

# 質問応答サイトにおける発言間の関係同定

Identification of relations between utterances in question answering communities

神保 一樹\*<sup>1</sup>    高村 大也\*<sup>2</sup>    奥村 学\*<sup>2</sup>  
Kazuki Jimbo    Hiroya Takamura    Manabu Okumura

\*<sup>1</sup>東京工業大学 総合理工学研究科 知能システム科学専攻  
Tokyo Institute of Technology, Department of Computational Intelligence and Systems Science

\*<sup>2</sup>東京工業大学 精密工学研究所  
Tokyo Institute of Technology, Precision and Intelligence Laboratory

Question answering communities on the web contain many questions and their answers allowing everyone to read them. There is a problem on these communities that some questions have too many answers for users to read them all. In this paper, we focus on relations between utterances in question answering communities to solve this problem. We define types of the relations between utterances and identify which type of relation each utterance pair has. We show that features which contain information of question types improve the performance of relation identification. Then we propose a method for summarization of answers in question answering communities. We show that relations between utterances improve the performance of summarization.

## 1. はじめに

Web 上には、ユーザからの質問に対して他のユーザが答えしていく、質問応答サイトと呼ばれるサイト (Yahoo!知恵袋\*<sup>1</sup> など) がある。質問応答サイトでは、あるユーザが質問を書き込むと、その質問が先頭となるスレッドが作成され、他のユーザは、そのスレッドに回答を次々と書き込める。回答の書き込まれたスレッドは保存され、誰でも閲覧することができる。

質問応答サイトでは、多くのユーザによるいろいろな観点からの回答が得られる反面、回答数の多いスレッドでは、ユーザにとってすべての回答を読むのが困難となる。そこで、回答内容の類似・対立などの関係に基づいてスレッド全体の構造をとらえれば、スレッドの内容をユーザにとってよりわかりやすく提示することができると思われる。

本研究では、質問応答サイトの各発言間の関係を同定し、さらに、発言間の関係の情報を利用して、スレッドの回答内容の要約を行う。まず、質問応答サイトの 2 発言間に成立する関係の種類を定義し、機械学習の手法を用いて、実際のデータ上で 2 発言間の関係の種類を同定する。発言間の関係は、質問・回答間について 2 種類、回答・回答間について 4 種類定義する。質問応答サイトの発言間の関係を同定するにあたり、素性として質問タイプの情報や質問の核文の情報を取り入れる。そして、同定した関係を用いて、質問応答サイトの内容を抜粋して要約する。発言間の関係の情報を利用することで、利用しない場合に比べてより良い要約を生成できることを示す。

## 2. 関連研究

質問応答サイトに関する研究は、幅広く行われている。栗山ら [7] は、質問応答サイトの質問と回答を分析した。質問が属する質問応答サイト内のカテゴリによって、頻出する質問タイプの傾向やベストアンサーの特徴が異なることを示した。Ding

ら [8] は、質問応答サイトの書き込みの中から、質問の背景を表す文脈、および質問に対する回答を Conditional Random Field を用いて抽出する手法を提案した。質問の背景の情報が、回答抽出の性能向上に貢献することを示した。Liu ら [2] は、質問や回答を分類する枠組みを提案し、これにより質問応答サイトの回答を自動で要約する手法を提案した。この研究では英語を対象としている。また、質問や回答を自動で分類する手法についてはふれていない。

田村ら [1] は、複数文からなる質問に対して、質問が求める回答の種類を表す質問タイプを同定する手法を提案した。答えを要求するもっとも重要な 1 文 (核文) のみに着目することが、質問タイプ同定の性能向上に効果的であることを示した。質問タイプは、質問が求める回答の種類を表すので、質問文と回答文の内容の関係を調べるために役立つと考えられる。

Radev [3] は、複数文書間に成り立つ関係を考える枠組みとして、Cross-document Structure Theory (CST) を提案した。この枠組みでは、同一・矛盾・続報など 24 種類の関係を定義している。ただし、ここで定義された関係は主に新聞記事などを想定したものであるため、質問応答サイトなどのコミュニティの発言間にそのまま適用することはできない。

## 3. 関係同定

### 3.1 問題設定

本研究で扱う発言間の関係は、同一スレッドに含まれる 2 発言間の関係である。ここで、スレッドとは、1 つの質問発言とそれに対する任意の個数の回答発言からなる集合である。

