

QA サイトにおける質問難易度を用いた最適な回答者提示

堀江 将隆 山本 和英

長岡技術科学大学 電気系

{horie, yamamoto}@jnlp.org

1 はじめに

近年、Yahoo!知恵袋¹ など、コミュニティベースの Q&A サイト (QA サイト) の利用者が増加している。QA サイトとは、コミュニティのユーザが疑問に思っていることや悩み等の質問を投稿し、他のユーザが回答することで知識を共有できる Web 上のサービスである。Web 検索サイトでのキーワード検索と違い、文章を用いた柔軟な検索が可能であり、需要が高いと言える。

QA サイトでは疑問を持っているユーザが質問を投稿し、その質問に対して回答可能な他のユーザが回答を投稿するので、1 つの質問につき 1 件以上の回答が得られる。しかし、質問者が知らない知識について質問しているので、回答を得られたとしてもそれが回答として正しいのか、また得られた回答の中でどれが疑問を解決できる回答として正しいのか判断できないことが考えられる。従って、現在の QA サイトでは質問者の疑問や悩みを十分に解決できていないと考える。

そこで我々は回答者の中から、質問に最適な回答を投稿できる回答者を選出し提示することを目的とする。本研究では Yahoo!知恵袋の質問回答を対象に、ある質問に対してベストアンサー (以下、BA) に選ばれたユーザを最適な回答者と仮定し、そのユーザを選出する手法を提案する。

まず、質問文の難易度情報を定義し、これを自動で推定する手法を提案する。次に対象の質問文、各回答者の回答履歴の質問文の質問難易度情報を使用し、最適な回答者を選出する手法を提案する。

2 関連研究

関連研究として、QA サイトにおけるベストアンサー推定や、質問推薦、質問分類等が挙げられる。ベストアンサー推定では石川ら [1] が、良質回答の特徴を考察し、最終的に重要な特徴 3 種類を素性として使用し機械学習で BA を推定した。

渡邊ら [2] は QA サイトにおける質問推薦の為に、質問分類する手法を提案している。初めに質問を正解ありと正解なしの 2 つに分類し、最終的に 5 種類の質問タイプに分類する事を目的としている。各品詞の単語と、質問タイプの分類に有効と思われる単語を素性として機械学習で質問分類を行っており、特定の単語が有効である点や、教師データが少ない、質問分類の難しさが要因で分類精度が高くなかったと考察している。我々の最終的な目的は最適な回答者を提示することであり、その為に質問情報や回答者情報を用いるのでこれらの研究とは異なる。

3 質問難易度

過去に我々 [3] は、質問文に含まれる専門用語を必要な知識と考えて、最適な回答者を選択する手法を提案したが精度は低かった。そこで、本研究では文の表層以外の情報を用いることを考えた。

すなわち、直感的には判断可能であるが、機械的には判断が難しいと思われる質問文への回答のしやすさを表す質問難易度情報を用いて、最適な回答者を探す手法を提案する。

質問難易度とは、質問に回答しやすいかを表す情報である。この質問難易度を推定することで、最適な回答者を見つける手がかりになると考えた。例えば、対象の質問の難易度が低い場合、そのような低難易度の質問に多く回答している回答者のほうが良い回答を期待できる。何故なら、低難易度の質問をするユーザーは、深い知識を求めているのではなく、簡単に理解できるような易しい回答を求めているからである。また高難易度の質問に多く回答しているユーザのほうが、高難易度の質問に合った回答が可能だと考えた。このように仮定した上で、質問難易度を推定できれば有用な情報となり得ると考えた。

3.1 質問難易度の定義

本研究における質問難易度を決定する要素を以下のように定義する。

- (1) 質問者が求めている回答の抽象度
- (2) 質問に必要な知識
- (3) 質問文の情報の不十分さ

質問者が求めている回答の抽象度は、質問者の求めている回答が抽象的かどうかを表す指標で、例えば特定の人名を問う質問のように回答が 1 つに定まっている場合は抽象度が低く、回答者は回答しやすい。一方、求めている回答の候補が複数ある場合、様々な情報を付加して質問者が納得するような回答をする必要がある為、回答しにくいと言える。従って抽象度が高いほど質問難易度は高い。

質問に必要な知識は、求めている回答はどの程度の知識を持っていれば答えられるのかを表す要素である。誰もが知っているような話題についての質問であれば回答しやすい、あまり知られていない話題の質問であれば知っている人しか回答できないので回答しにくい。従って、質問の話題があまり知られていない知識を要するほど質問難易度は高いと言える。

