

単語の類似度と感情表現を考慮した質問記事の判定手法

輪島 幸治[†] 佐藤 哲司[‡]

[†]筑波大学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

[‡]筑波大学図書館情報メディア系, 305-8550 茨城県つくば市春日 1-2

E-mail: [†]kwajima@ce.slis.tsukuba.ac.jp, [‡]satoh@ce.slis.tsukuba.ac.jp

あらまし 企業における顧客管理においては、顧客の窓口となる問合せ管理が重要とされており、重要度が高い問合せは早急な対応が求められている。しかし質問記事の重要度判別は、回答者の主観的な評価に依存しており、日々大量に寄せられる問合せから重要である質問記事を判別するにはコストがかかる。そこで本研究では質問記事内に含まれる顧客の要望と感情表現に着目し、質問記事を重要度の高い順に順位付けする手法を提案する。Word2Vec を用いて抽出した評価視点と、質問対象、感情表現で質問記事の主題が表現されているとし、それらの属性を用いて記事の重要度を算出する。オンラインサポートコミュニティに投稿された質問記事を対象に、提案手法を適用評価し、有効性を確認する。

キーワード 感情表現, LDA, Word2Vec

1. はじめに

インターネットインフラの普及やスマートデバイス利用者増加に伴い、企業に対する顧客の問い合わせ方法は、従来の電話を主体としたものから、メールやWEBフォーム、オンラインヘルプフォーラムなどテキストを通じて行われる機会が増加している。

顧客からの問い合わせテキスト（以後、質問記事）は多種多様であり、製品に対する使い方などの一般的な質問以外にも、提供サービスに対する不満やクレームなど、早急に対応しなければならない質問記事が混在している。例えば「パソコンのOSをバージョンアップしたいので、やり方を教えて欲しい。」のような質問記事がある一方で、「スマホの表面ガラスが割れてしまった。ひどい欠陥だ。二度と買わない。」「楽曲ダウンロードサービスを利用しているが、ダウンロードに時間がかかりすぎである。全く使えないので解約したい」などクレームのような質問記事も存在する。

後者のような質問記事への対応が遅れると、その顧客が離れていくだけでなく、Twitterなどのコミュニティメディアへの投稿によって悪評が蔓延するリスクへと繋がっていく。そのため企業には日々膨大に寄せられる質問記事の中から、迅速な対応を必要とする重要な質問記事を検出し、センシティブな対応をすることが求められている。しかしながら重要度が異なる大量の質問記事から、早急に対応すべき記事を人手で判別するには多大な負担がかかる。

そこで本論文では、製品やサービスに対して寄せられる大量の質問記事に対して、「記事中に表出される”質問対象”」、「評価対象」を同定し、「感情表現」を手がかりに質問記事の重要度を判定する手法を提案する。こ

こで、「パソコン」や「スマホ」「楽曲ダウンロード」などの製品やサービスのことを『質問対象』といい、質問対象の中で特定している箇所「OSバージョンアップ」「表面ガラス」「ダウンロード」のことを『評価対象』という。最後に、「教えて欲しい」「ひどい欠陥」「時間がかかりすぎ」など、質問者の感情が表出している箇所を『感情表現』という。

提案法では、教師無し学習手法として知られるLDAを用いて質問対象を抽出する。質問が言及している質問対象の部分、すなわち評価視点は、質問対象に対して言及する単語を抽出するため、類似性の高い単語を抽出できるWord2Vecを用いて実装する。最後に、評価表現辞書を用いて質問記事中の質問者の感情を感情表現として定量化する。

以下、本論文の構成を示す。2章では関連研究を概観し本論文の位置づけを明らかにする。3章では、質問記事の重要度を判別する提案手法を詳述する。4章では、提案法の実装と評価実験を、5章では実験の結果と考察を述べる。最後に6章でまとめる。

2. 関連研究

本研究は、大量な質問記事を対象として、個々の質問記事が言及している質問対象と評価視点を抽出する処理と、抽出した質問対象と評価視点の組に対して、質問者の感情である感情表現を抽出する処理からなる。以下、質問記事を対象とする研究と感情表現の抽出に関する研究に分けて、先行研究を概観する。

