

1 つの質問に投稿された異なる意見の回答の抽出

小島 正裕 南口 勝志 西村 涼 渡辺 靖彦 岡田 至弘

龍谷大学大学院 理工学研究科 情報メディア学専攻

{t10m101,t080456}@mail.ryukoku.ac.jp, r_nishimura@afc.ryukoku.ac.jp,
{watanabe,okada}@rins.ryukoku.ac.jp

1 はじめに

近年、質問を投稿しておくとのユーザが答えてくれる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。また、質問は投稿しないが、これまでに投稿された質問と回答を閲覧して、問題を解決するのに利用するユーザ (以後、2 次利用ユーザ) も多い。2 次利用ユーザは、その質問で最もよいと判定された回答 (以下、ベストアンサーあるいは BA) だけを読み、それ以外の回答 (以下、ノーマルアンサーあるいは NA) は読まないおそれがある。ノーマルアンサーでも場合によっては、2 次利用ユーザにとって重要な情報が含まれているかもしれない。特に、ベストアンサーとは異なる意見の回答があれば、それに目を通すことは重要である。例えば、以下の (質問 1) に対する (回答 1-BA) と (回答 1-NA) は意見が異なる。

(質問 1) 512M を 2 枚ざしと 1G を 1 枚ざしはどちらが性能がよくなるのでしょうか？

(回答 1-BA) 体感できる差は無いと思います。

(回答 1-NA) チップセットがデュアルチャンネル対応なら 512 二枚ざしの方が速いと思います。

この例では、2 次利用ユーザのチップセットがデュアルチャンネル対応である場合、(回答 1-BA) よりむしろ、(回答 1-NA) の意見が重要である。このように、ベストアンサーとは異なる意見の回答には、2 次利用ユーザにとって重要な情報が含まれていることがある。そこで、ベストアンサーとは異なる意見の回答をユーザに知らせるため、異なる意見の回答を抽出する方法について検討する。

これまで、意見文を対象としたテキストマイニングに関する研究がさかに行われており [1]、意見文から評判情報を抽出する研究 [2, 3] や意見文を利用した文書評価の研究 [4, 5] などがある。しかし、Q&A サイトにおいて 1 つの質問に投稿された異なる意見の抽出する研究はあまり行われていない。そこで本研究では、ベストアンサーとは異なる意見の回答をユーザに知らせるため、異なる意見の回答の抽出を機械学習を用いて行う。

2 Q&A サイトにおいて 1 つの質問に投稿された異なる意見の回答

本研究では、Q&A サイトの例として Yahoo!知恵袋^{*1} を取り上げる。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約 311 万件の質問、約 1347 万件の回答が国立情報学研究所から公開されている^{*2}。これらの質問には 1 件以上の回答が必ずあり^{*3}、そのうち 1 件がベストアンサーに選ばれている。この Yahoo!知恵袋のデータに収録されている以下の

- 「健康、病気、ダイエット」カテゴリ
- 「パソコン、周辺機器」カテゴリ
- 「レシピ、調理法」カテゴリ

に投稿された質問とその回答を対象に異なる意見の回答の抽出を行う。

この 3 つのカテゴリには、以下のようなベストアンサーとは異なる意見の回答があった。

回答で仮定している状況や環境などの条件が異なる 1 章の例がこのタイプの回答になる。2 次利用ユーザによっては、このタイプの回答の方が有効な場合がある。

回答者の状況や環境などの条件が異なる このタイプの回答は、意見や評判の収集を目的とした質問に対して投稿された回答に多い。これは、意見や評判の収集が目的の質問では、さまざまな回答者の意見が求められるからである。

(質問 2) みなさんのお宅の定番のお味噌汁の具は何ですか？

(回答 2-BA) 我が家では、きゅうりを味噌汁に入れます。

(回答 2-NA) キャベツの芯なんか、捨てるのはもったいない！千切りにして味噌汁の具にしてもいいですよ

^{*1} <http://chiebukuro.yahoo.co.jp/>

^{*2} <http://research.nii.ac.jp/tdc/chiebukuro.html>

^{*3} 回答が投稿されなかった質問は削除されるので、Yahoo!知恵袋の公開データに収録されている質問には 1 件以上の回答が必ずある。

回答が間違っている可能性がある このタイプの回答は、ベストアンサーがこの回答のどちらかが間違っている可能性があると考えられる。

(質問 3) 人の肺は 1 度汚れたら、汚れっ放しですか？

(回答 3-BA) タバコを吸って汚れた肺は、きれいにするのにすっていた期間と同じ時間がかかります。

(回答 3-NA) 汚れっぱなしです。

本研究では、回答の正当性や信頼性を評価しない。ただし、このタイプの回答を 2 次利用ユーザに知らせ、回答の正当性や信頼性を確認させることは重要である。

1 件の質問に複数の質問があり、答えた質問が異なるこのタイプの回答は、ベストアンサーが答えた質問とは異なる質問に答えているので、厳密には、ベストアンサーとは異なる意見の回答とはいえない。しかし、1 件の質問に複数の質問が含まれることはめずらしくない。

