

回答のクラスタリングによる 多様な主観的回答を行う相談システムの提案

畠山 恭明[†] 山名 早人^{‡§}

[†] 早稲田大学基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

[‡] 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

[§] 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: {hatake, yamana}@yama.info.waseda.ac.jp

あらまし インターネット上の掲示板に投稿される質問の中には、相談のような答えが一つに定まらないような質問が存在する。相談を投稿したユーザは、他のユーザからの返事を待つことで回答を得ることができるが、何も回答を得られないという場合も存在する。また、過去の類似している相談・回答を検索するとき、多様な意見を求めている場合は、大量の回答に目を通さなくてはならないという問題がある。そこで本研究では、過去の相談・回答を利用して、入力された相談文に対して回答文を出力するシステムを提案する。本システムは、回答を回答者の意見に基づきまとめあげることによって、多様な主観的回答を出力する。また、どのような理由で回答がヒットしたのかをシステム利用者に提示する。

キーワード non-factoid 型質問応答, 相談回答, 主観的回答

1. はじめに

Yahoo!知恵袋¹や 2ちゃんねる²といったインターネット上の掲示板に投稿される質問の中には、相談のような答えが一つに定まらないような質問が存在する。相談を投稿したユーザは他のユーザからの返事を待つことで回答を得ることができる。しかし、何も回答を得られないという場合も存在する。また、過去の類似している相談・回答を検索することは、多様な意見を求めている場合、大量の回答に目を通さなくてはならないという問題がある。また、回答が長文であった場合、回答の要点を把握することも困難である。そのため、相談文に対する似た回答をまとめ、回答のキーワードを提示するシステムは回答を効率的に閲覧するために有用であると考えられる。

入力された質問文に対して回答を出力するシステムは、大きく2つに分類できる。1つ目は Factoid 型質問応答システムと呼ばれる、回答の形式が単語となるシステムである。2つ目は Non-factoid 型質問応答システムと呼ばれる、回答の形式が文章であるシステムである。本研究で対象とするのは、文章を出力する Non-factoid 型である。Non-factoid 型質問応答システムの既存研究には、定義型や why 型, how 型といったある質問型に特化した研究[1][2][3], 質問型を特定せずに回答を行う Universal Question Answering(UQA)と呼ばれる研究[4][5], 最終的な出力回答を決定する研究[6][7]がある。

こうした既存研究では、類似回答のまとめを行っているが、回答の要点にそったまとめが行われているわけではない。そのため、回答が要点でない部分でまとめられてしまう場合がある。また、回答のキーワードなどを提示する仕組みは存在しないため、出力される回答の主張が何であるか、システム利用者は長文の回答を読まないと判断できない。

そこで本研究では、回答を回答者の意見に基づきまとめることにより、多様な主観的回答を出力することのできるシステムを提案する。提案システムでは、過去の相談・回答を利用して、入力された相談文に対して回答文を出力する。なお、本提案においては、質問分野は予め定められているとした上で、当該分野のキーワード辞書を予め作成し、同辞書を用いたクラスタリングを行う。キーワード辞書は過去の相談文の投稿タイトルから作成する。投稿タイトルには相談内容の要点を表現する語が含まれ、回答内容においても要点を表現する語となると考えられる。これらのキーワードをまとめ上げに用いることで、他の一般的な語の影響を避け、回答の要点によるまとめ上げを行う。また、長文の出力ではどのような理由で回答がヒットしたのかわからないため、回答のキーワードをシステム利用者に提示する。

また、過去の回答文には特定の状況に特化した回答も含まれるが、こうした特化した回答は一般的に回答としての確かな回答になることが少ない。これは、特化した状況での回答となるためである。そこで、こうし

¹ <http://chiebukuro.yahoo.co.jp/>

² <http://www.2ch.net/>

た特定の状況に特化した回答を出力しないようにするため、本システムでは、入力相談文に対して、過去のより単純な相談文に対する回答を出力することで、ある特定の状況に特化した回答を出力することを防ぐ。また、相談文の投稿者は、質問をしながらもすでに「投稿者自身が期待する回答」を持っている場合がある。本研究では「肯定的な回答」をその一つとし、回答文の肯定度により出力回答のスコア付けを行う。

