイプと回答の関係性について分析を行っている。

三浦ら [2] は、質問回答サイトへの参加動機と、質問に対する回答という行為でなされるコミュニケーションを、質問タイプ、性別、コミュニティへの参加スタイルから分析している。その過程で、質問には「正解あり」と「正解なし」の 2 タイプが存在することを指摘している。

栗山ら [3] は、質問に対して適切な回答の要素として質問タ

イプを考慮している。客観的な回答が得られる質問を「情報検索型」、適切な回答の基準がなく回答者の経験・価値観によって回答する質問を「社会調査型」と 2 タイプに分類している。更に、前者を「事実」、「真為」、「定義 / 記述」、「方法 / 手段」、「原因 / 理由」、「効果 / 結果」の 6 タイプに、後者を「助言」、「意見」、「嗜好」、「推薦」、「経験」の 5 タイプに細分化している。

後者として、Rodrigues ら [4] は、質問回答サイト内の質問者と回答者の交流関係を調査することを目的に、質問のタイプを分類している。回答、質問対象、質問と回答の関係性の 3 つの観点から、質問を 8 タイプに分類している。

QA コミュニティの質問・回答に対して、自然言語処理と機械学習を行うことで、質問に対する適切な回答の推定を行っている研究には、西原ら [5] や石川ら [6] がある。

西原らは、質問に適した回答とは質問に対する回答の文体の相性が良いことが重要であると仮定し、質問と回答の名詞、動詞、形容詞、文末表現を素性として機械学習を行っている。石川らは回答文字数、文末表現、URL、体験、経験の記述の有無、回答群の平均回答文字数、回答群の文末表現を素性として機械学習を行っている。

質問回答サイトにおいて、投稿された質問を回答者が選ぶことを考慮した質問タイプの分類はなされていない。また、本研究では質問が投稿された時点で質問を分類することを目指すため、質問文から得られる特徴量だけを機械学習の素性として用いることも特徴である。

3. 質問タイプの分類手法の提案

3.1 質問タイプの定義

本研究では、質問者が質問に対して回答に期待するタイプを質問タイプと定義し、表 1 のように、質問を 5 タイプに分類する。本研究ではまず、質問を「正解あり」と「正解なし」の 2 タイプに分類することを考える。質問者は質問を投稿する際に、ある程度自分の質問に対して、客観的な正解が存在するかどうかの予測を行う。質問者は「正解あり」タイプの質問に対しては、信頼性の高い、客観的な情報を求める場合が多く、回答の数が多くいかどうかはあまり問題ではない。一方で「正解なし」タイプの質問に対して、質問者は他の人の経験や意見を求め、多くの回答によって多面的に情報を集めたい、あるいは情報の信頼性を高めたいと考える場合が多い。このように「正解あり」と「正解なし」の質問では、質問者の求めるニーズが大きく異なる。「正解あり」タイプの質問を、事象の定義、真実、客観的な理由や手法を問う「事実」と、客観的な根拠、理由を問う「根拠」に分類する。「正解なし」のタイプについては、栗山ら [3] の分類を参考に、回答者自身の経験や体験がなければ回答できない「経験」、問題の解決方法を問う質問や情報提供を依頼する「提案」、推測や嗜好など、主観的に回答をして良い「意見」の 3 つに分類する。栗山らの分類と比較し、分類の区分は大きくなっているが、それは質問者だけでなく、回答者の視点に立ったときに回答できる質問というのは、最小単位の質問タイプだけではなく、ある程度の範囲の質問に対して可能であると考えたためである。

(注3): <http://it.nikkei.co.jp/internet/news/index.aspx?n=MMIT2E000002022009>

表 1 質問タイプの定義

質問タイプ	定義
事実	事象の定義, 真実, 客観的な理由や手法を問う質問
根拠	客観的な根拠, 理由を問う質問
経験	回答者の経験や体験がなければ回答できない質問
提案	問題の解決方法を問う質問や情報提供を依頼する
意見	推測, 嗜好など, 主観的に回答をしてよい質問

表 2 質問タイプの具体例

質問タイプ	具体例
事実	衛生管理者ってどんな会社でも、必ずおく必要がありますか？
根拠	蚊や虫に刺されると痒くなりますがなぜですか？ 蚊が吸う時に血を固まらせないように唾液が何かを同時に出し、その液で痒くなるのは師って知っているのですが、何で痒いのですか？
経験	子供を産むまでにどのくらい貯金されましたか？ また、その貯金額では厳しかったですか？余裕がありましたか？ 親の援助などについても合わせて教えてください。
提案	結婚記念日におすすめのレストランを教えてください。場所はできれば東京都内、東京近郊でお願いします。予算は二人で2万くらい。
意見	40万円もする時計を酔っ払って落とす男性って、どう思いますか？