(質問 4) ピーマンを大量に使う簡単レシピを教えてください。あと少し傷んだナスでできるレシピもあれば教えてください。

(回答 4-BA) 「3 色ピーマンのナムル」ピーマン 3 個、赤ピーマン 1/4 個、黄ピーマン 1/4 個で 2 人前です。

(回答 4-NA) なすは麻婆茄子がいいと思います。

このため、ベストアンサーとは異なる質問に答えた回答があるかどうかの情報は、2 次利用ユーザにとって重要である。そこで本研究では、このタイプの回答も取り扱うこととする。

本研究では、以上のような場合に異なる意見の回答として抽出を行う。

3 実験データの作成

実験データは、それぞれのカテゴリで 2 件以上の回答が投稿された質問を無作為に 1000 件ずつ取り出して作成する。それぞれのカテゴリから無作為に取り出した質問 Q_i ($i = 1, \dots, 1000$) には、ベストアンサーである回答 BA_i とノーマルアンサーである回答 NA_{ij} ($j = 1, \dots, L$) があるとする。実験データは、 Q_i -(回答 1)-(回答 2) という構造をもつ。まず、回答 1 が BA_i 、回答 2 が NA_{in} ($n = 1, \dots, L$) である実験データ $Q_i-BA_i-NA_{in}$ を作成する。この作成した実験データに対して、異なる意見の回答であるかのタグ付けを行う。次に、実験データの総数を増やすために BA_i と NA_{in} を入れ替え、回答 1 が

表 1 実験データの内訳

カテゴリ	データ数	異なる意見の回答数
健康、病気、ダイエット	5880	3800(64.6%)
パソコン、周辺機器	4906	1934(39.4%)
レシピ、調理法	8652	6322(73.1%)

表 2 異なる意見の回答の抽出に用いる素性

s1	質問に含まれる自立語と未定義語の 1-gram
s2	質問に含まれる自立語と未定義語の 2-gram
s3	質問に含まれる自立語と未定義語の 3-gram
s4	質問に含まれる自立語と未定義語の読み
s5	質問に含まれる自立語と未定義語の原形
s6	質問に含まれる自立語と未定義語の数
s7	質問に含まれる自立語と未定義語ごとの数
s8	回答 1 に含まれる自立語と未定義語の 1-gram
s9	回答 1 に含まれる自立語と未定義語の 2-gram
s10	回答 1 に含まれる自立語と未定義語の 3-gram
s11	回答 1 に含まれる自立語と未定義語の読み
s12	回答 1 に含まれる自立語と未定義語の原形
s13	回答 1 に含まれる自立語と未定義語の数
s14	回答 1 に含まれる自立語と未定義語ごとの数
s15	回答 2 に含まれる自立語と未定義語の 1-gram
s16	回答 2 に含まれる自立語と未定義語の 2-gram
s17	回答 2 に含まれる自立語と未定義語の 3-gram
s18	回答 2 に含まれる自立語と未定義語の読み
s19	回答 2 に含まれる自立語と未定義語の原形
s20	回答 2 に含まれる自立語と未定義語の数
s21	回答 2 に含まれる自立語と未定義語ごとの数
s22	回答 1 と回答 2 両方に含まれる自立語と未定義語の 1-gram
s23	回答 1 と回答 2 両方に含まれる自立語と未定義語の 2-gram
s24	回答 1 と回答 2 両方に含まれる自立語と未定義語の 3-gram
s25	回答 1 と回答 2 両方に含まれる自立語と未定義語の読み
s26	回答 1 と回答 2 両方に含まれる自立語と未定義語の原形
s27	回答 1 と回答 2 両方に含まれる自立語と未定義語の数

NA_{in} ($n = 1, \dots, L$)、回答 2 が BA_i である実験データ $Q_i-NA_{in}-BA_i$ を作成する。なお、 BA_i と NA_{in} を入れ替えた後の実験データのタグは、入れ替える前の実験データのタグと同じとする。このため、質問 Q_i には、ベストアンサー 1 件とノーマルアンサー L 件の合計 $(1 + L)$ 件の回答があるので、 $2L$ 件の実験データを作成することができる。表 1 に実験データの内訳を示す。なお、カッコ内の数字は異なる意見の回答が含まれるデータの割合を示す。

4 実験で用いる素性

機械学習で用いる素性を表 2 に示す。 $s1 \sim s7$ は異なる意見の回答を抽出する対象の質問から取り出す素性である。 $s8 \sim s14$ は異なる意見の回答を抽出する対象の回答 1 から取り出す素性である。 $s15 \sim s21$ は異なる意見の回答を抽出する対象の回答 2 から取り出す素性である。 $s22 \sim$