質問応答サイトの 2 発言間の関係を考えるとき、(1) 2 発言の一方が質問発言である場合 (「質問・回答」の組) と (2) 2 発言とも回答発言である場合 (「回答・回答」の組) とでは、発言間に存在しうる関係の種類が異なる。そこで、これら 2 通りの場合を別々の問題として扱う。

### 3.2 関係の種類

質問応答サイトのスレッドを要約するにあたって、発言間の関係の情報を利用することを考える。そこで本研究では、

連絡先: 神保一樹, 東京工業大学 総合理工学研究科 知能システム科学専攻 奥村研究室, 神奈川県横浜市緑区長津田町 4259 R2-728, 045-924-5295, jinbo@lr.pi.titech.ac.jp

\*1 <http://chiebukuro.yahoo.co.jp/>

Radev[3] が定義した文書間の関係の種類を参考にして、質問応答サイトにおける発言間の関係の種類を定義する。

質問・回答間の関係としては、以下の 2 種類を定義する。

直接回答 質問者の求める内容に対する直接の回答である

その他 上記に当てはまらない回答である

回答の発言の中には、話を横道にそらしたり、単に思いついたことを書いたりするものもある。このような発言は質問に対する回答としては有益でない。ある回答が質問に対する直接の回答であるか否かを判定することで、有益な回答とそうでない回答を分けることができると考えられる。なお、Radev の定義した関係は、新聞等の文書を対象としているため、質問に回答するという関係は含まれていない。

回答・回答間の関係としては、以下の 4 種類を定義する。

類似 質問に対する回答としての趣旨が類似する

情報追加 前になされた発言に含まれない情報を追加して述べる

対立 質問に対する回答としての趣旨が大きく異なる

その他 上記のいずれにも当てはまらない回答である

これらの関係を同定できれば、例えば、類似した回答が重複しないように発言を選ぶことにより、良い要約を生成できるようになると考えられる。ここで、「類似」は Radev らの「Equivalence (paraphrasing)」や「Agreement」、 「情報追加」は「Elaboration」や「Refinement」、 「対立」は「Contradiction」にそれぞれ対応する。

#### 4. 提案手法

本章では、関係同定の手法、およびそれを用いて要約を行う手法について、順に説明する。

##### 4.1 関係同定

発言間の関係同定は、二値分類器の一種であるサポートベクトルマシン (SVM) [5] を用いて行う。同一スレッド内のすべての 2 発言の組をそれぞれ 1 事例とし、one-vs-rest 法および pairwise 法を用いて実験を行い、結果を比較する。

学習および評価に用いるデータを表す事例ベクトルは、同一スレッド内の 2 発言の組すべてに対して 1 つずつ生成する。分類に用いる素性は、表 1 のとおりである。

表中、種類欄に「発言単体」とあるものは、発言ごとに値が求まる。この種類の素性は、1 事例につき 2 発言分を別々に入れる。「発言ペア」とあるものは、発言ペアごとに値が求まる。この種類の素性は、1 事例につき 1 ペア分を入れる。「スレッド」とあるものは、スレッドごとに値が求まる。この種類の素性は、1 事例に対して、該当スレッドに対する値を入れる。

表中、備考欄に「\*」印がある素性については、回答・回答間の関係同定において、質問の情報も用いる。この場合、「発言ペア」の素性については、2 回答のペアのほか、先の回答と質問のペア、後の回答と質問のペアの計 3 ペアを考え、それぞれのペアについて素性を入れる。また、「発言単体」の素性については、それぞれの事例について、2 回答に質問を加えた 3 発言分の素性を入れる。

備考欄に「+」印がある素性では、田村ら [1] が定義した質問タイプと核文 (質問を構成する文の中で、答えを要求する最も重要な文) の情報を用いている。これらの同定には、田村ら

