

質問回答サイトにおける回答文の客観的評価の推定

Estimation of Objective Assessments towards Answer Statements at Q&A Sites

横山友也† Yuya Yokoyama
 宝珍輝尚† Teruhisa Hochin
 野宮浩揮† Hiroki Nomiya

1. はじめに

インターネット上における質問回答サイトの利用者が近年急増している[1]. 質問回答サイトとは、インターネット上でユーザ同士が互いに質問と回答を投稿しあうコミュニティの一形態であり、様々な悩み事・相談事を解決する場であると同時に、膨大な知識が蓄積されたデータベースとして活用されるようになってきている. あるユーザが質問を投稿すると、他のユーザがその質問に対して回答を投稿する. 質問者は、質問文に対して最も適切と判断した回答文を「ベストアンサー」に選定し、その回答を行った回答者に謝礼として手持ちのポイントを贈与する. ここで、「ベストアンサー」とは、質問文に対する満足度が最も高いと質問者が主観的に判断した回答文である.

質問回答サイトの参加者が増え、また、投稿される質問数が膨大になると、回答者が自身の専門性や興味に合った適切な質問文を探し出すことが困難になるという問題が顕在化してくる. あるユーザが質問文を投稿しても、その質問文が必ずしも適切な回答者の目に留まり、回答を得られるわけではないという問題である. また、適切な回答者に巡り会えないミスマッチから、質問者にも不利益も生じる. つまり、質問回答サイトの課題は、日々投稿され続けている幾多の質問と、様々な興味・関心や専門性を有する回答者とを適切にマッチングすることであるが、質問者や回答者の努力に任せているのが現状である. そこで、ある質問文に適切な回答ができるユーザをその質問文に引き合わせる必要がある.

これまでの研究において、筆者らは、質問者に適切な回答者を引き合わせるために、質問者と回答者の相性を判断する手段として質問者と回答者の文章の印象評価を行ってきた[2]. Yahoo!知恵袋[1]に投稿された質問文と回答文の計 60 個の文章に対して、50 個の印象語を使用して、印象評価実験を行った. 得られた評価値に対して因子分析を行った結果、文章内容に関する因子が 9 個得られた[2].

しかし、ここで得られた因子得点は、評価実験を行った質問文と回答文の文章 60 個に対するもののみであって、他の多数の質問文と回答文に対する因子得点は得られていない. そこで、どのような質問文と回答文に対しても因子得点の推定を可能にすることを目的として、文章の特徴量からの文章の因子得点の推定を重回帰分析により試みてきた[2]. ここでは、名詞や動詞といった品詞や、英数字や漢字の語数ならびに出現率、のような構文情報の特徴量を使用した. さらに、単語心像性、文末表現、単語親密度、表記妥当性も特徴量として使用した[3]. 結果として、全般的に良好な推定精度が得られた. また、主要な特徴量を獲得することによって、因子得点の推定の妥当性を確認した[3].

ここで、国立情報学研究所[4]が回答文に対して客観的な評価を行い、グッドアンサーの選出を試みている[5].

「グッドアンサー」とは、良回答であると多くの人に同意される回答文である. この評価は、回答文が質問文に対してどの程度良好または適切であるかを示している. 評価の度合いは、4 人の評価者により 9 段階で決定される. 4 人の評価者が評価を行うので、これらの評価は客観的であると考えられる. この客観的評価値を質問文と回答文の因子得点から推定する事ができると、質問文と回答文の関係を評価できていることになり、質問者と回答者の相性の判断に向けての前進になると考えられる.

これまで、回答文の客観的評価値の推定を重回帰分析により試みてきた[6]. 既に獲得している重回帰式を使用することにより推定された因子得点から、これらの値を推定した. その結果、Yahoo!オークションとパソコンの文章の客観的評価値に関しては、やや良好な推定結果が得られ、恋愛相談・人間関係の客観的評価値に関しては、因子得点の自然対数を適用することにより、良好な推定精度が得られることを示している[6]. しかしながら、Yahoo!オークションとパソコンの 2 ジャンルに関しては、推定精度をさらに向上させる必要があると考えられる.

