

QAサイトにおける質問に適した回答の判定

西原陽子^{1,2} 松村真宏³ 谷内田正彦²

¹ 日本学術振興会 ² 大阪大学大学院基礎工学研究科 ³ 大阪大学大学院経済学研究科

1 はじめに

Yahoo!知恵袋 [1] や教えて!goo[2] など、自分の分からないことを質問すると、他の人が回答してくれる Web 上の QA サイトがある。QA サイトでは、質問者から回答者にお礼として、ポイントが付与されることが多い。ポイントは、回答者全員にではなく、質問者が気に入った回答文（ベストアンサー）を書いた数人にのみ付与される。回答者は、ベストアンサーとなる回答を書こうとするが、ベストアンサーとなる回答の基準は、明確に知られていない。

本稿では、質問と複数の回答が与えられたときに、ベストアンサーになる可能性が高い回答を判定する手法を提案する。提案手法では、ベストアンサーとなる基準には、質問と回答の文体の相性があると仮定する。文体の相性として、質問と回答の文末表現の組み合わせに注目し、使用されている文末表現によって質問と回答のペアをクラスタリングする。得られるクラスタの中で、ベストアンサーとそうでない回答を SVM を用いて学習し、回答の判定器を作成する。以下では、関連研究、提案手法の説明、評価実験について述べ、最後にまとめる。

2 関連研究

質問に適した回答を判定する手法と、文体の相性を評価する手法の従来研究をまとめ、提案手法との相違を示す。

2.1 質問に適した回答の判定

従来の、質問に対する適切な回答の評価手法には、質問のタイプを 5W1H で分類し、Web やコーパスでの検索結果から、正しい回答に近い文章を抽出するものが多かった [3, 4]。従来手法では質問に対して正しい回答を判定することはできるが、本研究で対象とする QA サイトでは、正しい回答であってもベストアンサーに選ばれるとは限らない。このため、従来手法ではベストアンサーとなる回答を判定することは難しいと考えられる。提案手法では、ベストアンサーに選ばれるには、質問と回答の文体の相性が重要と仮定し、文体の相性を評価するところが、これらの研究と異なる。

また、質問者が選ぶ良い回答に見られる傾向から評価関数を作成し、回答の判定を行おうとした研究もある [5, 6]。これらの手法で用いた傾向は、文章量、文章の構成、接続詞数、名詞数などであった。これらの手法では回答文しか考慮しておらず、また、質問者によって、同じ回答でも選ばれる場合とそうでない場合があり、判定の精度は高くなかった。提案手法では、回答に加えて質問も用いる。そして、質問と回答の相性を評価することによって、質問と回答の相性ごとに判定する評価関数を変えるところが、これらの手法と異なっている。

2.2 文体の相性評価

質問と回答の文体の相性は、質問の意図と回答の意図の対応とみることもできる。文末表現を使って、人間の意図をとらえようとした研究は多く、パソコンヘルプシステムに出されるユーザの質問の意図の

同定 [8, 9] や、テーマのある会話での人間の発話意図の同定 [10]、ブログの書き手の意図の同定 [11] などがある。本研究では従来研究と同様に、文末表現によって質問と回答の文体を表し、文体の相性の評価に用いる。

3 質問に適した回答の判定手法

提案手法の処理手順を説明する。提案手法での入力には質問と回答が書かれたテキスト集合（質問・回答文）とし、出力はベストアンサー（BA）になる可能性が高い回答とする。質問（Q）には複数の回答があり、回答はBAとそうでない回答（NA）のどちらかにラベル付けされているとする。質問・回答文が入力されると、BAとそれに対するQから文末表現を抽出し、抽出された文末表現を用いて、BAとQのペアをクラスタリングする。続いて、クラスタごとに、SVMを用いてBAとNAの傾向を学習し、得られた回答の判定器を使って、新たな回答の判定を行う。以下では、各処理を説明する。