具体的な質問タイプの例は、表 9 の通りである。

3.2 機械学習による分類

本研究では、SVM によって質問を質問タイプ別に分類する。SVM とは教師あり機械学習の 1 つで、与えられたデータに対して、マージンが最大にする基準によって分類をする。分類するデータは、QA コミュニティに投稿された質問記事に形態素解析を行い、素性として選んだ形態素の出現頻度による文章ベクトルを作成する。素性は、機械学習の対象とする質問記事から、品詞別に形態素を抽出し、下記の 4 種類の組み合わせを用いる。4 種類の素性で機械学習した際の、分類性能を比較する。

- (1) 名詞
- (2) 動詞 / 形容詞 / 副詞 / 助動詞
- (3) 名詞 / 動詞 / 形容詞 / 副詞 / 助動詞
- (4) 選択語

名詞によって抽出した語のうち、数値と記号は除外しておく。また、出現頻度が一度のみの語も除外しておく。名詞は全ての質問記事に含まれている品詞であり、分類性能のベースラインとして用いることができる。動詞、形容詞、副詞、助動詞は、質問記事によって含まれない品詞も存在するため、組み合わせで用いる。また、この組み合わせに名詞を加えた場合、どの程度結果に差が生じるかを比較する。選択語とは、経験則によって選択した、カテゴリによる依存せずに質問タイプの分類に有効だと考える 70 語である。選択語の一部を下記に示す。

「事実」に関連する語 方法 / 実際 / 可能

「根拠」に関連する語 理由 / 根拠 / どうして

「経験」に関連する語 経験 / みなさん / た

表 3 Yahoo!知恵袋データの詳細

収録期間	2004 年 4 月 1 日 — 2005 年 10 月 31 日
回答済みの質問	3,116,009 件
ベストアンサー	3,116,008 件
その他の回答	1,036,177 件

「提案」に関連する語 対処 / アドバイス / おすすめ

「意見」に関連する語 思う / 良い / 好き

4. 評価実験

本論文では、3.1 節で述べた質問タイプによる分類の可能性を確認するため、評価実験を行う。以下、4.1 節で判定者 1 名による分類、4.2 節で判定者 2 名による分類、4.3 節で機械学習による分類について述べる。

分析に用いるデータセット

本研究では、大学共同利用機関法人 国立情報研究所から提供を受けた、Yahoo!知恵袋のデータセットを用いる。データの詳細について、表 3 に示す。

収録期間は Yahoo!知恵袋のサービス開始の 2004 年 4 月 1 日から 2005 年 10 月 31 日までの、1 年 7 ヶ月間である。回答済みの質問とは、回答が 1 つ以上された質問である。回答済み質問には、質問の本文以外にも質問投稿日やカテゴリ名、回答数などの情報が付与されている。ベストアンサーとは、質問に対して行われた複数の回答のうち、質問者が最も参考になったとして選択した解答である。本研究では、回答済み質問だけを対象として研究を行う。

判定者 1 名と、判定者 2 名による分類では、表 3 に示すデータから、2005 年 9 月 1 日 0:00 以降で投稿順が連続する質問文を使用した。機械学習による分類では、NTCIR-8 CQA (コミュニティ QA テストコレクション)^(注4)に含まれる質問文のうち、2004 年 10 月 1 日から 2005 年 9 月 30 日までの期間に投稿された質問を対象とし、投稿順にかかわらずランダムな順序で使用した。事前に不適切な質問を除去し、1,020 件の質問文を抽出し、一定時間内で判定者ごとに異なる順序で判定させた結果、405 件の質問文に対して判定者 5 名から 30 名により判定を付与した。

4.1 判定者 1 名による分類

カテゴリと質問タイプとの関係性を確かめるために、判定者 1 名によって質問タイプの分類を行った。データセットから、利用者の多い「Yahoo!オークション」、「恋愛相談、人間関係の悩み」、「パソコンと周辺機器」、「政治、社会問題」(以下「オークション」、「恋愛相談」、「パソコン」、「政治と社会」)の 4 カテゴリ毎に 500 件、計 2,000 件の質問を選択し、事前に不適切な質問除外した上で、判定者 1 名が人手で分類を行った。その結果を表 4 に示す。