本稿では以下の構成をとる。まず第2節で関連研究について述べ、第3節で提案手法について述べる、第4節で評価実験を行い、最後に第5節でまとめを述べる。

2. 関連研究

本節では、Non-factoid型質問応答を扱う研究の中でも、UQAに分類される関連研究、出力する回答を決定する関連研究について述べる。

UQAに分類される研究として、石下ら[4]は質問の型分類を行わず、過去のQ&Aコーパスから類似する記述スタイルの質問を見つけ、対応する回答集合から回答表現を生成し、回答の抽出に利用した。また、回答の情報源には、検索エンジンより得られたWeb上の文書を用いた。しかし、相談回答といった情報は一般のWebページには多く存在せず、回答の情報源としては適さない。Chaturvediら[5]は質問クラスタリングと質問回答の関連性予測を統合したモデルを提案し、質問型によらない質問応答システムを構築した。しかし、回答の内容が不足の場合と過剰の場合を区別していないため、入力に対して関係のない情報が出力されてしまう問題がある。

出力する回答を決定する研究として、Achananuparpら[6]は回答間の関連性によりグラフを構築し、回答の多様化を行った。石下ら[4]は語の頻度ベクトルにより回答のクラスタリングを行い、回答の冗長性の排除を行った。しかし、これらの手法は語の重要度を考慮していないため、回答の要点でない部分で回答の多様化・冗長性の排除が行われる可能性がある。Sakamotoら[7]は回答間のグラフ構造を構築し、回答に多くの観点含まれるものからランキングを行った。しかし、回答に多くの観点が含まれているほど、入力相談文と対応していない内容が多く含まれる回答となってしまう可能性がある。

3. 提案手法

本節では、相談文を入力し回答文を出力する提案システムについて述べる。既存研究においては、以下の

問題点が存在した。

- 検索エンジンによりWeb上の情報を取得し回答に用いている[4]が、相談の回答のような内容は一般のWebページには多く存在せず、回答を見つけるには適さない。
- 回答の情報が不足である場合と過剰である場合を区別していない[5]ため、入力に対して関係のない情報が出力されてしまう。
- 語の重要度を考慮せず回答の多様化を行っている[4][6]ため、回答の要点でない部分で多様化された回答が出力される場合がある。

これらを踏まえて提案手法では、

- 質問文（入力）をもとに、単純な内容の類似相談文をみつけ、その回答を利用することで関係のない情報の出力を抑える
- 回答文の内容を端的に表現するキーワードにより、回答の要点による回答選択を行う。

とした。

本研究では、過去の相談・回答文として「教えてgoo!³」の恋愛相談カテゴリからデータを収集し、利用した。収集したデータは、相談文10,645件、回答文49,456件である。

提案手法は以下の流れをとる。

step 1. 過去の相談・回答文データセットからキーワードを抽出

step 2. 入力質問文に対する回答候補取得

step 2.1. 入力に対する類似相談文の検索

step 2.2. 2.1.で取得した類似相談文に紐づく回答を回答候補とする

step 3. 出力する回答の選択

step 3.1. 回答候補のクラスタリングを行う

step 3.2. 回答候補に対し文章の肯定度をスコア付けする

step 4. 最終的な出力

以下、各々のstepについて順に説明する。

3.1. キーワード辞書の構築

キーワード辞書の構築では、相談文検索に用いる相談文キーワード辞書と、回答選択に用いる回答文キーワード辞書の構築を行う。

本研究では、相談文に付与される投稿タイトルに、相談文の内容を端的に表現するキーワードが含まれていると考える。そこで、投稿タイトルから形態素解析器Juman[8]を用いて名詞・形容詞・動詞を抽出してキーワードと定義し、キーワード辞書を作成した。Juman