3.1 入力：質問・回答文の集合

入力する質問・回答文の集合は、QAサイトの質問カテゴリの中で、同一のカテゴリに属するものとする。回答が1つしかない質問・回答文や、最も早く投稿された回答がBAとなっている質問・回答文は、質問者によって回答が適切に判定されなかったとして、入力から外す。

3.2 質問・回答文からの文末表現の抽出

質問・回答文の集合から、各文末に含まれる助詞・助動詞の1,2,3-gramを文末表現として抽出する。抽出された文末表現を用いて、質問と回答の文体の相性を評価するため、ここで回答文はBAのみとする。形態素解析器はChaSenを用いる [12]。文末の判定は読点（.）、または句点（.）、記号（?または!）とする。抽出された文末表現は同じ1,2,3-gramであっても、

BAから抽出されたもの、Qから抽出されたものを別々に扱う。

3.3 質問・回答文のクラスタリング

抽出された文末表現を用いて、QとBAのペアをクラスタリングする。このとき、全ての文末表現を用いると、クラスタリングに多くの時間がかかるため、使用する文末表現を絞り込む。ここでは、含まれるQとBAのペアの数が多い、上位20%をクラスタリングに用いる。クラスタリングにあたり、QとBAのペアを、含まれる文末表現を属性に持つベクトル QB で表す。

$$QB(q_i, b_i) = (m_1, m_2, \dots, m_{N1}) \quad (1)$$

ただし、 q_i はQ、 b_i はBA、 m は抽出された文末表現、 $N1$ は出現頻度上位20%の文末表現の数とする。ベクトル QB の文末表現 m_j の評価値は $tfidf$ 値とする。

$$tfidf(m_j, q_i, b_i) = tf(m_j, q_i, b_i) \times \left(\log \frac{N}{df(m_j)} + 1\right) \quad (2)$$

ただし、 tf は文末表現 m_j の q_i, b_i での頻度、 N は QB の数、 $df(m_j)$ は文末表現 m_j が含まれる QB の数とする。

ベクトル QB を階層的にクラスタリングし、クラスタリング結果のデンドログラムを参照し、人が判断してクラスタ数を決定する。クラスタ間の距離はWard法により測る。Ward法では式(3)によって、クラスタ C_1 と C_2 の距離が測られる。

$$C(C_1, C_2) = E(C_1 \cap C_2) - E(C_1) - E(C_2) \quad (3)$$

ただし、 c をクラスタ C のセントロイドとし、 $E(C)$ は式(4)によって評価される。

$$E(C) = \sum_{QB \in C} (D(QB, c))^2 \quad (4)$$

3.4 回答文の機械学習

クラスタごとに、線形 SVM を使って、分類器を機械学習する。先に得られたクラスタに含まれる回答文は BA だけなので、クラスタ中に含まれる質問文 Q に対応する NA を追加する。このとき、BA に似た NA が含まれると学習がしづらいため、BA と NA のコサイン値を求め、値が最も高い NA は追加しない。BA と NA のコサイン値は式 (5) で評価される。

$$\cos(b_i, n_k) = \frac{b_i \cdot n_k}{|b_i| \times |n_k|} \quad (5)$$

1 つの Q に対し、NA¹ を 1 つだけクラスタに入れ、クラスタごとに回答文の判定器を作成する。

SVM で処理する際は、BA を正例、NA を負例とする。回答文 (BA、NA の両方) a_i に対し、SVM で処理するベクトルを式 (6) で表す。

$$A(a_i) = (x_1, x_2, \dots, x_l) \quad (6)$$

ベクトル $A(a_i)$ の素性 x_l は、名詞、動詞、形容詞の 1.2.3-gram、文末表現の 1.2.3-gram とする。

3.5 出力：質問に適した回答の候補

得られた判定器を使って、質問に適した回答の候補 (BA の候補) を出力する。

4 評価実験

提案手法でのクラスタリングの効果と、BA に似た NA の除去の効果を調べた。

4.1 実験概要

使用したデータは QA サイトの「Yahoo!知恵袋」に投稿された質問・回答文の集合になる。その中から投稿件数の多い 3 つのカテゴリ「恋愛相談、人間

関係の悩み」「パソコン、周辺機器」「政治、社会問題」を選び、実験で用いた。実験には、BA の選ばれ方が不適切なデータを除き、各カテゴリから 5,000 件ずつを取り出して用いた。