質問文情報の不十分さとは、単純で曖昧な質問よりも、回答に関する詳細な条件が設定されており、その回答を導きやすい質問文のほうが、回答者にとって答

¹Yahoo!知恵袋. <http://chiebukuro.yahoo.co.jp/>

えやすい。従って、質問文が不十分なほど質問難易度は高いと言える。

以上より、3つの要素が質問難易度を決定する要素であると考えた。

4 質問難易度推定手法

初めに、質問難易度情報の要素を推定する手法を提案する。定義したように要素は3種類あるが、質問難易度を決定するのに最も関係している要素を選択する予備実験を行った。その結果、質問が求めている回答の抽象度(以下、抽象度)が重要であることが分かり、この要素を自動で推定する実験を試みた。他の2つの要素は簡単な実験により、情報を抽出する。

4.1 抽象度の定義

抽象度の定義は、渡邊ら[2]の質問タイプを参考に定義した。質問タイプと同様の定義にしなかった理由は、求めている回答の抽象度合いを表しており、質問タイプとは完全に一致しないと考えた為である。抽象度は3段階と設定した。多く分類してしまうと、分類基準が複雑になり、人手での分類が難しい。定義と例文を以下に示す。

抽象度1 回答候補が1種類で、事象の定義や説明、客観的な理由を問う質問

例文 TVゲーム「テトリス」を考えたのはどこの国の人?

抽象度2 回答候補が複数で、回答者の経験等から回答する質問、手法や解決方法を問う質問

例文 MDに入った曲をCDに録音する方法はありますか?

抽象度3 回答候補が複数で、情報提供を依頼する質問や、推測等の主観的に回答する質問

例文 今年4月から開始されるアニメでどれが一番お勧めですか?

抽象度1は、回答候補が1種類の質問である為、求めている回答の抽象度は低い。抽象度2と抽象度3は、回答候補が複数の質問であるが、抽象度2はある程度求めている回答の種類は絞られており、抽象度3の大量の回答候補よりは抽象度が低いと考えた。

4.2 予備実験

抽象度が実際の回答とどのような関係であるかを調べる。10人の回答者が過去に回答したそれぞれ100件の質問、合計1000件に対して人手で抽象度を設定し、実際の回答と関係あるかを調査した。回答者10人は、5つのカテゴリにおいてカテゴリマスターを2人ずつ選出した。カテゴリマスターとは、カテゴリごとの回答数やBAに選ばれた割合を考慮したランキング上位のユーザである。同カテゴリから2人選んだ理由は、カテゴリにおいて抽象度の偏りの違いがあるかを調査する為である。

質問は、3件以上の回答が得られたものに限定する。これは、1件や2件の回答からBAに選ばれて、BA率が高かったとしても信頼性が低い回答であると考えた為である。抽象度は定義に従い判定者1名が分類する。

各ユーザの抽象度ごとの質問数、BA数、BA以外の質問(NA)数、BA率(=BA数/質問数)を算出した。例を表1に示す。質問数が最も多い抽象度を Abs_{max} 、質問数が最も少ない抽象度を Abs_{min} と表す。表1の

表 1: 回答者の抽象度ごとの質問数とBA率

抽象度	質問数	BA	NA	BA率
1	80	32	48	0.40
2	14	3	11	0.21
3	5	1	4	0.20
合計	99	36	63	0.36

場合、質問数が80件の抽象度1が Abs_{max} 、質問数が5件の抽象度3が Abs_{min} に該当する。

抽象度と質問数の関係について調べる。結果より、偏った抽象度の質問に回答しているユーザが多いことが分かった。 Abs_{max} の質問数と、 Abs_{min} の質問数との差が2倍以上あるユーザは10人中9人であった。従って、ほとんどのユーザが一定の抽象度の質問に偏って回答していることが分かる。理由として同じカテゴリの質問に多く回答している為とも考えられる。

栗山ら[4]は、質問タイプを客観的な回答が得られる質問の情報検索型、適切な回答の基準がなく回答者の経験・価値観によって回答する質問の社会調査型と大きく2タイプに分類しており、これらはカテゴリに依存するものだと分かっている。この為、カテゴリに依存して偏ったのかを調べるため、同カテゴリのユーザ2人を比較する。その結果、各ユーザの Abs_{max} が一致したのが5カテゴリ中1カテゴリだけであった。従って、カテゴリ間で多少の偏りはあるが、ユーザごとに特定の抽象度に回答しやすい傾向があると言える。