³ <http://oshiete.goo.ne.jp/>

の辞書は Juman7.0.1 のものを用いた。また、抽出された単語から一般的な語を除去する。一般語の除去には以下 2 つの処理を行った。

- 数字のみの語の除去
- idf 値による除去

idf 値による除去では、まず「投稿タイトルとして一般的な語ではなく、日本語として一般的な語の除去」を行いたい。そこで、Wikipedia 記事 1,119,485 件に対する単語の文書頻度 $df(w)$ を利用して idf 値を計算する。単語 w の idf 値は以下の式を用いて計算した。

$$idf(w) = \log_e \frac{N}{df(w)} \quad (1)$$

ただし、 N は Wikipedia の記事数である。idf 値が 2.0 以下の単語を一般語として、キーワード辞書から除去した。

次に、投稿タイトルにおける文書頻度が非常に高い・低いため、文書を特定することに適さない単語の除去を行う。式(1)による idf の式において、 N を投稿タイトル数として、

$$\begin{cases} idf(w) < 2.0 \\ idf(w) > 5.0 \end{cases} \quad (2)$$

となる単語を除去した。これらのしきい値は、値を変化させていき、結果を人目で確認することによって決定した。これらの操作後に生成されたキーワード例を表 1 に示す。また、抽出されたキーワードは 881 個であった。

次に、回答文キーワード辞書の構築について述べる。相談文キーワード辞書の構築では相談投稿タイトルの情報を用いたが、回答文にはタイトルが存在しない。そのため、回答文の前半部分に回答の要点が含まれるとして、回答文先頭の 2 文から回答文キーワード辞書を構築する。キーワード語の抽出法、一般語の除去法は相談文キーワード辞書構築と同様に行う。ただし、Wikipedia における idf 値のしきい値は 2.0 とした。また、投稿タイトルにおける文書頻度の代わりに、回答文全文における idf 値を用いて、

$$\begin{cases} idf(w) < 1.6 \\ idf(w) > 5.2 \end{cases} \quad (3)$$

となる単語の除去を行った。これらのしきい値は、同様に結果を人目で確認することによって決定した。生成されたキーワード例を表 2 に示す。また、抽出されたキーワードは 1,090 個であった。

表 1 相談文キーワード例

相談文キーワード
会社, 卒業, 高校, 友人, 年上, 片思い, 後悔

表 2 回答文キーワード例

回答文キーワード
変だ, 抱える, 詫びる, 普通だ, 勘違い, 立派だ

3.2. 類似相談文検索

類似相談文の検索では、3.1 項で抽出したキーワードを用いる。類似相談文を検索する際、被検索相談文に余分な要素が含まれていると、紐づく回答にも意図しない余分な要素が多く含まれてしまうことが考えられる。そこで、提案手法では、まず入力相談文 $query$ に含まれるキーワードの集合を K_{input} 、被検索データセット中の相談文 $querydb$ に含まれるキーワードの集合を $K_{querydb}$ とする。このとき K_{query} に含まれているが $K_{querydb}$ に含まれていない語は許容するが、 K_{input} に含まれておらず $K_{querydb}$ に含まれる語があった場合は特別なペナルティを与える。これを踏まえて、以下の式(3)で類似相談文のスコア付けを行う。ここで、 α はペナルティ項の重みパラメータである。

$$\begin{aligned} score_{querydb}(query) &= \frac{|K_{query} \cap K_{querydb}|}{|K_{query}|} \\ &* \left(1 - \alpha \frac{|K_{query} \cap K_{querydb}|}{|K_{querydb}|} \right) \end{aligned} \quad (3)$$

ただし、 $\overline{K_{query}}$ はキーワード辞書全体 K に含まれているが、 K_{query} には含まれていない語の集合である。

式(1)は、($querydb$ が $query$ のキーワードを網羅しているか) * ($querydb$ に余分なキーワードが含まれていないか) という形式をしている。

3.3. 回答クラスタリング

先行研究の質問応答システムでは、類似する回答による出力の冗長性を解消するために、回答候補をクラスタリングし、出力を決定する[4]。提案手法もそれにならない、相談者に返す回答に多様性をもたせるためにクラスタリングを行う。同一クラスタになった回答は同様の要点を持つ回答であるとし、後の処理でクラスタ内の代表となる回答のみを出力させる。クラスタリングには回答文 $answer$ 中のキーワード集合 K_{answer} を用いる。これは、キーワードが回答の要点を端的に表現するものであると考えられるためである。回答文 $answer_1$ 、 $answer_2$ の距離関数は Jaccard 係数を用いて以下の式(4)とした。また、クラスタリングのアルゴリズムは完全リンク法による階層的クラスタリングを用いた。

$$d(answer_1, answer_2) = 1 - \frac{|K_{answer_1} \cap K_{answer_2}|}{|K_{answer_1} \cup K_{answer_2}|} \quad (4)$$