クラスタリングの効果と、BA に似た NA 除去の効果を調べるため、以下の 3 つの手法を用意した。

- ・BL: 名詞、動詞、形容詞、文末表現の 1,2,3-gram を素性として、SVM
- ・CL: クラスタリングして、BL と同じ素性を用いて、SVM
- ・PP: BA と似た NA を除去し、クラスタリングして、BL と同じ素性を用いて、SVM

BL と CL の結果を比較することで、クラスタリングの効果の測り、CL と PP の結果を比較することで、BA と似た NA の除去の効果の測る。

実験で用いた SVM は SVMlight[7] になる。データは 10 等分し、10 分の 9 を学習データ、10 分の 1 をテストデータとし、分類の精度、適合率、再現率を求めることを 10 回行った。

4.2 実験結果

実験で得られた、分類の精度、適合率、再現率の平均を図 1 に示す。クラスタの数は、「恋愛関係、人間関係の悩み」では 4、「パソコン、周辺機器」では 3、「政治、社会問題」では 3 とした。

4.2.1 クラスタリングの効果

BL と CL で得られた図 1 の結果を比較すると、精度、適合率、再現率の全てで CL での数値が高かった。これは Q と BA に含まれる文末表現でクラスタリングすることによって、SVM で BA と NA の特徴を学習しやすくなったためと考えられる。

表 1 に「恋愛関係、人間関係の悩み」の 4 つのクラスタに見られた、特徴的な属性と解釈された相性を示す。表 1 では、クラスタの特徴的な属性から、Q

¹複数の NA の中から、学習する NA はランダムに選ばれる。

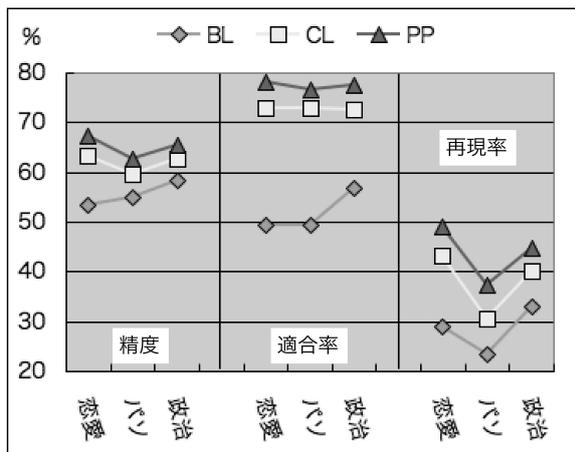


図 1: 分類の精度、適合率、再現率の平均

とBAの相性が解釈できている。回答は同じ単語が使われていても、Qによって、BAになるときとNAになるときがあり、クラスタリングしないままSVMで学習しても、精度の良い判定器は得にくい。しかし、文体の相性によって、あらかじめ回答をクラスタリングし、BAの傾向を学習させることで、精度の良い判定器が得られると考えられる。そのためQとBAの文末表現を用いて、回答をクラスタリングすることによって、判定精度が上昇したと考えられる。以上より、QとBAに含まれる文末表現によって、クラスタリングを行うことで、BAの判定精度が上がると分かった。

4.2.2 BAに似たNA除去の効果

CLとPPで得られた図1の結果を比較すると、精度、適合率、再現率の全てでPPでの数値が高かった。これは、BAに似たNAを除去することで、BAに特徴的な素性を学習しやすくなったためと考えられる。したがって、BAに似たNAを除去することで、判定精度が上がると分かった。