質問記事に関する研究には、質問記事の評価に関する研究[1]、質問と回答のマッチングに関する研究[2][3]などがある。文献[1]は、品詞の比率、単語心像

性、文末表現などの質問記事の特徴量に着目し、質問回答サイトにおけるベストアンサー推定のための因子得点の推定手法を提案している。文献[2]は、Why型質問と呼ばれる「なぜ～ですか?」のような文体の回答に、回答検索による回答の絞込と教師あり機械学習手法を用いた回答のリランキングの2ステップで、最適な回答を提示する手法を提案している。文献[3]は、質問記事のセマンティックフレームのセットを取得後、質問の意味グラフを比較し、類似の文書を検索する手法を提案している。

感情表現に関する既存の研究には、自然言語処理的アプローチと心理学的なアプローチを用いるものがある。自然言語処理によるアプローチの代表的な研究として、単語を肯定と否定という2値極性の評価極性に抽出する研究が知られている[5]。Turneyら[4]は、コーパスから共起情報を元に抽出している。高村ら[6]は、単語の極性を電子スピントとみなし、語彙ネットワークをモデル化したものから抽出している。また小林ら[7]は、2値極性の情報に加え、客観的・主観的といった情報を単語にタグ付けしている。なお、評価極性を用いた文書分類では、Turney[4]らはコーパスから得られた評価極性の値を元に文書の極性の平均値を算出し、平均値の符号から文書を分類する方法やPangら[8]のように肯定極性文比率と呼ばれる肯定極性文の比率による類似度関数を用いる方法などがある。

本研究は、文献[1]に類する質問記事の特徴量に着目した研究である。だが人間が重要と判断すべき質問記事を抽出すべきという目的から、質問対象ごとの評価に言及する箇所の同定と質問記事中の感情表現の抽出に注力している。またオンラインコミュニティ上のデータは口語やスラングを含むことが多く、必ずしも正しい文体で構成されているとは限らないため、文献[2][3]に示されている構文構造を手がかりとする手法の適用は難しいと考えている。

3. 提案手法

3.1. 概要

本研究では、質問記事の核となる3要素を抽出することを目的としている。一般的に質問記事は、図1に示すように、質問が言及している製品やサービスを表す『質問対象』、質問対象における評価の視点となる『評価視点』、および質問者が感じている『感情表現』の3要素からなる。これら3要素は、質問記事中でその出現位置や順序に規則性はなく、図1に示すような限られた構文パターンでは十分な精度で抽出することはできない。このため、本論文では、これら3要素のそれぞれについて、機械学習の手法を適用することで人手に頼らずに各要素を抽出する手法を提案する。

(1) (質問対象) (助詞) (評価視点) {が/は/も/に/を} (感情表現)
例: iPhone の アップデート が 煩わしい

(2) (質問対象) (評価視点) {が/は/も/に/を} (感情表現)
例: Wi-Fi 設定 が 難しい

図 1 典型的な質問記事の例

- 質問対象
質問対象とは、質問記事が言及している製品やサービスなど、問い合わせの対象を識別する要素である。本論文では、予め質問対象を特定しないことから、教師なし機械学習として知られているLDA (Latent Dirichlet Allocation) [9]を用いて文書群より得られる潜在的なトピックを質問対象とする。その際、LDAによって得られるトピック分布に基づいて、最も生成確率が高いトピックを質問対象とする。
- 評価視点
評価視点とは、質問対象に対して質問者が要望を示している要素である。LDAで得られたトピック毎の単語分布を基に、Word2Vec[10]を用いて類似性が高い語を抽出し、各質問対象の評価視点とする。このため、評価視点は、Word2Vecで得られる単語が、質問記事中に生成されているかによって抽出する。
- 感情表現
感情表現とは、質問対象や評価視点に対して質問者が表出する感情を表す要素である。本論文では、質問記事内に含まれる形容詞を感情表現の候補とする。感情表現の定量化は、既存の評価表現辞書で日本語に対応した単語感情極性対応表[6]を用いて行うことから、辞書に含まれる単語が感情表現の候補となる。