表 3 実験結果

学習データ	分類器	テストデータ											
		健康、病気、ダイエット				パソコン、周辺機器				レシピ、調理法			
		正解率	再現率	適合率	F 値	正解率	再現率	適合率	F 値	正解率	再現率	適合率	F 値
健康、病気、 ダイエット	ME	75.4%	0.795	0.789	0.792	55.9%	0.511	0.379	0.435	60.5%	0.663	0.745	0.702
	SVM	75.0%	0.785	0.790	0.787	39.6%	0.921	0.346	0.503	70.6%	0.974	0.712	0.823
パソコン、 周辺機器	ME	45.4%	0.131	0.692	0.220	79.4%	0.650	0.708	0.678	38.9%	0.165	0.805	0.274
	SVM	41.3%	0.003	0.750	0.006	79.7%	0.660	0.710	0.684	30.2%	0.001	0.667	0.002
レシピ、 調理法	ME	53.6%	0.410	0.673	0.510	61.2%	0.263	0.379	0.311	84.0%	0.902	0.873	0.887
	SVM	59.1%	0.918	0.600	0.726	43.8%	0.838	0.354	0.498	83.6%	0.893	0.876	0.884

s27 は異なる意見の回答を抽出する対象の回答 1 と回答 2 の両方から取り出す素性である。素性に利用した形態素を自立語と未定義語にしたのは、付属語は異なる意見の回答を判定するのに重要でないと考えたからである。また、Yahoo!知恵袋に投稿される質問や回答には、正しく付属語が使われていないことがある。これらのことから、素性の情報量が増えすぎないようにするためにも、付属語を利用しなかった（付属語を利用した実験も行ったが概ね正解率や F 値に差はないことを確認している）。なお、形態素解析には JUMAN [6] を用いた。

5 実験と検討

機械学習である最大エントロピー法（以下、ME）とサポートベクトルマシン法（以下、SVM）による実験を行った。なお、ME には maxent を用いた。また、SVM には TinySVM の線形カーネルを利用し、ソフトマージンパラメータを 1 とした。

異なる意見の回答を抽出する実験を 10 分割クロスバリデーションにより行った。また、学習データとテストデータのカテゴリが異なる場合においても、異なる意見の回答が抽出できるかの実験を行った。表 3 に正解率と再現率、適合率、F 値を求めた実験結果を示す。なお、学習データとテストデータが同じカテゴリの場合（太枠で囲まれている部分）は、10 分割クロスバリデーションによる実験の結果を示している。10 分割クロスバリデーションによる実験では、3 つのカテゴリすべてで正解率がおおよそ 80% となった。このことから、投稿される質問と回答の内容が似ているカテゴリを学習データとテストデータに用いると、正解率がおおよそ 80% で意見の異なる回答を抽出できると考えられる。一方、学習データとテストデータのカテゴリが異なる場合の実験では、カテゴリが同じ場合と比べて正解率が低い。これは、カテゴリが異なると使われる形態素も異なるため、学習データとテストデータの両方に含まれ

る形態素が少ないからと考えられる。また、「パソコン、周辺機器」カテゴリを学習データ、異なるカテゴリをテストデータに用いた場合は再現率が低い。これは、3 章の表 1 で示したように「パソコン、周辺機器」カテゴリは異なる意見の回答が少ないためと考えられる。

本研究の実験で用いた素性のうち、異なる意見の回答であるデータ、または異なる意見の回答でないデータと判定するのに重要な素性の分析を行った。分析には、ME を用いた 10 分割クロスバリデーションによる実験で求められる値を利用して行った。具体的には、村田らの方法 [7] と同様の処理を行った。表 4 に判定するのに重要な素性の分析結果を示す。表 4 では、Dif は異なる意見の回答であるデータと判定するのに重要な素性、Sim は異なる意見の回答でないデータと判定するのに重要な素性を示している。なお、デフォルト素性とは、常に用いられる素性であり、判定するのに重要な素性が特になければこの素性が上位にくる。また、s15～s21 の素性は、回答 1 と回答 2 を入れ替えたデータも実験データとしているため、s8～s14 の素性と同じであるので結果から省いている。表 4 より、「健康、病気、ダイエット」カテゴリと「レシピ、調理方法」カテゴリの両方の上位に、s27:0 がある。これは、回答 1 と回答 2 の両方に含まれる自立語と未定義語が 0 個という素性である。この素性が上位にあることから、自立語と未定義語が 1 つも一致しない回答は意見が異なる場合が多いことがわかる。一方、「パソコン、周辺機器」カテゴリでは、s27:0 が上位にない。これは、「パソコン、周辺機器」カテゴリには他の 2 つのカテゴリに比べ、