そこで、本論文では、回答文の客観的評価値の推定精度の向上を図る. ここでは、因子得点に対してステイヴンスの法則とフェヒナーの法則を適用する. また、多重共線性を考慮するにあたり、説明変数になるべく多く残るように削減する. 推定の結果、パソコンの文章に関しては推定精度を向上させることができた.

本論文の構成は次の通りである. 2.では関連研究について述べる. 3.では、これまでの研究として、文章の因子得点の推定精度の結果を概説する. 4.では、国立情報学研究所が試みているグッドアンサーについて説明し、5.でこれまで行った回答文の客観的評価値の推定結果を示す. 6.では新たに行った推定法とその結果を示す. 最後に、7.で論文をまとめる.

2. 関連研究

これまで、「ベストアンサー」を推定する研究が行われてきている[7-11]. Blooma らは、非テキスト特徴量とテキスト特徴量を用いて、「ベストアンサー」の推定を試みている[7]. Agichtein らは、内容や語法の特徴量を使用することによって、質問文と回答文の質の評価を試みている[8]. また、類推による手法も提案されている[9]. この手法では、過去の知見における質問文と回答文のリンクを使用することによって、「ベストアンサー」を探索する. Kim らは、「ベストアンサー」の選択基準を提案している[10]. 情報型の質問は、文章内容の特徴量が重要であり、提案型の質問には有用性が重要であり、選択型の質問には社会的な感情が重要であるとしている. 西原らは、ある 1つの質問文に対する回答文群より、「ベストア

ンサー」になりやすいものを検出する手法を提案している[11].

これらの研究は「ベストアンサー」の推定に重点が置かれた研究である。これに対して、本研究では、「グッドアンサー」を考慮することにより、質問文に適切な回答を施すことができる回答者を質問者に引き合わせることを図っている。

3. 文章の因子得点の推定

3.1 文章の因子

2005年9月にYahoo!知恵袋[1]に投稿された12組60個の4大ジャンル（Yahoo!オークション、パソコン、恋愛相談・人間関係、政治・社会問題）の質問回答文に対して、印象評価実験を実施した。実験結果に対して因子分析を施したところ、文章に関する因子が9個得られた[2]。因子とは、複数の印象語により説明された文章の性質を意味する。的確性、不快性、独創性、容易性、執拗性、曖昧性、感動性、努力性、熱烈性の9因子が得られた。因子に対応する印象語を表1に示す。また、因子得点も得られている。これらの得点は、文章の特徴を表すのに使用される。

3.2 文章の特徴量

3.1で得られた因子得点は、実験で使用した60個の質問回答文から得られたものだけである。そこで、どのような文章の因子得点の推定も可能とするために、文章の特徴量に対し重回帰分析を適用した。表2に記載された77個の特徴量を使用している。以下、各特徴量を説明する。

3.2.1 構文情報

文章の数や長さ、名詞や動詞といった品詞の数や割合、などといった構文情報を抽出した。感嘆符や疑問符といった具体的な記号も特徴量として使用している[3]。さらに、語数に対する語彙数の割合を意味するTTR（Type Token Ratio）も考慮している。これらの特徴量を表2のg1-g36に示す。

3.2.2 単語心像性

単語心像性も文章の特徴量として使用している。単語心

表1: 9因子と対応する印象語

因子	印象語
第1因子(的確性)	説得力がある 流暢な 重要な
	素晴らしい 好ましい 巧みな
	真実味がある 清々しい 妥当な
	充実した 美しい 的確な
第2因子(不快性)	丁寧な
	不快な 憤慨した 非常識な
	残念な 不当な 呆れる
第3因子(独創性)	幻滅した 怖い
	独創的な 予想外な 特殊な
第4因子(容易性)	斬新な 不思議な
	易しい 明瞭な 難しい
第5因子(執拗性)	細かい しつこい 長い
第6因子(曖昧性)	曖昧な 不十分な
第7因子(感動性)	心温まる 感動的な
第8因子(努力性)	涙ぐましい
第9因子(熱烈性)	熱い 力強い