表 4 から、以下の特徴がわかる

- 根拠タイプの質問は、他のタイプに比べて少ない
- 「パソコン」では、半数以上が事実タイプの質問

(注4): <http://research.nii.ac.jp/ntcir/permission/ntcir-8/perm-ja-CQA.html>

表 4 判定者 1 名による分類結果

カテゴリ	質問タイプ					計
	事実	根拠	経験	提案	意見	
Yahoo!オークション	142	2	55	127	149	475
恋愛相談、人間関係の悩み	10	0	54	150	264	478
パソコンと周辺機器	308	3	8	118	54	491
政治、社会問題	114	8	13	37	295	467
計	574	13	130	432	762	1911

表 5 カテゴリ別 kappa 係数

カテゴリ	kappa 係数
Yahoo!オークション	0.66
恋愛相談、人間関係の悩み	0.62
パソコンと周辺機器	0.47
政治、社会問題	0.69

• 「恋愛相談」「政治と社会」では、半数以上が意見タイプの質問

• 「オークション」では、他のカテゴリと比較すると質問タイプの分布の偏りが少ない

4.2 判定者 2 名による分類の一致度

質問タイプを分類するに当たり、本研究の提案する質問タイプの分類が一般的に正しく分類が可能か実験を行った。前節の同様のデータから、カテゴリ毎に質問 100 件を選択し、判定者 2 名によって人手で分類を行った。判定者は共に情報学分野の学生である。2 人の分類した結果から、分類判定の一致度として kappa 係数 [7] を用いる。kappa 係数は次の式によって求められる。 P_o は、判定者間で質問タイプの分類が一致した質問数と対象とした質問数から算出した観測された一致率である。 P_c は、判定者がそれぞれタイプ分類の数から観測した、偶然による一致率である。

$$kappa = \frac{P_o - P_c}{1 - P_c} \quad (1)$$

kappa 係数の判定基準は次の通りである。

- 0.0 - 0.2: slight
- 0.21 - 0.4: fair
- 0.41 - 0.6: moderate
- 0.61 - 0.8: substantial
- 0.81 - 1.0: almost perfect

上記で示した kappa 係数の計算式と判定基準を用いて、カテゴリ別に kappa 係数を計算した結果を表 5 に示す。

表 5 から「オークション」「恋愛相談」「政治」では substantial (かなりの一致)、「パソコン」では moderate (中程度の一致) となったことから、人手による質問タイプの分類は、カテゴリによらず安定して行えることが示唆された。「パソコン」での一致率がやや低くなった理由としては、判定者に専門知識が要求される質問が多く含まれていたことが挙げられる。すなわち、判定者が質問に対する正解を知っているかいないかによって、質問タイプを「事実」とするか「提案」とするかが分かれるため、他のカテゴリよりも一致度が低くなったと考えられる。

表 6 判定者 5—30 名による分類結果

	事実	根拠	経験	提案	意見	計
Yahoo!オークション	13	0	5	2	10	30
恋愛相談、人間関係の悩み	0	0	7	0	21	28
Yahoo!知恵袋	3	2	0	1	13	19
パソコン、周辺機器	12	0	0	5	2	19
言葉、語学	9	2	2	2	0	15
政治、社会問題	8	0	0	0	7	15
テレビ、ラジオ	8	0	0	0	5	13
子育ての悩み	1	0	5	3	4	13
自動車	5	0	4	2	0	11
国内	2	0	3	2	3	10
妊娠、出産	3	0	5	0	1	9
病気、症状、ヘルスケア	3	1	0	3	1	8
レシピ、調理法	3	0	2	3	0	8
動物、植物、ペット	5	1	0	1	1	8
交通、地図	3	0	1	0	3	7
野球	1	1	0	0	4	6
話題の人物	4	0	0	0	3	7
数学、サイエンス	5	0	0	1	0	6
健康、病気、ダイエット	1	0	1	3	0	5
芸能人、タレント	0	0	0	1	4	5
その他 (76 カテゴリ)	68	6	18	24	45	163
計	157	13	53	53	129	405

4.3 機械学習による分類

SVM はツールとして LibSVM^(注5)を使用し、デフォルトの設定のまま使用した。

教師データと評価データ

前節の判定者 1 名による分類では、質問タイプはカテゴリによって特徴的な分布を表すということが判明した。また、判定者 2 名による分類では質問に対し、ある程度の一致したタイプを付与することができるが、判定が難しい質問も存在することが判明した。そこで、機械学習に用いる教師データ、評価データは、NTCIR-8 CQA に含まれる質問から、全カテゴリを対象に 405 件をランダムに抽出し、判定者 5—30 名によってタイプの付与をした質問を用いた。