3.2. 質問対象の抽出方法

質問対象の候補は、企業が提供している製品やサービスである。このため、自社で提供している製品やサービスなどを質問対象とすることもできるが、一般に数多く提供している製品やサービスを網羅的に列挙することは難しく、またそれらの全てが質問対象となることもない。また、同一の製品であっても、その呼称、すなわち質問記事中で言及される質問対象としてのラベルが複数存在することもある。さらに競合する他社製品の登場など、時間的な経緯で質問対象となる製品やサービスが大きく変化することも多い。

筆者らは、このように変化する質問対象を、投稿されている質問記事から動的に抽出する手法として、教師無し学習手法 LDA [9]を適用することを検討している[15]. LDA は潜在的なトピックの存在を仮定した確率モデルであり、質問記事が言及している製品やサービスをトピックに、LDA で算出されるトピックの高確率なタームを質問記事中のその製品に対する呼称（ラベル）に対応させる。これにより、質問記事集合を LDA で学習しトピックを抽出することで、質問記事集合全体が言及している質問対象の集合を得ることができる。

優れる	: すぐれる	: 動詞	: 1
良い	: よい	: 形容詞	: 0.999995
喜ぶ	: よろこぶ	: 動詞	: 0.999979
褒める	: ほめる	: 動詞	: 0.999979
めでたい	: めでたい	: 形容詞	: 0.999645
(～中略～)			
ない	: ない	: 助動詞	: -0.999997
酷い	: ひどい	: 形容詞	: -0.999997
病気	: ひょうぎ	: 名詞	: -0.999998
死ぬ	: しぬ	: 動詞	: -0.999999
悪い	: わるい	: 形容詞	: -1

図 2 評価表現辞書[6]

3.3. 評価視点の抽出方法

本論文では、評価視点の抽出手法に word2vec を適用することを検討する。Mikolov らによる word2vec は、CBOW と skip-gram と呼ばれるニューラルネットワークのモデル利用した手法である。word2vec では、文脈情報から単語の特徴をベクトルで表現し、特定の単語との類似度や単語同士の関連性をベクトルの加減算によって算出できる。

評価視点は、質問記事中で質問対象と高い関連性があることから、word2vec によって評価視点を抽出できると期待される。そこで、LDA で求めた各トピックの高確率ターム、すなわち質問対象の候補語との関連性が高いタームを word2vec によって求め、評価視点の候補とする。

3.4. 感情表現の抽出方法

感情表現は、質問記事中に質問者が記述した不満や要求の表れである。ここでは、個々の質問記事に出現する単語の語感を定量化するために評価極性値を用いる方法を提案する。評価極性値とは、記事中の単語が一般的に良い印象を持つか(肯定)、悪い印象を持つか(否定)を表した属性を数値として定量化されたものである。例えば、「優れる」、「めでたい」などは肯定的な極性、「悪い」、「酷い」などは否定的な極性を持つ単語である。

評価極性値が定義された辞書は、一般的に評価表現辞書と呼ばれる。本研究では、高村ら[6]の単語感情極性対応表の極性値を用いることとした。この辞書には、図 2 に示すように、各用語に対して品詞属性だけでなく、極性属性として肯定(値 1)から否定(値-1)まで付与されている。

4. 評価実験

4.1. 実験環境及び評価対象データ

提案法の有効性を確認するために、CentOS release 7 上に Python 2.7.5 を用いて実装と評価を行った。その際、LDA および word2vec アルゴリズムの実装は、Řehůřek ら[11]によって開発された Gensim を用いて行った。

一方、評価に使用するデータは、Apple Inc.が提供している 2008 年 10 月 1 日から 2014 年 1 月 24 日までの Apple サポートコミュニティ[12]の質問文書 10,391 件を使用した。評価に先立って、以下のデータクレンジングを行った。

- 正規表現による記号の除去
- Unicode 正規化による単語の標準化
- 1 文字の単語を除去
- 単語の活用形を標準形に変換

4.2. 正解データの作成および比較方法

質問記事の重要度を判定する本研究では、業務に精通した人手による判定に近い結果をシステムが出すことが目的となる。そのため、サポート業務の経験を有する 20 代～40 代の者 8 名をアノテータとして、正解となるラベル付け作業を行った。