次に抽象度とBA率の関係について調べる。 Abs_{max} のBA率が、他の抽象度のBA率よりも高い場合、回答者は多く回答している抽象度の質問は得意であり、 Abs_{min} のBA率が、他の抽象度の質問より低ければ、普段回答しない抽象度の質問は得意ではないと言える。結果より、 Abs_{max} のBA率が最大のユーザが4人、最小のユーザが1人、 Abs_{min} のBA率が最大のユーザが2人、最小のユーザが5人であった。つまり、回答数が多い抽象度の質問に回答した場合のBA率が他より高く、回答数が少ない抽象度の質問に回答した場合はBA率が低いということである。

この結果より、対象の質問に適した抽象度の質問に多く回答している回答者は、適していない抽象度の質問に多く回答している回答者よりも、良い回答を出せると言える。

4.3 抽象度推定手法

質問文を定義した3段階の抽象度に分類することを考える。

初めに抽象度ごとの単語の出現割合リストを作成する。出現割合リストとは、各抽象度における単語の出現数と、その出現割合で構成されており、それぞれ以下のように算出する。

- 各抽象度での単語の質問文ごとの出現確率

$$P_{abs_i}(w) = f_{abs_i}(w) / Q_{abs_i}$$

- 出現確率の各抽象度に占める割合

$$R_{abs_i}(w) = P_{abs_i}(w) / \sum_{i=1}^3 P_{abs_i}(w)$$

- 出現割合リストの構成

$$w, f_{abs_i}(w), R_{abs_i}(w)$$

abs は抽象度、 w は単語、 Q は質問数、 f は出現頻度、 P は出現確率、 R は割合を表す

出現割合リストの例を表2に示す。この出現割合リストを使用して、質問文に含まれる単語がどの抽象度に多く属しているかで対象の質問の抽象度を決定する。

表 2: 抽象度ごとの単語の出現割合リストの例

	f_{abs_1}	R_{abs_1}	f_{abs_2}	R_{abs_2}	f_{abs_3}	R_{abs_3}
名前	28	0.812	1	0.045	4	0.143
対処	1	0.035	16	0.878	2	0.087
好き	4	0.054	3	0.063	53	0.883

表 3: 文章表現辞書

抽象度 1	は+(どうい どのような どんな) +(意味 いみ)+(でしょうか ですか)
抽象度 2	(どうしたら どうすれば) +(いい 良い)+(でしょうか ですか)
抽象度 3	(どう どのように) +(思いますか 思われますか)

抽象度判定は出現割合リストを用いて以下の手順で行った。

- (1) 対象の質問文に対して形態素解析を行う。
- (2) 全単語に対して出現割合リストを参照し、対応する出現割合 $R_{abs_i}(w)$ を抽象度スコア $Score_{abs_i}$ に加算する。
- (3) 3つのスコア $Score_{abs_i}$ のうち、最も高いスコアの抽象度 abs_i を、質問文に付与する。

スコア改善のため、(2)において、 $R_{abs_i}(w)$ を以下のように設定する。

- ・出現割合 $R(w)_{abs_1}$ が閾値 R_{th} 以上の時、 $R(w)_{abs_i}$ と出現頻度 $f(w_{abs_i})$ の積とする。
- ・ $R(w)_{abs_i}$ と重み $weight$ の積とする。

4.4 文章表現辞書を用いた手法

予備実験において抽象度を人手で判断した結果、特徴的な単語単体を見て抽象度を判断するというより、文章表現から特定の抽象度であると判断する場合が多かった。従って、判断する為の基準となった質問内容を含む文の文末表現と、よく出現すると感じた特定の表現を用いて質問の抽象度を判断することを考える。

Yahoo!知恵袋の質問から、各分類においてよく出現し抽象度を判断できると考えた文章表現を人手で抽出し、各抽象度における文章表現辞書を作成した。また、抽出した文章表現を、似たような文章表現にも照合するように汎化した。例を表 3 に示す。

文章表現辞書中の表現が対象の質問文に含まれている場合、質問文に対応する抽象度と判定する。ただし、文章表現辞書中の表現を含んでいない質問も存在するので、抽象度を付与できなかった質問に対して、前節の抽象度別出現割合リストを用いた手法を適用する。