3.4. 回答の肯定度によるスコア付け

相談者が相談文を投稿する際、相談者には既に期待する回答が存在する場合がある。本研究では「肯定的な回答」を相談者が期待する回答であるとして、このような回答に対し高いスコア付けを行う。このスコア付けには、熊本らの印象値計算手法[9]を用いる。熊本

らの手法では、文章の印象を事前に与えた印象語群を用いて、「楽しい⇔悲しい」といった印象尺度ごとに印象値を計算する。本システムでは、「肯定的⇔否定的」という印象尺度に手法を適用し、得られた印象値を肯定度スコアとする。熊本らの手法の流れは以下のようになる。

1. 印象尺度に対応する印象語群を設定する
2. 印象語群を用いて、文章中単語の印象値を計算し、印象辞書に登録する
3. 印象辞書の印象値を用いて、文章の印象値を計算する

肯定度スコアの計算には、事前情報として、「肯定的な印象語群 IW_L 」と「否定的な印象語群 IW_R 」が必要となる。これらの印象語群は、データセット中回答文に含まれる単語から文脈に依存する単語を除外し、文書頻度が一定以上であるものから選択した。選択された印象語群を表 3 に示す。

表 3 選択された印象語群

	印象語
肯定的な印象語群 IW_L	良い, 素敵, 素晴らしい, 正しい
否定的な印象語群 IW_R	悪い, ダメ, ひどい, 間違い

次に、単語 w の印象値 $v(w)$ を計算する。データセット中で、 IW_L を IW_R より多く含んでいる回答文の集合を S_L 、 IW_R を IW_L より多く含んでいる回答文の集合を S_R とする。そして S_L 、 S_R に含まれる助詞、連体詞、指示詞以外のすべての単語を抽出する。このとき、単語 w の S_L における文書頻度を $N_L(w)$ 、 S_R における文書頻度を $N_R(w)$ とする。また、 S_L の回答文数を N_L 、 S_R の回答文数を N_R とする。このとき単語 w の印象値 $v(w)$ は、以下の式(5)によって計算できる。

$$v(w) = \frac{P_L(w) * \log_{10} N_L}{P_L(w) * \log_{10} N_L + P_R(w) * \log_{10} N_R} \quad (5)$$

ただし、

$$P_L(w) = \frac{N_L(w)}{N_L} \quad (6)$$

$$P_R(w) = \frac{N_R(w)}{N_R} \quad (7)$$

である。

最後に、印象辞書の印象値を用いて、回答文の肯定度を計算する。文章の肯定度は、回答文中に含まれる単語の平均印象値とする。このとき、同じ単語が含まれていた場合も単純に足し合わせることで平均を計算する。

3.5. 最終的な出力の決定

3.3 項での各クラスタ内の回答に対し、3.4 項のスコア付けを適用する。各クラスタ内のスコア上位 n 件の回答をクラスタ内の代表回答として採用し、出力とす

る。つまり、クラスタ数が k であるとき、システムは $n \times k$ 個の回答を出力する。また、各出力回答内に含まれるキーワードをヒットした理由として利用者に提示する。

4. 評価実験

本研究は、入力された相談文に対して、(1)一般的な主観的回答を行う、(2)多様な主観的回答を行う、ことを目的としている。そのため、システムの最終的な出力が(1)、(2)にそった回答となっているかについて評価実験を行う。

4.1. データセット

収集した相談・回答データから 10 件の相談文と紐づく回答を選択し、テストデータとした。

4.2. 評価実験方法

システム全体の評価は、(1)一般的な主観的回答が出力されているか、(2)多様な主観的回答が出力されているか、の 2 点について評価実験を行う。(1)、(2)のいずれにおいても、回答選択におけるクラスタ数を 5、クラスタ内出力数は 1 とした。また、回答が適切なものであるかは考慮しない。