表2: 因子得点の推定に使用している特徴量

g	特徴量
g1	助動詞(語彙数)
g2	接頭詞
g3	記号(語彙数)
g4	文数
g5	文の長さ平均(字数)
g6	カタカナ(語数)
g7	全角記号(語数)
g8	全角英数字(語数)
g9	形容詞(語数)
g10	副詞(語数)
g11	連体詞(語数)
g12	接続詞(語数)
g13	感動詞(語数)
g14	ひらがな(%)
g15	漢字(%)
g16	カタカナ(%)
g17	記号(%)
g18	TTR
g19	全角記号(%)
g20	英数字(%)
g21	全角英数字(%)
g22	名詞(%)
g23	形容詞(%)
g24	副詞(%)
g25	連体詞(%)
g26	接続詞(%)
g27	感動詞(%)
g28	「!」の数
g29	「?」の数
g30	句点の数
g31	読点の数
g32	中点の数
g33	3点リーダの数
g34	鍵括弧の数
g35	括弧の数
g36	「/」の数
g37	単語心像性4点台(語数)
g38	単語心像性6.5以上7.0未満(語数)
g39	か(語数)
g40	な(語数)
g41	し(語数)
g42	たい(語数)
g43	ない(語数)
g44	だ(文末語数)
g45	か(文末語数)
g46	な(文末語数)
g47	し(文末語数)
g48	です(文末語数)
g49	ます(文末語数)
g50	たい(文末語数)
g51	ない(文末語数)
g52	ぞ(%)
g53	だ(%)
g54	よ(%)
g55	ね(%)
g56	か(%)
g57	です(%)
g58	ます(%)
g59	ない(%)
g60	か(文末%)
g61	ですか(語数)
g62	不是吗(語数)
g63	ますか(語数)
g64	ました(語数)
g65	単語親密度該当単語率
g66	単語親密度6.5以上7.0未満(語彙数)
g67	単語親密度4点台(語数)
g68	単語親密度5点台(語数)
g69	単語親密度5.5以上6.0未満(語数)
g70	単語親密度6点台(語数)
g71	単語親密度6.0以上6.5未満(語数)
g72	表記妥当性該当単語率
g73	表記妥当性3点台(語数)
g74	表記妥当性3.5以上4.0未満(語数)
g75	表記妥当性4点台(語数)
g76	表記妥当性4.0以上4.5未満(語数)
g77	表記妥当性5点台(語数)

像性とは、単語から喚起される様々なイメージが、どの程度思い浮かべやすいかを示す主観的特性である。例えば、「りんご」という言葉を聞くと、赤・黄・緑の丸い形の果物、甘くみずみずしい味・匂い、サクッとした音や歯ざわり、持った時の感触を思い浮かべることができるため、「りんご」の単語心像性の特性値は高いといえる。単語心像性の特性値は、1-7の7段階で評価されている[3]。これらの特徴量を表2のg37とg38に示す。

3.2.3 文末表現

文末表現として、「ぞ」「だ」「よ」「ね」「か」「な」「し」「です」「ます」「たい」「ない」といった語を使用している[3]。例えば、「ぞ」を例にすると、「ぞ(語数)」は文中に出現する「ぞ」の数であり、「ぞ(文末語数)」は文末に出現する「ぞ」の数である。また、「ぞ(%)」や「ぞ(文末%)」は、これらの含有率である。これらの特徴量を表2のg39-g64に示す。

3.2.4 単語親密度

単語親密度は、単語にどの程度なじみがあるかを表す指標である[12]。ここでは、単語は1-7の7段階で視覚的にも聴覚的にも評価されている。ここでは、視覚的に評価されたデータのみを使用する。例えば、「たんに」という単語を例にすると、「単に」を意味する場合は、単語親密度の特性値が5.312となる。一方で、「乾電池の単二」を意味する場合は、特性値が3.594となる。これらの値は、被験者の評価値を平均した値である。これらの特徴量を表2のg65-g71に示す。

3.2.5 表記妥当性

表記妥当性は、単語表記の妥当性を表すもので、1-5の5段階で評価されている[12]。単語には複数の意味や表記がある。例えば、「こしょう」という単語を例にすると、「故障」「呼称」「湖沼」などの意味が考えられる。また、漢字やひらがな、カタカナで表記される場合や、または漢字やひらがなを混ぜて表記される場合もある。表記妥当性のそれぞれの場合が記録されている。例えば、「ごらく」という単語を例にすると、ひらがなで書かれた場合は表記妥当性の特性値が2.95であり、カタカナで書かれた場合は特性値が1.95であり、漢字で書かれた場合は特性値が4.90である。