4.5 評価実験及び考察

使用するデータは、Yahoo!知恵袋データ「第二版」²の全カテゴリからランダムに選択した 1900 件を用いる。この 1900 件のデータにはあらかじめ人手で抽象度をタグ付けしている。1900 件のデータを、準備の為のデータセット A(950 件) と実験の為のデータセット B(950 件) に分ける。

データセット A を用いて、抽象度別出現割合リストと文章表現辞書を作成する。初めにデータセット A を用いて実験を行い、その結果から追加条件の為、精度が高くなるような単語出現割合の閾値 R_{th} と抽

表 4: 抽象度推定実験結果

実験データ	データセット A	データセット B
追加条件無し	70.6%	48.2%
追加条件設定	88.3%	60.3%
+文章表現	86.4%	63.2%

象度の重み $weight$ を決定する。抽象度推定実験にはデータセット B を用いる。

実験は、出現割合リストを使用した手法と、文章表現辞書を使用した手法に出現割合リストを用いた手法を適用した。まず、文章表現辞書を用いた手法をデータセット A に対して行い、精度の良い文章表現のみを選択し新たに辞書を構築する。この辞書を用いて実験し、抽象度を付与できなかった残りの質問には出現割合リストを使用した手法を適用する。質問文に付与されている抽象度タグと、出力した抽象度タグが一致すれば正解とし、精度を以下の式で算出する。

$$\text{精度} = \text{正解数} / \text{質問数}$$

追加条件を設定せずに実験したものと、最適な R_{th} 、 $weight$ を設定して実験したものと、文章表現を用いた手法の結果を表 4 に示す。

結果より、重みづけを行うことで精度が 12 % 向上した。出現割合リスト手法において、抽象度ごとに出現割合が高い語の数にばらつきがあり、抽象度 2 では出現割合が高い語が多く、抽象度 1 では少なかった。これより、抽象度 1 のスコアが大きくなるように $weight$ を設定することで精度向上へと繋がったと考えられる。

文章表現を用いた実験では、抽象度を判断可能な表現のみを使用した為、抽象度を付与した質問に対しての精度は 78 % と高い。しかし、照合する表現が少なくなる為、抽象度を付与できた質問数は少なかった。高精度で抽象度を付与した質問と、抽象度を付与できなかった残りの質問に対して、出現割合リストを用いた手法を適用し組み合わせることで精度を 3 % 向上することができた。

出現割合リストを用いた実験時のスコア付けの詳細を調べた結果、各抽象度のスコアの差が大きい場合は全体の精度と比較して、高い精度で抽象度を判断できていることが分かった。これより、特徴が大きい語によって抽象度は判断できると言える。

単語数が多い質問に対して正しく判断できていないことが多かった。これは関係のない単語が邪魔をしていて、全体のスコアが大きくなってしまった為と考えられる。また、2 つ以上の質問内容を含んでいる質問に対して正しく判断できていないことが分かった。これは 2 つの定義を満たしている場合があるので、人手で判断する際には重要そうな質問に関する抽象度を設定した為、自動で判断するのは難しいことが分かった。

5 回答者選択実験

5.1 手法

質問と、回答者が普段多く回答している質問の質問難易度が近い場合、良い回答を投稿できると考えた。質問の難易度情報の 3 要素を、以下のようにスコアとして表現する。

²Yahoo!知恵袋データ「第二版」．国立情報学研究所

表 5: 回答者選択実験

	ランダム	抽象度	内容語	検索 hit
精度	27.8%	28.5%	27.9%	26.9%

抽象度: 出現割合リストを用いた手法で判定した値

重要な専門用語の検索ヒット数: 質問文に含まれる最も重要な専門用語の検索ヒット数
・専門用語は、TermExtract³ で、最も重要と出力された語を使用し、検索ヒット数は Yahoo! API のウェブ検索⁴で取得した。

質問文の情報の不十分さ: 質問文に含まれる内容語数
・内容語は ChaSen⁵を使用して抽出

質問難易度情報を用いた手法の手順を以下に示す。

- (1) 回答者が過去に回答した全ての質問に対して、それぞれのスコアを算出する。
- (2) 抽象度の場合最も頻度の高い抽象度とし、他の 2 つはそれぞれのスコアの平均を回答者の質問難易度情報とする。
- (3) 対象の質問文の質問難易度情報スコアをそれぞれ算出する。
- (4) 質問文のスコアと回答者のスコアとの差が最も小さい回答者を最適な回答者として出力する。
回答者候補で同じ抽象度を持っている場合、複数の回答者の中で抽象度の質問の割合が高い回答者を選択する。