表 1: 関係同定に用いる素性の一覧

素性	種類	備考
発言が回答の先頭であるか 2 発言の発言順が連続しているか どちらの発言がより長い (文字数の比較) 2 発言の時間間隔 発言間で互いに反対語となる単語が存在するか 同一の URL が出現するか 質問タイプ	発言単体 発言ペア 発言ペア 発言ペア 発言ペア 発言ペア スレッド	* * * * +
unigram と bigram 単語 unigram 単語 bigram 各文の先頭 5 単語の unigram 各文の先頭 5 単語の bigram 各文の末尾 5 単語の unigram 各文の末尾 5 単語の bigram	発言単体 発言単体 発言単体 発言単体 発言単体 発言単体	* *
質問の核文・全文の注目名詞と 同じ意味カテゴリーの名詞 発言内に存在するか 発言間で同一のものが存在するか 発言間で異なるものが存在するか 質問の核文・全文の注目名詞と同一の名詞 発言内に存在するか 発言間で同一のものが存在するか	発言単体 発言ペア 発言ペア 発言単体 発言ペア	+ + + + +
固有表現 発言内に存在するか 発言間で同一のものが存在するか 発言間で異なるものが存在するか 質問タイプに対応する固有表現 発言内に存在するか 発言間で同一のものが存在するか 発言間で異なるものが存在するか	発言単体 発言ペア 発言ペア 発言単体 発言ペア 発言ペア	+ + +
かき括弧内の表現 発言間で同一のものが存在するか 発言間で異なるものが存在するか	発言ペア 発言ペア	
類似度と包含度 単語 unigram のコサイン類似度 単語 bigram のコサイン類似度 文節 unigram のコサイン類似度 単語の意味カテゴリー unigram のコサイン類似度 名詞のみの unigram のコサイン類似度 文ごとの単語 unigram コサイン類似度の最大値 名詞の包含度	発言ペア 発言ペア 発言ペア 発言ペア 発言ペア 発言ペア 発言ペア	* * * * * * *

の用いた SVM による分類手法を用いた。田村らの手法では本来、複数の項目を同時に尋ねる質問を対象外としているが、ここでは対象外の質問を除外せずに、そのまま手法を適用した。

意味カテゴリーはソーラスから獲得する。本研究では日本語語彙大系 [9] の意味体系を利用する。また、反対語の情報は角川類語新辞典 [10] に含まれる反対語情報を利用する。形態素解析、係り受け解析、固有表現抽出には CaboCha<sup>\*2</sup> を用いる。

##### 4.2 回答要約

発言間の関係を利用して、回答要約に採用する発言を選択する手順を、Algorithm 1 に示す。ここでは要約に採用する発言の組み合わせだけを求め、出力する順序は考えない。

まず (2 行目)、「質問に対する直接回答でない発言は、回答要約に入れるにはふさわしくない」という考えから、質問に対して「直接回答」でない発言を不採用とする。次に (3~10 行目)、「情報追加する発言の内容は情報追加される発言の内容を包含している」と仮定して、情報追加する発言 (集合 A) を要約に採用し、情報追加される発言 (集合 B) を不採用とする。このとき、「内容が類似する発言を複数採用するのは冗長である」「内容をもとに不採用と判断された発言と類似する発言は、同様に不採用とすべきである」との考えから、これらの発言と内容が類似し、対立する要素がない発言 (集合 C) を、同様に不採用とする。そして (11~16 行目)、「内容が類似する発言を回答要約に複数入れるのは冗長である」という考えから、発言を 1 つ選んで採用し、採用した発言と類似する発言 (集合 D) を不採用とすることを繰り返す。

\*2 <http://chasen.org/~taku/software/cabocha/>

**Algorithm 1** 要約生成アルゴリズム

```

1: すべての回答発言の採否は未定である．
2: 質問に対して「直接回答」でない(「その他」である)すべての発言を不採用とする．
3: 次の条件をすべて満たす発言の集合を  $A$  とする．
   • 前の発言に「情報追加」する
   • 後の発言から「情報追加」されない
4: while  $A$  のいずれかの発言の採否が未定である do
5:    $A$  に含まれかつ採否が未定である発言のうち任意の 1 つを選び、これを  $u$  とする．
6:    $u$  から「情報追加」される発言の集合を  $B$  とする．
7:   次の条件をすべて満たす発言の集合を  $C$  とする．
     •  $B \cup \{u\}$  に含まれるいずれかの発言と「類似」の関係にある
     •  $B \cup \{u\}$  に含まれるいずれの発言とも「対立」の関係にない
8:    $u$  を要約に採用する．
9:    $B$  および  $C$  に含まれるすべての発言のうち、採否が未定であるものをすべて不採用とする．
10: end while
11: while 採否が未定である発言が存在する do
12:   採否が未定である発言のうち任意の 1 つを選び、これを  $u$  とする．
13:    $u$  と「類似」の関係にある発言の集合を  $D$  とする．
14:    $u$  を要約に採用する．
15:    $D$  に含まれる発言のうち、採否が未定であるものをすべて不採用とする．
16: end while
17: 採用された発言をすべてあわせて、要約とする．