単語には異なった意味や読みがある場合もある。例えば、「アース」には「地面」や「地球」、「殺虫剤」の意味がある。形態素解析されたデータにそのような単語が含まれている場合は、その単語の文脈に応じて判断することにより手動で決定する。これらの特徴量を表2のg72-g77に示す。

3.3. 因子得点の推定

印象評価実験で使用された60個の質問回答文に関して、表2で示した77個の説明変数を基盤とした281個の二次項(説明変数同士の積)を説明変数とし、表1で示す9因子の因子得点を目的変数として、重回帰分析を実施した。重回帰式が結果として得られた。例えば、第5因子の因子得点の推定式は式(1)により表される。

$$y_5 = 0.182g_{18}g_{18} + 0.000280g_5g_6 + 0.00467g_{24}g_{60} - 0.0467g_{23}g_{58} + 0.00985g_4g_{58} + 0.102g_{23}g_{29} + 0.339g_{33}g_{45} - 0.201g_{66}g_{71} - 0.0149g_{51}g_{72} - 0.266g_{51}g_{54} - 0.672 \quad (1)$$

紙面が限られているため、全ての重回帰式の組をここには掲載しない。推定精度の良好性を示す重相関係数を表3に示す[4]。全ての値が0.9以上であるため、9因子とも推定精度が非常に良好であるといえる。

4. グッドアンサー

「グッドアンサー」を決定することを目的として、国立情報学研究所では回答文の質の決定を試みている[5]。「グッドアンサー」とは、多くの人により良回答であると同義される回答文である。ここでは、4人の評価者が1500個の質問文と7443個の回答文に対して、A,B,Cの3段階で評価を行っている。Aは質問文に対して十分な回答文、Bは部分的に関連する回答文、Cは質問文に関連の無い回答文、を意味する。これらの評価に重みを定義するために、A,B,Cに対してそれぞれ2,1,0の得点が与えられる。4人の評価者の合計値は0から8となる[5]。これらはそれぞれL0,L1,...,L8と表記され、「Good Answer with Weights」(GAWs)と呼ばれる[5]。4人の評価者が評価を行ったため、GAWsは客観的評価値であると考えられる。従って、質問者と回答者のマッチングに有用であると考えられる。

質問回答文とGAWsの例を表4に示す[13,14]。各回答文にはGAWが付与されており、PA04-01/LA06-01と表記されている回答文は実際に「ベストアンサー」に選択された回答文である。残りの回答文は投稿時間に基づいて時系列順に並べられている。表4の例より、「ベストアンサー」が必ずしもその組の中で最大のGAW値が付与されているとは限らないことがわかる。

5. GAWsの推定

5.1 推定手法

既に獲得している重回帰式から得られた因子得点より、GAWsの値の推定を試みる。ここでは、GAWsの値を推定するのに再度重回帰分析を適用する。回答文の因子得点の値そのもの(“VALUES”と表記)に加え、回答文と質問文または他の回答文との差を考慮するために、以下の値を重回帰式の説明変数として使用する。

表3: 重相関係数

因子	重相関係数
第1因子(的確性)	0.989
第2因子(不快性)	1.000
第3因子(独創性)	0.999
第4因子(容易性)	1.000
第5因子(執拗性)	0.925
第6因子(曖昧性)	1.000
第7因子(感動性)	0.963
第8因子(努力性)	0.950
第9因子(感動性)	1.000