タイプを付与した結果を表 6 に示す。

根拠タイプの質問は 13 件である。機械学習を行うには十分な量を得ることが出来なかった。従って、機械学習には「事実」「経験」「提案」「意見」の 4 タイプを対象にした。タイプ毎に 53 件の質問を選択し、40 件を教師データ、13 件を評価データとして用いた。また、これら 212 件の質問に対して形態素解析を行い、提案手法で示した品詞毎に、素性となる語を抽出した。抽出された品詞の組み合わせ別の単語数を表 7 に示す。

素性毎に、質問タイプ別の精度 (Precision)、再現率 (Recall)、F 値を以下の式に基づいて算出し、F 値のマクロ平均を算出した。

$$Precision = \frac{\text{質問タイプに正しく分類された質問数}}{\text{質問タイプに分類した質問数}} \quad (2)$$

(注5): <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 7 品詞別の単語数

素性品詞	単語数
名詞	1,285
動詞 + 形容詞 + 副詞 + 助動詞	452
名詞 + 動詞 + 形容詞 + 副詞 + 助動詞	1,737
選択語	70

表 8 精度, 再現率, F 値, 平均 F 値

素性	質問タイプ	精度	再現率	F 値	平均 F 値
名詞	事実	0.00	0.00	0.00	0.24
	経験	0.33	0.08	0.13	
	提案	0.00	0.00	0.00	
	意見	0.25	0.92	0.39	
動詞 + 形容詞 + 副詞 + 助動詞	事実	0.25	0.08	0.12	0.13
	経験	0.33	0.31	0.32	
	提案	1.00	0.08	0.14	
	意見	0.26	0.77	0.39	
名詞 + 動詞 + 形容詞 + 副詞 + 助動詞	事実	0.00	0.00	0.00	0.26
	経験	0.20	0.08	0.11	
	提案	0.00	0.00	0.00	
	意見	0.27	0.92	0.41	
選択語	事実	0.57	0.31	0.40	0.42
	経験	0.46	0.46	0.46	
	提案	0.57	0.31	0.38	
	意見	0.33	0.62	0.43	

$$Recall = \frac{\text{質問タイプに正しく分類された質問数}}{\text{質問タイプの質問数}} \quad (3)$$

$$F \text{ 値} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

上記の式によって算出した結果を, 表 8 に示す.

表 8 から, 品詞では「意見」の再現率が高く, 分類の漏れが少ないことがわかる. 選択語では「意見」の再現率が下がるが, 他のタイプの再現率が上がる「事実」と「提案」の上がり方は大きく, 選択語を用いることで精度, 再現率の結果が共に良くなっている. 「経験」も選択語による結果が一番高い値を示しているが, 名詞をいれない品詞の組み合わせの場合も多少良い結果を示している.

5. 考 察

本研究では, QA コミュニティの質問を, 機械学習によって質問タイプ毎に分類を行った. その結果, 名詞のみを素性を素性とした場合に, 動詞/形容詞/助動詞/副詞を素性とした場合と比較して, 平均 F 値が高い値を示した. 機械学習による分類の評価実験では, カテゴリによる質問の選択を行っていないため, カテゴリに依存するような名詞は少ない. 従って, 名詞を用いることが, 動詞/形容詞/副詞という内容語と, 助動詞という機能語を組み合わせるよりも効果的であることがわかる. さらに, 名詞/動詞/形容詞/助動詞/副詞と組み合わせることで, 平均 F 値が高くなっていることから, 動詞/形容詞/副詞/助動詞も, 多少は質問タイプを分類する上で影響があると考えられる.

また, 経験則によって選択した 70 語を素性とした場合, 名

詞/動詞/形容詞/助動詞/副詞を素性とした各場合と比較して, 分類性能を示す平均 F 値が 2 倍近くの結果を示した. すなわち, 本研究が提案する質問タイプの分類において, 質問タイプの分類は特定の語が強く影響していることが判明した. しかし, 平均 F 値の 0.42 という値は, 今後より多くの質問を自動で分類することを考えると, 高い値ではない. 平均 F 値が低い要因として, 教師データの不足と質問タイプの判定の難しさが考えられる.