- URL のみを書いて URL 以外の形態素がない回答
- 質問について不明な点を指摘している回答

が多いからである。これらの回答は、問題の解決方法についての質問に投稿されることが多く、同じ質問に投稿されている他の回答では、問題の解決方法についての意見が述べられている。本研究では、これらの回答は、自立語と未

表 4 判定するのに重要な素性の分析結果 (上位 3 つずつ)

カテゴリ	判定するのに重要な素性	
	Dif	Sim
健康、病気 ダイエット	s27:0	s26:できる
	s27:1	デフォルト素性
	s6:6	s13:2
パソコン、 周辺機器	s13:11	デフォルト素性
	s25:いち	s13:0
	s22:もの	s26:ない
レシビ、 調理法	s27:0	s26:食べる
	s25:とき	デフォルト素性
	s26:むく	s8:h t t p

定義語が 1 つも一致していなくても意見が異なる回答でないとしている。このため、「パソコン、周辺機器」カテゴリでは、s27:0 が上位になかったと考えられる。一方、異なる意見の回答でないデータと判定するのに重要な素性では、3 つカテゴリすべてにおいて上位にデフォルト素性がある。このことから、本研究で利用した素性のうち、異なる意見の回答でないデータと判定するのに特に重要な素性はなかったと考えられる。

6 今後の課題

今後の課題は以下の 2 つである。

- 異なる意見の回答のタイプを考慮した分類
- 新たな素性の考案

本研究では、異なる意見の回答は、4 つのタイプがあることを示した。しかし、これらを分類することはしていない。ただし、これらを分類することができれば、異なる意見の回答を 2 次利用ユーザが利用しやすくなると考えられる。例えば、異なる意見の回答が、「回答で仮定している条件が異なる」ということが知らされていれば、2 次利用ユーザの条件と一致する意見は重要であるが、条件と一致しない意見は重要でないと考えられる。また、「間違いの可能性がある」ということが知らされていれば、その質問で投稿されている回答は、正当性や信頼性を考慮して読むべきである。このように、異なる意見の回答のタイプを考慮した分類ができれば、2 次利用ユーザにとって Q&A サイトをより利用しやすくなると考えられる。このため、異なる意見の回答のタイプを考慮した分類を行うことを考えている。

また、本研究では、異なる意見の回答を抽出するのに機械学習で用いた素性は、質問とその回答に含まれる自立語と未定義語を利用して作成した。しかし、学習データとテ

ストデータのカテゴリが異なると正解率が低い結果となった。これは、カテゴリが異なると使われる形態素が異なるからである。また、素性の分析を行ったところ、意見の異なる回答でないデータと判定するのに特に重要な素性はなかった。これらのことから、形態素を利用した素性以外の素性も用いることで、さらに正解率や F 値が向上すると考えられる。このため、新たな素性を考案しようと考えている。特に、学習データとテストデータで使われる形態素が異なるカテゴリにおいても、判定するのに重要な素性を考案することが重要である。

7 おわりに

本研究では、Q&A サイトにおいて 1 つの質問に投稿された異なる意見の回答をユーザに知らせるため、どのような異なる意見の回答があるかを調査し、異なる意見の回答を抽出する方法について検討した。

本研究で提案した手法で抽出できる異なる意見の回答を 2 次利用ユーザに知らせることは、さまざまな意見があることを気づかせることに役立つ。そして、1 つの意見だけでは得られなかった情報から、問題を多角的に検討や評価することができるようになって考えている。また、本稿では、異なる意見の回答を 2 次利用ユーザに知らせることを目的として、実験や検討の結果を述べたが、異なる意見の回答でないものは、わざわざ読まなくてもよい可能性がある。このことから、2 次利用ユーザの負担を減らすのにも本研究で提案する手法は役立つと考えている。

参考文献

- [1] 乾, 奥村: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp.201-241, (2006).
- [2] 小林, 乾, 松本, 立石, 福島: 意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No.3, pp.203-222, (2005).
- [3] 峠, 大橋, 山本: 繰り返し学習を用いた話題に順応する意見文抽出, 情報処理学会研究報告, Vol.2004, No.119, pp.43-50, (2004).
- [4] 乾, 奥村: テキスト評価分析の技術とその応用, 情報処理, Vol.48, No.9, pp.995-1000, (2007).
- [5] 河原, 黒橋, 乾: 主要・対立表現の俯瞰的把握 - ウェブの情報信頼性分析に向けて: 情報処理学会研究報告, Vol.2008, No.67, pp.49-54, (2008).
- [6] 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).
- [7] 村田, 内元, 馬, 井佐原: 機械学習手法を用いた名詞句の指示性の推定: 言語処理学会, Vol.7, No.1, pp.31-50, (2000).