ラベル付け作業は、第一著者が選定した 10 件の質問記事に対して、アノテータの主観で最も重要（早急に対応すべき）から最も重要でないまでの 10 段階で順位付けを行った。その後、正規化順位法によってアノテータの結果を正規化し相互比較を行った。スコアの算出方法は、集計した値に各順位の 2 乗数を乗算し、すべての順位の値を加算してスコア値とした。

$$\text{Sensibility Score} = \sum_{R=1}^R R^2 * \text{Count}$$

4.3. 質問記事の重要度算出方法

3章に述べた手法で抽出した質問対象，評価視点，感情表現を用いて，質問記事の重要度を算出する．ここで，質問対象は質問記事のカテゴリ化のための要素であることから，重要度算出の基準には含めないものとする．

質問記事の重要度を算出するにあたり，評価視点と感情表現は独立した要素であり，二つの異なる軸によって質問記事の重要度が決定されるものとする．評価軸の一つとなる評価視点は，Word2Vecによって算出された類似度の値の総和で算出されるものとする．一方，感情表現のスコアは，質問記事中に出現する形容詞の評価値を評価表現辞書から求め，それらの総和値，平均値，中央値，最頻値をそれぞれ算出することとした．

提案法の優劣は，人手で作成した正解データとの順位相関係数によって評価する．質問記事の順位付けは，上述のスコアを用いて行い，スピアマンの順位相関係数 ρ を用いて比較評価を行う．

5. 実験結果

5.1. 質問対象の抽出および分布

実験に使用する Apple サポートコミュニティのデータに，トピック数を 40 として LDA を適用した結果を図 3 に示す．図の縦軸は，各トピックに出現する高確率タームが出現する質問記事の個数である．図の結果から，最も多いトピックの割当数は 1,745 件であった．この結果から，評価視点と感情表現の抽出実験は，この割当数が最も大きなトピックを対象として行うこととした．

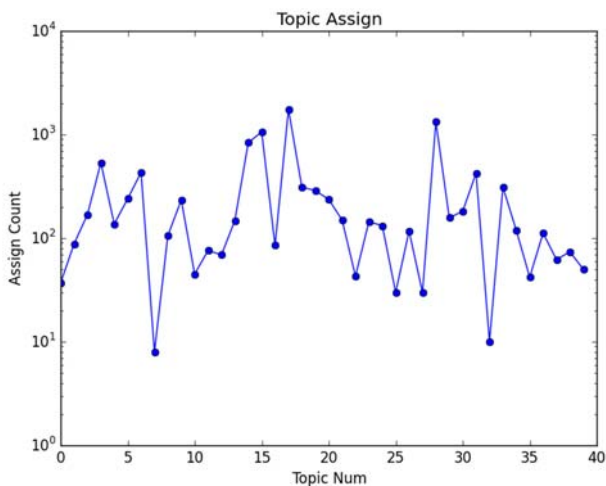


図 3 質問記事群へのトピックの割当

5.2. 評価視点の抽出と評価

前項で抽出した質問対象の単語分布から，評価視点となる単語の集合を抽出する．対象とする各質問記事に対して評価視点を割り当てることで，各質問記事の

評価視点に関するスコアを算出する．記事毎のスコアは，割り当てられた評価視点のスコアの総和とした．

評価対象とする質問記事は，人手によって正解データをラベル付けした 10 件を使用した．質問記事毎に算出した評価視点のスコアを図 4 に示す．図の横軸は質問記事であり，人手によって抽出した結果を Sensibility Score の図に，先行研究 [13] において KeyGraph と呼ばれる重要語抽出手法を用いた場合を KeyGraph Score の図に，本論文で提案する word2vec を用いた場合を Word2Vec Score の図に示している．この図において KeyGraph による評価では最も高く評価すべき質問記事（記事 ID=8）のスコアが 3 位，4 位と同値程度となっていたのに対し，提案手法では最も高い値となっており，2 位の質問記事とともに適切に上位記事としてスコア化され，人手による評価と類似したグラフの形状となっている．

得られたスコアで質問記事を順位付けした結果を表 3 に示す．この表は，人手によって付与された順位を基準（Correct Ranking）として，KeyGraph および Word2Vec によって得られた順位を示している．例えば，