特に, 質問タイプの判定は, 判定者を多くしても, 判定が割れる質問が存在した. 判定が割れた質問には, 表 9 のようなものがあつた. 表 9 の 1 の質問では, 質問者がサッカー観戦の「経験」を持つ人に対し, その経験を基にした情報を得たいのか, それとも経験がなくとも考えられる「提案」をたくさんもらえればいいのかは, 投稿者でなければ判断ができないところがある. 表 9 の 2 の質問では, 「事実」である客観的な医学知識を提供してもらいたいのか, あるいは質問者がとるべき次の行動について, 参考程度の「意見」を集めたいのかは, この文面だけでは判断が難しい. 以上のような点から, 質問タイプの分類が難しい質問も, QA コミュニティの中には存在している. 表 9 の 3 の質問も同様に, 「経験」と「意見」に判定者によって判定が割れた質問である. また, 文章が短いために抽出できる文章ベクトルはが疎行列に近くなってしまう, 機械学習による分類も困難であると考えられる.

以上のことから, 今後の課題として, 教師データを増加と質問タイプの判定基準の明確化が挙げられる.

6. おわりに

本研究では, 質問回答サイトの質問を, 質問者が回答に期待するタイプを質問タイプと定義し, 事実, 根拠, 経験, 提案, 意見の 5 タイプに分類した. 判定者 1 名による分類では, カテゴリに応じて質問タイプの分布が異なることが確認できた. 判定者 2 名による分類では, 質問を提案するタイプに分類することが一般的に可能であることが確認できた. 最後に機械学習による分類を行い, 品詞の組み合わせによる形態素と, 経験則から質問タイプを特徴づけると考えられる形態素を素性とした場合で, その分類性能の比較を行った. 機械学習による分類では, 経験則によって選択した 70 語を素性として分類を行った結果, F 値の平均が 0.42 と, 名詞や動詞などの品詞による単語を素性とした場合と比較して 2 倍程度良好な結果が得られた. このことから, 質問タイプはある特定の語によって強く影響を受けていると考えられる.

分類性能を向上させるための今後の課題として, 教師データを増加と質問タイプの判定基準の明確化が挙げられる.

謝 辞

本研究の一部は科研費 (21500091) の助成を受けたものである. 本研究の実装・評価に際し, 大学共同利用機関法人 国立情報学研究所から提供を受けた, Yahoo!知恵袋のデータおよび NTCIR-8 CQA (コミュニティ QA テストコレクション) を利

表 9 判定の難しい質問の例

1. 明日初めてサッカーの試合を観戦しに行くのですが、なにせ初めてなのに加え、W 杯予選ということもあって、どのような状況になるのか予想が付きません。また、観戦するにあたって、重要なマナーや、「サッカー観戦するならこれは当然！」ということがあれば是非教えてください気をつけたほうがいいこと、やっておいたほうがいいこと、などもあれば教えてください。よろしくお願いいたします。
2. 旦那のことですが、先日インフルエンザで高熱を出してから味覚が変だと言っています。一部のものを除いて、何を食べても納豆のような味がするらしいのです。すぐ治るだろうと軽く考えていましたが、なかなか治らず不安になってきました。自然と治るのを待つしかないのでしょうか？
3. 恋は追いますか？追われますか？

用している．ここに記して謝意を示す．

文 献

- [1] Soojung Kim and Jung Sun Oh and Sanghee Oh. Best answer selection criteria in a social Q&A site from the user oriented relevance perspective. *American Society for Information Science and Technology (ASIS&T) 2007 Annual meeting*, 2007.
- [2] 三浦麻子, 川浦康至. 人はなぜ知識共有コミュニティに参加するのか：質問行動と回答行動の分析. *社会心理学研究*, Vol. 23, No. 3, 2008.
- [3] 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*, Vol. 2009-FI-95, No. 19, 2009.
- [4] Eduarda Mendes Rodrigues and Natasa Milic-Frayling. Socializing or Knowledge Sharing? Characterizing Social Intent in Community Question Answering. *CIKM2009*, Vol. 23, , 2009.
- [5] 西原陽子, 松村真宏, 谷内田正彦. QA サイトにおける質問に適した回答の判定. *NLP 若手の会 第 2 回シンポジウム*, 2007.
- [6] 石川大介, 栗山和子, 酒井哲也, 関洋平, 神門典子. QA サイトにおけるベストアンサー推定の分析とその機械学習への応用. *情報知識学会 第 18 回 (2010 年度) 年次大会*, Vol. 20, No. 2, 2010.
- [7] J.Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, 1960.