

文書ジャンルを横断した回答意見の検索のための分析

関 洋平[†]

豊橋技術科学大学[†]

seki@tut.jp

1. はじめに

Web は、時代に応じて利用形態を変化させつつ、公開される文書数は増加し、文書ジャンルは書き手の目的に応じて多様化の一途をたどっている。文書ジャンルには、不特定多数の読者に向けて事件を報道するニュース、興味を共有する読者に向けて書き手の考えや体験を表現するブログ、自分の情報要求に対する回答を求めて読み手と積極的にやり取りをするコミュニティ QA や、その場の体験を一方向的に公開するマイクロブログがあり、目的に応じて文章の書き方、表現は異なり、出現する意見の傾向も異なる。

一方、ユーザが Web から情報を得ようとする方法や情報要求も多様化している。近年では、多くのユーザから積極的に情報を収集する集合知を利用して情報を得る方法が盛んであり、Wikipedia やコミュニティ QA がその代表と考えられる。コミュニティ QA は、分野を特化することで、その分野に特に興味がある人あるいは専門家に向けて自由に質問することができ、ユーザの情報要求として、専門家から得られる回答情報を検索するタイプの質問や、多数の回答者からアンケートのような社会調査目的で回答を収集するタイプの質問がある[1]。また、方法を問う情報検索型の質問の別解を検索するために、質問語と Web 文書中の単語との共起頻度を利用してクエリ拡張を行い、回答を補完する研究が試みられている[2]。しかし、意見を問う質問に対しては、別のアプローチが必要と考える。

本研究では、意見を問う社会調査型の質問と、理由を調査する情報検索型質問に焦点を当て、個別の意見を多く含むブログから回答意見を収集するための検索手法について分析を行い、考察する。特に、回答意見の、ドメイン (カテゴリ) や質問タイプとの関係について調査した。

2. 意見を問う質問の収集と回答意見の傾向

本節では、Yahoo!知恵袋に出現する意見を問う質問の選択手順を述べた後、回答として現れる意見のタイプについて議論する。

2.1 意見を問う質問の収集

まず、意見を問う質問の傾向を明らかにするために、Yahoo!質問検索API¹を利用し、質問の収集を行った。質問は、利用するブログデータ (ICWSM 2009 Spinn3r Blog Dataset²) の収集時期に合わせて、2008 年の 8 月または 9 月に起こった出来事 (ロシアグルジア侵攻、福田首相辞任、リーマンショック、中国食品汚染 (天洋食品、粉ミルク問題)、北京五輪 (北島、ボルト)) を検索キーワードとし、2008 年 8 月または 9 月に書かれた質問に限定して収集した。質問の絞り込みは、以下の手順で行った。

1. 質問を形態素解析器 mecab を用いて内容語 (固有・普通・サ変・普通名詞) を抽出し、検索システム lemur を利用し、ブログデータに適合文書がある質問を選択した。
2. 質問の Yahoo!知恵袋における回答数が、2 以下のものを削除した。
3. 1 つの話題に対して、複数の質問が現れている場合、その時期の話題性が高いと考えて、選択した。
4. 残った質問のうち、意味のある内容を問わないものや、情報検索型の質問を削除した。

以上の戦略により、多くの回答を必要として、ブログ中でも話題になっており、様々な質問が出るような話題性の高い質問として、7 つの話題についての 16 の意見を問う質問を収集した。

2.2 Yahoo!知恵袋の回答に出現する意見タイプ

Yahoo!知恵袋に出現する回答 (ベストアンサー) について、意見のタイプの出現傾向を調査した。

Analysis for Cross-Genre Opinion Retrieval

[†]Yohei Seki, Toyohashi University of Technology

¹

<http://developer.yahoo.co.jp/webapi/chiebukuro/chiebukuro/v1/questionsearch.html>

² <http://www.icwsm.org/2009/data/index.shtml>

まず, BCCWJ2009 コーパス¹から, 頻出する 7 つのカテゴリである, “ヤフオク”, “ヘルスケア”, “インターネット”, “恋愛相談”, “PC”, “政治”, “テレビ番組”に含まれる 251 の文書を対象として, 意見タイプとして, 自発的な感情の表明 (affect), 人間の振る舞いの判断や批評 (judgment), 事物や事象に対する評価 (appreciation) を人手で判定した結果を表 1 に示す. なお, 判定は, 重複判定や 3 つのタイプ以外の意見とすることを認めている. 判定の精度として, 複数の判定者間の一致率 (κ 係数) は, 新聞で 0.7~0.9, 知恵袋で 0.4~0.5 である.

表 1 Yahoo!知恵袋の代表的 7 カテゴリのベストアンサーに出現する意見タイプの分布

ID	カテゴリ	質問数	意見数	感情	批評	評価
1001	Yahoo!オークション	58	148	13.51	12.84	37.16
1002	病気・症状・ヘルスケア	21	92	22.83	5.43	31.52
1003	インターネット	24	21	14.29	14.29	42.86
1004	恋愛相談・人間関係の悩み	52	233	21.03	17.6	30.9
1005	パソコン・周辺機器	48	88	12.5	5.68	55.68
1006	政治・社会問題	25	75	14.67	4.8	33.33
1007	テレビ・ラジオ	23	35	11.43	20	62.86

上記結果から, 回答の意見タイプにはカテゴリに依存性があり, ヘルスケア, 恋愛相談では感情を表すような回答意見が多く, 政治問題であれば批評, インターネット, PC, テレビ番組では, 評価を表す回答意見が多いことがわかった.