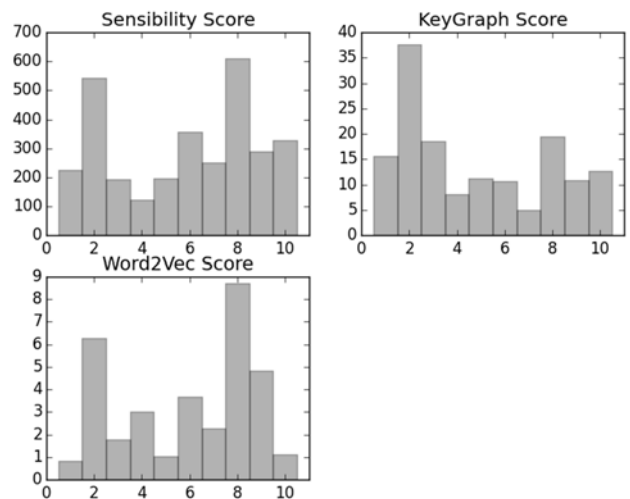


図 4 評価視点のスコア

表 3 評価視点によるランキング

Correct Ranking	Ranking with Review Points	
	KeyGraph	Word2Vec
1	2	1
2	1	2
3	8	4
4	5	6
5	7	10
6	10	5
7	4	8
8	6	3
9	3	9
10	9	7

人手で第1位とした質問記事を、KeyGraphは第2位としWord2Vecは第1位としたことを示している。また、人手で10位と判定された質問記事は、KeyGraphで第6位に、Word2Vecで第5位にランキングされており、自動化が難しい質問記事であることが分かる。

提案手法(Word2Vec)とKeyGraphとの順位比較では、ともに1位、2位の上位文書は上位の質問記事として評価されているが、KeyGraphは第3位に人手判定で8位の記事をランキングしている。これに対して提案手法では4位、5位、6位の質問記事が中位文書として適切に評価されており、KeyGraphによる評価視点の抽出よりも提案手法による抽出方法に優位性があることが分かる。

得られた評価視点に基づくランキング(表3)の結果から順位相関係数を算出した結果を表4に示す。Word2Vecを用いて評価視点を抽出する提案手法は、人手による順位と類似した順位となっており、相関係数 $\rho = 0.636$ であった。検定の結果は $P < 0.05$ で有意差があると判断できる。

表4 評価視点に基づく順位相関係数

	ρ	P
KeyGraph	0.406	n.s.
Word2Vec	0.636	$0.01 < P < 0.05$

以上の結果から、評価視点にKeyGraphを用いる方法よりもWord2Vecを用いた提案手法が優れており、また正解データとも有意に相関があることが明らかとなった。

5.3. 感情表現の抽出と評価

評価視点と同様に10個の質問記事を対象として、感情表現に対してもランキング付けを行った。筆者らの先行研究である文献[14]によって、該当の文書群のほぼ全てが負の感情表現であることが明らかとなっている。これは、サポートセンターに寄せられる質問記事の多くが、何らかの不具合を解決することを目的として投稿されていることに起因している。そこで、本論文の感情表現のスコア評価では、抽出される感情表現のうちポジティブな単語の有無によって順位の結果が変化するかを確認する。スコアの算出方法に際しては、各質問記事に出現する感情表現の単語スコアの値の平均値、総和、中央値、最頻値を用いて比較した。結果を表5および表6に示す。

表5 感情表現における順位相関係数

Rank	With Positive Words		With Positive Words	
	ρ	P	ρ	P
Sum	0.563	< 0.05	0.563	< 0.05
Mean	0.418	n.s.	0.260	n.s.
Median	0.115	n.s.	0.309	n.s.
Mode	0.406	n.s.	0.406	n.s.