5.2 評価実験

使用するデータは、Yahoo!知恵袋データの質問文 1000 件と回答した複数の回答者 ID とその回答履歴である。データセットには回答者が 3 人から 5 人の質問のみ使用する。平均回答者数は 3.74 人である。質問難易度情報の要素それぞれのスコアを用いた実験を行った。選択したユーザが BA を回答したユーザである場合を正解とし、精度を以下の式で算出する。結果を表 5 に示す。

精度 = 正解数 / 質問数

5.3 考察

表 5 から、全てにおいてランダムとほぼ同程度の精度であり、本稿での質問難易度情報を使った手法では最適な回答者選択ができていないことが分かった。実際に抽象度がどのように影響しているかを調べた結果、実験データの質問と BA の回答者の抽象度が一致しているものが 1000 件中 511 件あることを確認した。ランダムに抽象度を選択して、BA 回答者の抽象度と一致する確率は 3 分の 1 なので、511 件一致したということは、最適な回答者を選択することに抽象度が有効であると言える。しかし、質問に対して抽象度が一致している回答者が複数存在する質問が、351 件ある為

³TermExtract.

<http://genshen.dl.itc.u-tokyo.ac.jp/termextract.html>

⁴Yahoo!API ウェブ検索. <http://developer.yahoo.co.jp/>

⁵日本語形態素解析システム「茶釜」.

<http://chasen.naist.jp/hiki/ChaSen/>

そこから絞ることができなかったのが精度が低かった要因であると考えた。この時、BA 回答者とその他の回答者の抽象度が一致している回答者数の平均は 2.6 人である。

提案手法で、最も多い抽象度の質問を回答した回答者を選択した後、同じ抽象度の回答者から一人に決定する際に、過去の質問におけるその抽象度の割合で最終的に決定した。この手法を改善することにより 511 件に対しての精度は向上できると考えた。例えば、比率のみではなく質問数も考慮する手法が挙げられる。

また、BA 回答者の平均順位を算出した結果、2.3 位であった。平均回答者数が 3.74 人であるのでランダムに選択した場合、1.85 位になる。さらに、質問と BA 回答者の抽象度の一致が 511 件あることから、本手法は特定の質問に対しては最適な回答者を提示できるが、質問全体に対しては有効ではないと言える。特定の質問とは、抽象度に依存している回答者が多い質問であり、抽象度とは関係なく好きな分野の様々な質問に回答する回答者が多い質問に対しては、本手法は適用できないと考えた。

6 おわりに

本研究では、最適な回答者を提示するために質問難易度情報を用いる手法を提案した。質問難易度情報として抽象度、必要な知識、情報の不十分さの 3 つの要素を定義し、重要であると考えた抽象度を自動で推定する実験を行った。単語の出現割合と人手で選択した文章表現を使用し推定した結果、60%とランダムで選択した場合の約 2 倍の精度を得ることができた。

そしてこの結果から質問難易度情報を用いて、対象の質問に最適な回答者を提示する手法を提案した。結果より、精度が低いことから本研究の手法が有効であるとは言えない結果となった。しかし、解析結果から抽象度が最適な回答者を選択する為に有益であることが分かった。

今後の課題は、本研究で簡易に実験した質問難易度情報の 2 つの要素について考察することと、質問難易度情報を有効的に使用する手法の提案が挙げられる。

謝辞

本研究において、ヤフー株式会社が国立情報学研究所に提供した「Yahoo! 知恵袋データ (第 2 版)」を利用致しました。

参考文献

- [1] 石川大介, 栗山和子, 酒井哲也, 関洋平, 神門典子. QA サイトにおけるベストアンサー推定の分析とその機械学習への応用. 情報知識学会誌, Vol. 20, No. 2, 2010, pp.73-85.
- [2] 渡邊直人, 島田諭, 関洋平, 神門典子, 佐藤哲司. QA コミュニティにおける質問者の期待に基づく質問分類に関する一検討. データ工学と情報マネジメントに関するフォーラム 2011, 2011.
- [3] 堀江将隆, 山本和英. Q&A サイトにおける専門用語を用いた最適な回答者提示. 自然言語処理学会第 17 回年次大会, pp228-231, 2011.
- [4] 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, Vol. 2009-FI-95, No. 19, pp.1-8, 2009.