表 4: 質問回答文と GAW の例

ID	文章	GAW
PQ04	4月に新しくパソコンを買いました。 それと同時にウイルス駆除ソフトも買いました。 が、……………今まで検索で引っかかった事はありません。 今となって後悔しています。 だって、メールとかなら(Hotmail)届く前に削除されるし。 皆さんって、なぜソフトを使っているのですか??? Yahooとか見ていて感染することは無いですよね?	-
PA04-01	最近のウイルスはインターネットに接続しただけで、 侵入してくるタイプもあるからです。 昔からあるウイルスも、ホームページを見ただけで感染したり、 ディスク感染したりしましたけど。	L7
PA04-02	ネタでしたら、許してあげる……………	L2
PA04-03	油断して、感染して、友達に送っちゃったりしたら嫌だから。 友達のところにおくっても、そっちで防御しちゃうかもしれない いけどね。	L5
PA04-04	あなたのところは、平和だなあ… うらやましい限りです。	L3
PA04-05	知らないよ!! 馬鹿!! そんな質問するな!! クソ!!!!!!!!!!!!!!!!!!!!	L2
PA04-06	もし、良かったらソフトを無効にして一週間ほどネットしてみてください。 感染すると思いますよ。 マイクロソフトのサイトでも感染して事あるのに…。	L4
PA04-07	ウイルス駆除ソフトは、転ばぬ先の杖です。5年以上パソコン やっけて、2、3回変なリンク先で引っかかる程度ですが、 ヤフーだけなら、まず、安心出来ます、が感染しない保障はありません。	L8
ID	文章	GAW
LQ06	「人間顔じゃないよ!」ってカッコいい人や美人な人に言われても いまいち励みにならない気しません? これって、ひねくれてるんでしょうか?	
LA06-01	カッコいい人や美人な人は顔だけで見られて中身を見てもらえないのかも。 普通の人が言うよりも説得力がある気がします。	L7
LA06-02	はい。その通りです。 知性や内面に自信がありませんか?	L6
LA06-03	「人間の顔じゃないよ!」って言われたらショックだよ……………	L3
LA06-04	逆に、あんたはいいよね……、とむかつくような気がしますね。	L5
	～くろすけ～	
LA06-05	でも、実際にかっこよくない人とお付き合いしていたときは、 「ほんとはね、カレミたら納得できた(笑)」って言われましたよ。 ほんとに中身が大事なので、中身を選びますが、釣り合っていないって よく言われます。 なんであいつなの?ってカッコいい人からも言われます。 でも、それでも、かっこよくなくても、その人が好きだったので、 本気で人は顔じゃないなって思ってます。	L7

- DIS: ある回答文の因子得点とその質問文の因子得点の差の絶対値
- DIS_MAX: ある回答文の因子得点とその組における回答文の最大値との差 (回答文の因子得点そのものが最大値であるならば, DIS_MAX の値は 0.00)
- DIS_MIN: ある回答文の因子得点とその組における回答文の最小値との差 (回答文の因子得点そのものが最小値であるならば, DIS_MIN の値は 0.00)

ここでは3大カテゴリー (Yahoo!オークション, パソコン, 恋愛相談・人間関係) に着目する. 説明変数の識別子は h_{A1} , h_{A9} , h_{A10} , h_{A18} , h_{A19} , h_{A27} , h_{A28} , h_{A36} であり, それぞれ Yahoo!オークション, パソコン, 恋愛相談・人間関係を表す. 例として, Yahoo!オークションを表す説明変数 h_{A1} - h_{A36} を表 5 に示す. 添え字の "A" は, パソコンの場合は "P", 恋愛相談・人間関係の場合は "L" に置換される.

質問回答文の因子得点の推定を行った場合と同様に, 多重共線性と二次項を考慮している. 目的変数を国立情報学研究所が提供した GAWs の値とし, 説明変数を "VALUEs," "DIS," "DIS_MAX," "DIS_MIN" に基づいたもの

として, 重回帰分析を行った.

サンプル数や回答文の数, 説明変数の数, 重相関係数を表 6 に示す. 重相関係数は, 値が 0.9 以上であれば推定精度が良好であることを示し, 値が 0.7 以上 0.9 未満であれば推定精度がやや良好, 値が 0.7 未満であれば推定精度が不良であることを表す. この基準より, Yahoo!オークションとパソコンはやや良好な推定精度が得られたものの, 恋愛相談・人間関係は推定精度が不良であるといえる.

5.2 自然対数の考慮

恋愛相談・人間関係の文章の客観的評価値の推定精度を向上するために, フェヒナーの法則 [15] に基づいて, DIS の得点に自然対数を適用した. 刺激強度 E が変化する時, E に対応する感覚量 S との関係は, 比例定数 k を用いて式 (2) で表される.

$$E = k \log S \quad (2)$$

従って, 心理的な感覚量は, 刺激量の強度の対数に比例して知覚されるという関係式を導出した形が, フェヒナーの法則である [15].