表6 感情表現におけるランキング

Correct Ranking	With Positive Words				With Positive Words			
	Sum	Mean	Median	Mode	Sum	Mean	Median	Mode
1	2	1	4	3	2	4	4	3
2	1	4	2	2	1	3	2	2
3	8	9	9	4	8	6	5	4
4	3	2	3	5	3	2	3	5
5	5	5	5	6	4	5	6	6
6	10	10	10	8	10	10	10	8
7	4	3	6	9	6	7	7	9
8	9	8	7	7	9	9	9	7
9	6	7	8	10	5	8	8	10
10	7	6	1	1	7	1	1	

表5に示す順位相関係数 ρ の値を見る限り、ポジティブな単語の有無による感情表現の順位に大きな差異は見られなかった。また、検定の結果、正解データとの相関関係が確認できたのは総和のみであった。

このことから感情表現のスコア値の算出方法は、ポジティブな単語を含む方法で、かつ、割当られた感情表現の単語のスコア値の総和が適しているといえる。

6. 評価視点と感情表現による評価マップ

本論文では、質問記事の重要度は評価視点と感情表現の二軸で判定できるとして、5.2節および5.3節でそれぞれ独立に評価をしてきた。ここでは、横軸を評価視点の軸とし縦軸を感情表現の軸として二軸を同時に評価する評価マップを作成する。

評価視点の抽出にKeyGraphを用いた場合の評価マップを図5に、Word2Vecを用いた場合を図6に示す。図中のROはRanking by Operatorの略で、アノテータ

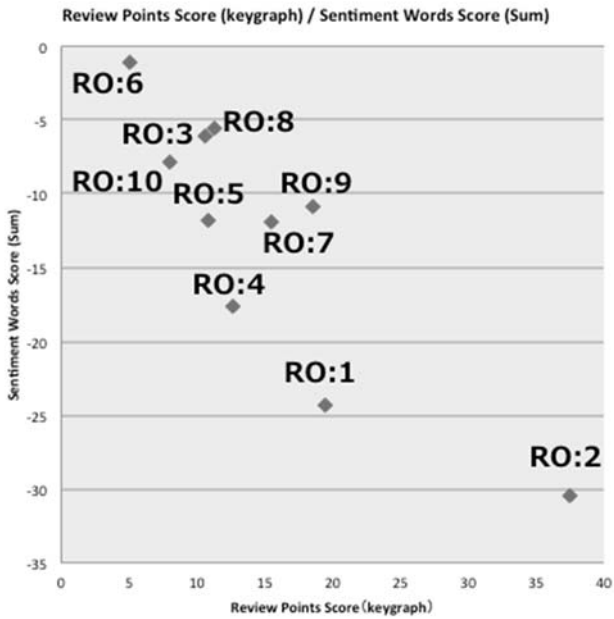


図 5 Review Points(key graph) + Sentiment Words

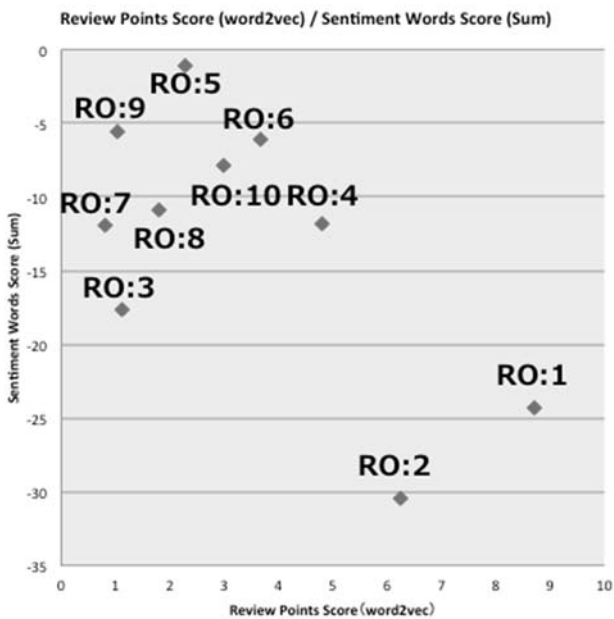


図 6 Review Points(word2vec) + Sentiment Words

によって付与された正解ランキング値である。

質問記事として重要度が高い RO:1 および RO:2 の記事は評価視点に KeyGraph を用いた図 5 でも右下にプロットされているものの、評価視点に Word2Vec を用いた図 6 の方が重要度の高い記事とそれ以外の記事とを明確に分離していることが分かる。第 3 位(RO:3)以下の質問記事に関しては、Word2Vec の方がやや塊の周縁部に上位の記事 (RO:3, RO:4, RO:5) が配置されている様に思われるが、これらの記事に対する順位付けは今後の課題である。