(1)について、各回答に対し、1(一般的でない)~5(一般的)の 5 段階で人手により回答の一般性スコアの評価を行う。ここで、一般的というのは、ある特定の相談に特化した回答でなく、他の類似した相談の回答となり得るものとする。類似相談文検索のベースラインとして、Jaccard 係数による類似相談文検索を比較対象とした。

(2)について、出力された回答リストに対し、要点が異なる回答の出力数を評価する。出力された回答に対して、人手により同様の観点の回答をグループにし、作られたグループ数を異なる要点の数とした。ベースラインとして、入力された相談文に実際に行われた回答(正解回答)の、異なる要点の数を比較対象とした。

4.3. 評価実験結果

回答の一般性についての評価実験結果を表 4 に示す。また、回答の要点数についての評価結果を表 5 に示す。

表 4 相談文に対する回答の一般性スコア

検索手法	平均スコア
Jaccard 係数	3.00
提案手法 $\alpha=1.0$	3.12

表 5 相談文に対する回答の要点数

回答選択手法	要点数
正解回答(10件) +人手まとめ	5.6
提案手法 (5件出力)	4.1

5. おわりに

本稿では、入力された相談文に対し出力を行うシステムを提案した。提案手法では、相談文のような長文で主観的である入力に対し、

- 入力よりも単純な類似相談を見つける
- 回答のクラスタリングにより、多様な回答を出力する
- 相談者が求める回答の一つとして、肯定的な回答を出力する

という処理を行った。

恋愛相談カテゴリを対象にした評価実験の結果、一般的な回答の出力では、提案手法がベースラインより回答の一般性を向上させ、提案手法の有効性を示した。多様な回答の出力では、提案手法で回答を5件出力したうち、重複したのは平均1件未満となった。

今後の課題として、相談文解析の詳細化、正答を出力するための回答選択が挙げられる。

参 考 文 献

- [1] Hang Cui, Min-Yen Kan and Tat-Seng Chua: "Generic soft pattern models for definitional question answering," Proceedings of the 28th annual international ACM SIGIR conference on Research and Development in Information Retrieval, pp.384-391, 2005.
- [2] Ryuichiro Higashinaka and Hideki Isozaki: "Corpus-based Question Answering for why-Questions," Proceedings of IJCNLP, pp.418-425, 2008.
- [3] 佐々木智, 藤井敦: "取るべき行動を理由と共に答える質問応答システム-how型とwhy型の統合", 言語処理学会第15回年次大会発表論文集, pp.36-39, 2009.
- [4] 石下円香, 佐藤充, 森辰則: "Web 文書を対象とした質問の型に依らない質問応答手法." 人工知能学会論文誌, 24 卷 4 号, pp. 339-350, 2009.
- [5] Snigdha Chaturvedi, Vittorio Castelli, Radu Florian, Ramesh M. Nallapati and Hema Raghavan: "Joint question clustering and relevance prediction for open domain non-factoid question answering," Proceedings of the 23rd International Conference on World Wide Web, 2014.
- [6] Palakorn Achananuparp, Xiaohua Hu, Tingting He, Christopher C. Yang, Yuan An and Lifan Guo: "Answer diversification for complex question answering on the web," Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, pp.375-382, 2010.
- [7] Kotaro Sakamoto, Keiichi Nagao, Hayato Kobayashi, Hideyuki Shibuki, Tatsunori Mori and Noriko Kando: "Re-ranking answer candidates based on exhaustiveness of variety of answer viewpoints in

non-factoid QA," Proceedings of the SIGIR Workshop on Web Question Answering: Beyond Factoids, 2015.

- [8] JUMAN - KUROHASHI-KAWAHARA LAB, <http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?JUMAN>.
- [9] 熊本忠彦, 河合由起子, 田中克己: "新聞記事を対象とするテキスト印象マイニング手法の設計と評価", 電子情報通信学会論文誌, Vol.J94-D No.3, pp.540-548, 2011.