また, 推定精度が向上するかどうかを調べる為に, 質問回答文のサンプル数を増加した. 5.1. での試行と同様に, 多重共線性と二次項を考慮している.

サンプル数や回答文の数, 説明変数の数, 重相関係数を表 7 に示す. 自然対数を考慮した場合, 恋愛相談・人間関係はかなり向上して推定精度が良好となっている. 一方で, パソコンは推定精度がやや良好ではあるが推定精度が少々低下した. また, Yahoo!オークションはかなり低下して不良な推定精度となってしまった.

6. GAWs の推定精度の向上

6.1 スティーヴンスの法則とフェヒナーの法則

新たな説明変数として, 表 5 に示す h_{A1} - h_{A36} , ならび

表 5: Yahoo!オークションの場合の説明変数

変数	説明
h_{A1} - h_{A9}	第1因子~第9因子のVALUEs
h_{A10} - h_{A18}	第1因子~第9因子のDISs
h_{A19} - h_{A27}	第1因子~第9因子のDIS_MAXs
h_{A28} - h_{A36}	第1因子~第9因子のDIS_MINs

表 6: 分析結果 (自然対数を考慮する前)

カテゴリー	サンプル数	説明変数	重相関係数
Yahoo!オークション	13組91回答文	181	0.710
パソコン	15組80回答文	158	0.727
恋愛相談・人間関係	16組119回答文	168	0.478

表 7: 分析結果 (自然対数を考慮した後)

カテゴリー	サンプル数	説明変数	重相関係数
Yahoo!オークション	19組198回答文	279	0.552
パソコン	19組140回答文	250	0.701
恋愛相談・人間関係	16組208回答文	163	0.927

に $h_{p1}-h_{p36}$ に対し、スティーヴンスの法則[16]を適用した。刺激強度 E が変化するとき、 E に対応する感覚量 S との関係は、べき関数の指数部 b を用いて、式(3)で表される。

$$E = \begin{cases} S^b & (x \geq 0) \\ -|S|^b & (x < 0) \end{cases} \quad (3)$$

また、フェヒナーの法則に関しては、5.2にて式(2)で示しているが、VALUE, DIS_MAX, ならびに DIS_MIN の値が 0.00 の場合、式(2)では値をとることが不可能である。そこで、値が 0 の場合に対応できる式として、刺激強度 E が変化するとき、 E に対応する感覚量 S との関係は、対数関数の底 α を用いて、式(4)で表される[17]。

$$E = \begin{cases} \log_{\alpha}(S+1) & (x \geq 0) \\ -\log_{\alpha}(|S|+1) & (x < 0) \end{cases} \quad (4)$$

式(4)を $h_{A1}-h_{A9}$, $h_{A19}-h_{A36}$, $h_{p1}-h_{p9}$, $h_{p19}-h_{p36}$ に適用している。スティーヴンスの法則におけるべき関数は $b=0.3$ 、フェヒナーの法則におけるべき関数は $\alpha=4$ 、とそれぞれ実験的に設定して、 $h_{A1}-h_{A36}$, ならびに $h_{p1}-h_{p36}$ に対して適用した。例として、Yahoo!オークションを表す説明変数 $h_{A1}-h_{A36}$ を表 8 に示す。添え字に関する説明は、5.1 で示している。

6.2 説明変数の選択方法

表 8 に示す説明変数に関して、5.1 で示した推定手法と同様に、ここでも二次項と多重共線性を考慮している。多重共線性を考慮するにあたり、以下の工程を反復する。

- (1) 以下の設定をする必要がある。
 - ・ n 行 n 列の相関行列：説明変数間同士のもの
 - ・ $(n+1)$ 行目：各列の値が 0.7 以上（多重共線性の閾値）の個数から 1 を減算した値を設定
 - ・ $(n+2)$ 行目：1, 2, ..., n の通し番号を設定
 - ・ $(n+1)$ 列目：各行の値が 1 である個数を設定
 - ・ $(n+2)$ 列目：1, 2, ..., n の通し番号を設定
- (2) $(n+1)$ 行目の値の昇順に、データ列を並び替える。
- (3) $(n+1)$ 行目の値に関して、適当な閾値 $p (< n)$ 以上の値の列を k 列削除する。
- (4) $(n+2)$ 行目の値の昇順に、データ列を元通りに並び替える。
- (5) $(n-k+1)$ 列目の値を昇順に、データ行を並び替える。
- (6) $(n-k+1)$ 列目の値が 0 である行を k 行削除する。
- (7) $(n-k+2)$ 列目の値を昇順に、データ行を元通りに並び替える。