ただし、BAと最も似ているNAを除去しても、PPの精度・適合率・再現率はCLのそれと大きく変わらなかった。これは、3つのカテゴリで、NAと

表 1: クラスタの特徴的な属性と解釈された相性。カテゴリは「恋愛相談、人間関係の悩み」

| クラスタ | 属性と解釈された相性 |
|------|--|
| 1 | 「Aか、Qです、Qます」が多い 解釈：質問者は丁寧な誠意のある回答が欲しい |
| 2 | 「Aだ、Aです、Aない、Aですか、Qだ、Qたい、Qですか、Qない」が多い 解釈：質問者の意見とは、異なった面からの意見が欲しい |
| 3 | 「Aないです、Aました」が多い 解釈：質問者は正解ではなく、意見に対する感想が欲しい |
| 4 | 「Aないです、Aました、Qだ、Qです、Qない」が特に少ない 解釈：質問者は簡潔に答えだけが欲しい |

BAのコサインの平均値は0.6と高く、BAと最も似ているNAを除去しても、BAと似ているNAが含まれていたためになる。したがって、回答の学習で、BAに似たNAの処理方法を改善する必要があると分かった。

5 まとめ

本稿では、質問に適した回答を判定する手法を提案した。提案手法は、ベストアンサーに選ばれる回答は、質問との相性が良いと仮定し、相性を質問と回答文で使用される文末表現から評価する。相性ごとに質問と回答文をクラスタリングし、クラスごとに機械学習することによって、ベストアンサーになる可能性が高い回答が選び出される。

実験において、クラスタリングの効果とベストアンサーに似た回答の除去の効果調べたところ、両方とも判定精度の向上に効果があると分かった。今

後の課題には、ベストアンサーに似た回答が含まれている中で、学習方法を考えることがある。

謝辞

ヤフー株式会社様から「Yahoo!知恵袋」の質問・回答データの使用許可をいただきました。ここに記して感謝します。

参考文献

- [1] (URL)<http://chiebukuro.yahoo.co.jp/>
- [2] (URL)<http://oshiete.goo.ne.jp/>
- [3] Koji Eguchi: NTCIR-5 Query Expansion Experiments using Term Dependence Models, Proceedings of the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access, (2005).
- [4] Eugene Agichtein, Steve Lawrence and Luis Gravano, Learning search engine specific query transformations for question answering, The 10th International Conference on World Wide Web, pp.169–178, (2001).
- [5] 中瀬達也・砂山渡・橘啓八郎: 相互支援型コミュニティサイトにおける質問に適した回答の抽出, 電子情報通信学会知能ソフトウェア工学研究会資料, Vol.104, No.725, pp.61 – 66, (2005).
- [6] 渡辺凡・砂山渡: 電子掲示板におけるユーザの性質の評価, 電子情報通信学会知能ソフトウェア工学研究会資料, Vol.105, No.652, pp.25 – 30, (2006).
- [7] T. Joachims, Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, (1999).
- [8] 美馬秀樹・泓田正雄・林淑隆・青江順一: 自然言語インタフェースにおける間接発話文の意図理解法, 電子情報通信学会論文誌, Vol.J78-D-II, No.5, pp.803–810, (1995).
- [9] 熊本忠彦・伊藤昭・海老名毅: 支援対話におけるユーザ発話意図の認識-ユーザ発話文の解析に基づく統計的アプローチ- 電子情報通信学会論文誌, vol.J77-D-II, no.6, pp.1114–1123, (1994).
- [10] 西原陽子・砂山渡・谷内田正彦: 発話テキストからの人間の仲の良さと上下関係の推定, 電子情報通信学会, (2007)(採録決定).
- [11] 松村真宏・三浦麻子: ブログにおける書き手の意図とモダリティ表現, 第 19 回人工知能学会全国大会, (2005).
- [12] 松本裕治・北内啓・山下達雄・平野善隆・松田寛・高岡一馬・浅原 正幸: 日本語形態素解析システム『茶筌』 version 2.2.1 使用説明書, (2000).