以上のことから、重要度が高い質問記事を質問記事集合から抽出する課題に対して、評価視点に Word2Vec を用いる提案手法が KeyGraph を用いるよりも優れているといえる。

7. まとめ

質問記事を重要度の高い順に順位付けするために、質問対象・評価視点・感情表現を用いて評価する手法を提案した。その際に、評価視点に word2vec を用いることで、順位相関係数で有意性がある結果を得ることができた。また感情表現に関しても、スコア値の総和を用いることで、順位相関係数で有意性のある結果となった。また評価視点と感情表現を 2 軸としてプロットした評価マップにて評価した際、質問記事集合から重要な質問記事を抽出する際に有意な結果となった。

質問対象と評価視点の組に対して、強いネガティブな感情を持つ質問記事ほど緊急性を要する重要な質問記事として評価を行う。提案手法を用いて質問記事を重要度に応じて評価することで、大量に寄せられる質問記事を適切な順序で対応することが期待できる。

謝辞

本研究は、JSPS 科研費 25280110 の助成を受けたものです。ここに記して謝意を示します。

参考文献

- [1] 横山 友也,宝珍 輝尚,野宮 浩揮,佐藤 哲司,“文章の特徴量を用いた質問回答文の印象の因子得点の推定”,日本感性工学会論文誌 12(1), 15-24, 2013.
- [2] 呉 鍾勲,鳥澤 健太郎,橋本 力,川田 拓也,デサーガ ステイン,風間 淳一,王 軼謳,“意味的極性と単語クラスを用いた Why 型質問応答の改善”,情報処理学会論文誌 54(7), 1951-1966, 2013.
- [3] Minoru Harada, Yuhei Kato, Kazuaki Takehara, Masatsuna Kawamata, Kazunori Sugimura, and Junichi Kawaguchi, “QA System Metis Based on Semantic Graph Matching”, Proc. of the 6th International Conference on NII Test Collection for IR Systems(NTCIR6), Tokyo, Japan, pp.448-459, 2007.
- [4] Turney, Peter D., “Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews”, ACL '02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417-424, 2002.
- [5] 乾 孝司,奥村 学,“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 1340-7619 The Association for Natural Language Processing,13-3, pp. 201-241, 2006.
- [6] 高村大也,乾孝司,奥村学,“スピンモデルによる単語の感情極性抽出”,情報処理学会論文誌ジャーナル, Vol.47 No.02 pp. 627-637, 2006.
- [7] 小林のぞみ,乾健太郎,松本裕治,立石健二,福島俊一. 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.

- [8] Pang, B. and Lee, L., “Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales.” In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005), 2005.
- [9] Blei et al, “Latent Dirichlet Allocation,” Journal of Machine Learning Research, Vol.3, pp.993-1022, 2003.
- [10] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, "Efficient Estimation of Word Representations in Vector Space," CoRR, abs/1301.3781, 2013.
- [11] Radim Řehůřek and Petr Sojka, “Software Framework for Topic Modelling with Large Corpora,” Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp. 45-50, 2010.
- [12] アップルジャパン合同会社, “Apple サポートコミュニティ”, <https://discussionsjapan.apple.com/>, 最終閲覧日: 2014/1/5.
- [13] 大澤 幸生, ベンソン ネルス E, 谷内田 正彦, “KeyGraph: 語の共起グラフの分割・統合によるキーワード抽出”, 電子情報通信学会論文誌. D-I, 情報・システム, I-コンピュータ J82-D-I(2), 391-400, 1999.
- [14] 輪島 幸治, 小河 誠巳, 古川 利博, 嶋田 茂, “潜在的ディリクレ配分法を用いたネガティブ要因分析”, 電子情報通信学会 他共催 第6回データ工学と情報マネジメントに関するフォーラム DEIM2014 論文集, A9-3, 2014.
- [15] Wajima, K. and Satoh T., “Urgent Question Detection based on the Review Points and Sentiment Words,” The 17th International Conference on Information Integration and Web-based Applications & Services, iiWAS2015, pp. 307 – 311, 2015.