以後、(2) ~ (7) の工程を繰り返していくことにより、(3) の段階で適当な閾値 p' ($< p$) を設定した上で、説明変数の数を削減していく。2 度目以降の閾値は、(2) の時点で $(n+1)$ 行目の値のうち最大値から 5 少ない値の範疇で、列を削除していく（例えば、1 回目は $p=20$ と設定した場合、2 回目は $p'=15$ として作業を工程する）。

表 8: Yahoo!オークションの場合の説明変数

変数	説明
$h_{A1} - h_{A9}$	第1因子～第9因子のVALUEs
$h_{A10} - h_{A18}$	第1因子～第9因子のDISs
$h_{A19} - h_{A27}$	第1因子～第9因子のDIS_MAXs
$h_{A28} - h_{A36}$	第1因子～第9因子のDIS_MINs
$h_{A37} - h_{A45}$	DISsを自然対数化
$h_{A46} - h_{A54}$	VALUEsにスティーヴンスの法則を適用
$h_{A55} - h_{A63}$	DISsにスティーヴンスの法則を適用
$h_{A64} - h_{A72}$	DIS_MAXsにスティーヴンスの法則を適用
$h_{A73} - h_{A81}$	DIS_MINsにスティーヴンスの法則を適用
$h_{A82} - h_{A90}$	VALUEsにフェヒナーの法則を適用
$h_{A91} - h_{A99}$	DIS_MAXsにフェヒナーの法則を適用
$h_{A100} - h_{A108}$	DIS_MINsにフェヒナーの法則を適用

表 9: 分析結果

試行	p	カテゴリー	反復回数	説明変数	重相関係数
試行1	5	オークション	5	204	0.444
		パソコン	5	216	0.576
試行2	20	オークション	8	487	0.577
		パソコン	8	520	0.761
試行3	35	オークション	10	590	0.565
		パソコン	10	659	0.813
試行4	50	オークション	11	656	0.588
		パソコン	11	675	0.863

6.3 推定結果

ここでは、4 度の試行を行っている。

- ・ 「試行 1」：6.2 の (3) において、最初に $p=5$ として、表 8 に示す説明変数の単項と二次項との計 5994 個の説明変数の削減を行った時の試行である。この時の p の値、反復回数、説明変数、重相関係数を表 9 の「試行 1」に示す。（以下の各試行についても同様）
- ・ 「試行 2」：最初に $p=20$ として、説明変数の削減を行った時の試行である。
- ・ 「試行 3」：最初に $p=35$ として、説明変数の削減を行った時の試行である。
- ・ 「試行 4」：最初に $p=50$ として、説明変数の削減を行った時の試行である。

4 度の試行を通して、Yahoo! オークションは依然として推定精度が不良のままであったが、パソコンの場合は「試行 2」から「試行 4」までを通じて、やや良好な推定精度が得られている。また、 p の値を大きくしていくにつれ、説明変数は徐々に増加し、重相関係数の値が概ね向上している。重相関係数の値が大幅に増えた理由としては、「試行 1」で $p=5$ 以上に設定すると削除されてしまった推定に有効な説明変数が、「試行 2」以降で p の値を大きくしたことにより、結果的に増えたためであろうと考えられる。

7. まとめ

本論文では、質問回答サイトに投稿された回答文の客観的評価値 GAWs[5]の推定精度の向上を試みた。ここでは、Yahoo!オークションとパソコンの2ジャンルの文章に関して、因子得点に対してステイヴンスの法則とフェヒナーの法則を適用した。また、多重共線性を考慮するにあたり、説明変数になるべく多く残るように削減した。その結果、パソコンはやや良好な推定精度が得られたが、Yahoo!オークションは不良な推定精度となってしまった。