2.3 意見を問う質問に対する回答意見タイプ

次に, 2.1 節で収集した意見を問う質問に出現する回答意見についての判定結果を表 2 に示す. なお, 表 1 の結果とは異なり, 判定対象は, 質問に対する回答になっている意見に限定した.

表 2 意見を問う質問の回答意見タイプの分布

トピック	質問	回答意見数	感情	批評	評価
グルジア 政治	ロシアとグルジアは、なぜ戦闘しているのですか？悪いのはどっちの国ですか？	2	0	100	0
	ロシア軍機、グルジア爆撃を開始しました。本格戦闘突入し戦争になる可能性が非常に高くなると思いますが、今後どのような展開になると思われますか？	2	0	0	100
福田首相 政治	福田首相の辞任には反対ですか？それとも、賛成ですか？理由をつけてお願いします。	4	0	50	25
	福田首相が辞任するらしいですが次の総理大臣は誰ですか？福田首相はなぜ総理を辞任したんですか？	4	25	50	0
リーマン ショック 経済	リーマンブラザーズはなぜ破綻したんですか？	1	0	0	0
	リーマンが破綻した事で、日本経済にはどんな影響があるのですか？絶対に無傷ではないですよね…	4	25	0	75
中国粉ミルク 事件	リーマン破綻、何故騒がれるの？米大手証券であるリーマンの破綻、これが就職氷河期の再来、又は世界恐慌の予兆とまで言う人がいます。これは本当なのですか？	5	20	0	100
	中国の粉ミルクにメラミンが混入していた件で、世界各国がすでに中国産の乳製品の輸入禁止措置に踏み切っています。EU、インド、そして韓国でも日本は何にも無し…般決意は？また中国で最悪な事件ですか、有害物質メラミンで汚染された粉ミルクを飲んだ乳児が腎臓結石になった事件が…	1	100	0	0
天洋食品 事件	回収された天洋餃子を食べた中国の方が中毒にかかったと発表されましたが、これに対して日本側は抗議をしないのでしょうか？	6	50	0	33.33
	天洋冷凍餃子の中毒者が中国国内でも発生していたようですが、中国政府はそれでも凍餃子は日本国内で混入されたと言っているんですかね？	5	20	80	0
北島ス ポーツ	金メダル 北島選手についてこの人は確かに、すごいと思いますよ。…しかし、関が出来てないと思います…みなさん、彼のこと、どう思いますか？	1	0	100	33.33
	競泳の北島選手が、北京大会限りで引退することを明らかにしました。私としては、引退するのは惜しいと思っています。皆さんの考えをお聞かせください。	3	0	100	0
ウサイン・ ボルトス ポーツ	北京オリンピックの男子200メートル決勝でまたもやウサイン・ボルトがやってくれましたね！…予想はしていましたが、やっぱり見てみると衝撃は隠せません。みなさんはどうですか？	1	0	100	0
	北京五輪陸上男子100mを制するのほどの選手でしょうか？ウサインボルト・アサファバウエル・タインティの3種の争いになると予想します。	1	0	0	0

¹ <http://www.tokuteicorpus.jp/>

この結果から, 人間の振る舞いの批評が比較的多いが, 質問ごとの揺れも大きいとわかった.

3. ブログからの回答意見検索

前節では, 質問に対する回答の意見タイプには, カテゴリ依存性があることと, 個別の質問に応じて偏りがあることを見た. 次に, 意見を問う質問を分類し, ブログ中に現れる回答となりうる意見について簡単な調査を行った.

3.1 調査方法

表 2 に提示した質問を, “見通しの提供”, “賛否態度収集”, “感想収集”, “理由・原因の推測”, “説明提供”, “議論”の 6 つに分類し, ブログ中の回答意見について, Appraisal Theory[3]に基づく詳細な意見タイプの出現傾向を調査した.

3.2 質問タイプごとの回答意見の出現傾向

- 賛否態度: “切望・敬遠”などの感情意見が多く, 人については, “満足・不満”, 戦争については, “調和・混乱”や“明瞭・複雑”などの評価意見が見られた.
- 見通し: “安全・危険”などの感情意見と評価意見が見られた. 評価意見の詳細は, トピックが戦争か人かで揺れが見られた.
- 理由推測: リーマンショックや福田首相辞任といったトピックへの依存性から, “調和・混乱”などの評価意見が見られた.

4. おわりに

本論文では, Yahoo!知恵袋に現れる意見を問う質問の, ブログからの回答意見の検索について検討を行い, カテゴリや質問タイプと回答意見との関係について簡単な分析を行った.

今後は, 意見タイプを利用した回答意見検索システムの自動化と評価を進める予定である.

謝辞

この研究の一部は, 科学研究費補助金若手研究 (B) (課題番号 21700268), 特定領域研究 (課題番号 21011003), 基盤研究 B (課題番号 21300029) を受けて遂行された.

参考文献

- [1] 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 FI-95/DBS-148, pp.1-8 (2009)
- [2] 高田夏希, 山本祐輔, 小山聡, 田中克己: 質問応答コンテンツに対する Web からの別解情報検索, 第 2 回知識共有コミュニティワークショップ, pp.53-62 (2009).
- [3] Martin, J.R. and White, P. R. R.: *The Language of Evaluation: Appraisal in English*, Palgrave Macmillan (2005).