Yahoo!オークション、パソコンのいずれの文章に対しても、回答文のGAWsの推定精度を向上させることが今後の課題である。これまでは、文章の因子得点とそれに基づいた値を説明変数として使用してきた。しかしながら、因子得点に基づいた説明変数だけでは、GAWsの推定を試みることに一定の限界があると考えられる。恋愛相談・人間関係の文章は、自然対数を適用したことにより、良好な推定精度が得られたため、主観的な価値判断に基づいた回答がグッドアンサーとして選択されると考えられる。一方で、Yahoo!オークションやパソコンは、客観的事実に基づいた回答がグッドアンサーとして選択されると考えられる。そこで、今後の方針としては、質問文と回答文の両者に共通する「キーワード」への着目を検討している。「キーワード」の語数や、「キーワード」がそれぞれの文章中の名詞に対してどの程度の割合で出現するかを調べていき、説明変数に追加して調べていきたいと考えている。

また、GAWsを用いてベストアンサーを推定することも、今後の課題である。このために、質問者の特性を解明し使用する必要があると考えている。さらに、質問文に対して適切な回答が見込める利用者を選出するプロトタイプシステムの構築・評価を行う予定である。また、特徴量の大部分が日本語に特化しているため、他言語への対応も今後の課題である。

謝辞

本研究は、一部、科研費(25280110)の助成を受けて行われたものである。また、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!知恵袋のデータを利用している。

参考文献

- 1) Yahoo!知恵袋(URL), <http://chiebukuro.yahoo.co.jp/>
- 2) 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 質問回答サイトの質問文と回答文の印象評価とベストアンサーの推定, 日本感性工学会論文誌, Vol.10, No.2, pp.221-230, 2011.
- 3) 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, Vol.12, No.1, pp.1-10, 2013.
- 4) 国立情報学研究所(URL), http://www.nii.ac.jp/cscenter/idr/yahoo/chiebkr2/Y_chiebukuro.html, 2012.
- 5) Sakai et al.: Overview of the NTCIR-8 Community QA Pilot Task (Part II): System Evaluation, NTCIR-8 Proceedings, pp.433-457 (2010), <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings8/NTCIR/02-NTCIR8-0V-CQA-SakaiT.pdf>
- 6) 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司: 質問回答サイトに投稿された回答文の客観的評価値の推定, 第15回日本感性工学会大会, 2013.
- 7) Blooma, M.J. and Chua, A.Y.K. and Goh, D.H.L.: A Predictive Framework for Retrieving the Best Answer, Proc. of 2008 ACM Symposium on Applied Computing (SAC08), pp.1107-1111, 2008.
- 8) Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G.: Finding High-Quality Content in Social Media, Proc. of the Int'l Conf. on Web Search and Web Data Mining (WSDM08), pp.183-194, 2008.
- 9) Wang, X. J., Tu, X., Feng, D. and Zhang, L.: Ranking Community Answers by Modeling Question-Answer Relationships via Analogical Reasoning, Proc. of 32nd Int'l ACM SIGIR Conf., pp.179-186, 2009.
- 10) Kim, S., Oh, J. S. and Oh, S.: Best-Answer Selection Criteria in a Social Q&A site from the User-Oriented Relevance Perspective, Proc. of American Society for Information Science and Technology (ASIS&T) 2007 Annual Meeting, 2007.
- 11) 西原陽子, 松村真宏, 谷内田正彦: Q&A コミュニティでの質疑応答パターンの理解, 第22回人工知能学会全国大会, 1H2-7, 2008.
- 12) 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳格, 田中正之, 天野成昭, 近藤公久: 単語心像性①, NTT データベースシリーズ日本語の語彙特性 第3期(第8巻), (社)三省堂, 2005.
- 13) 解決済みのYahoo!知恵袋(URL), http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q10850500
- 14) 解決済みのYahoo!知恵袋(URL), http://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q102608985
- 15) 田中良久: 心理学的測定法 第2版, (社)東京大学出版会, pp.143-144, 1985.
- 16) 大山正, 詫摩武俊, 中島力, "新版 心理学," 有斐閣双書, pp.137-142, 1993.
- 17) 北川高嗣, 中西崇文, 清木康: 静止画像メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的画像検索への適用, 情報処理学会論文誌:データベース. Vol.43. No.SIG 12(TOD16), pp.38-51, 2002.