```

表 2: 質問・回答間のタグの割合

関係	ペア数
直接回答	1524 (90.6%)
その他	159 (9.4%)
合計	1683

表 3: 回答・回答間のタグの割合

関係	ペア数
類似	635 (20.3%)
情報追加	526 (16.9%)
対立	1448 (46.4%)
その他	512 (16.4%)
合計	3121

## 5. 実験

関係同定の性能を評価し、さらに発言間の関係が回答要約の性能向上に役立つかどうかを調べるため、実験を行う。

### 5.1 データ

実験には、Yahoo!知恵袋のデータ [6] のうち、回答が複数書き込まれたスレッドのみを 500 スレッド用いる。1 スレッドあたりの平均回答数は 3.37 (最小 2, 最大 22) である。

関係同定の正解データ作成は、1 人で行った。最終的に各ペアに付いたタグの割合を、表 2,3 に示す。

回答要約の正解データは、これら 500 スレッドのうちランダムで選択した 50 スレッドについて作成した。作成にあたっては、2 名のアノテータが独立に、同一の 50 スレッドに対して、「なるべく少ない個数の発言で、質問に対する回答となる内容をもれなく網羅する」という基準のもとで、要約に採用すべき発言を選択した。この 2 名はいずれも、関係のタグ付けを行った人物とは異なる。内容が同等の発言が複数あつてどちらを選択しても良い、という場合もあるので、「回答 1 または回答 2」のような選び方も許した。

正解データの一致度を調べたところ、完全一致したものは 50 スレッド中 22 スレッドあつた。完全一致はしないものの、両方の正解データとともに正解と判断される発言の選び方が存

在するものは 7 スレッドあつた。また、2 つの正解データが共通する発言を全く含まないスレッドは存在しなかった。

回答要約の評価にあたっては、ここで得られた 2 つの正解データを別々に用いて、評価尺度の値を 2 通り求め、それらの値の平均をとって評価する。

### 5.2 実験設定

実験では、SVM の実装として TinySVM<sup>\*3</sup>を用いる。カーネル関数は、2 次の多項式カーネルを用いる。

5.1 節で説明した 500 スレッドのデータを、50 スレッドずつに 10 分割して、交差検定を行う。この分割では、スレッド数は等分されるが、事例ベクトル数は等分されない。

### 5.3 評価尺度

関係同定の性能評価は、正解率によって行う。これは、発言のペアごとに、同定した関係が正解しているか否かを判断し、正解しているものの割合を求めたものである。

回答要約の性能評価は、適合率、被覆率、F 値、冗長さの 4 尺度によって行う。これらの尺度は、平尾ら [4] が定義した抜粋による複数文書要約の評価尺度を参考にして定義する。実験の評価においては、4 尺度の値をスレッドごとに求め、それぞれを全スレッドについて平均した値を用いる。

適合率は、抜粋された発言のうち正解要約の要素となるものの割合を表す。また、被覆率は、抜粋が正解要約の内容を網羅している割合を表す。正解データによって正解と判断される選び方のリストを  $C_1, \dots, C_m$ 、システムが抜粋した発言集合を  $E$  とすると、適合率および被覆率は次の式で表される：

$$\text{適合率} = \max_{1 \leq i \leq m} \frac{|E \cap C_i|}{|E|}, \quad \text{被覆率} = \max_{1 \leq i \leq m} \frac{|E \cap C_i|}{|C_i|}.$$

適合率および被覆率は、いずれも値が大きいほどより良い要約であるといえる。理想的な抜粋に対しては、いずれも値が 1 となる。

F 値は、適合率と被覆率の値の調和平均である。この値が大きいほど、より良い要約であるといえる。理想的な抜粋に対しては、この値は 1 となる。

冗長性は、被覆率を下げずにどれだけ抜粋を減らせるかを表す。回答内容の要約として  $n$  個の内容が必要であり、 $i$  番目の内容について  $f_i$  個の冗長さな発言を抜粋  $E$  が含むとき、冗長性は以下の式で表される：

$$\text{冗長さ} = \frac{\sum_{i=1}^n f_i}{n}.$$

冗長性は、値が小さいほどより良い要約であるといえる。理想的な抜粋に対しては、この値は 0 となる。

### 5.4 比較手法

関係同定におけるベースラインは、最も多く出現する関係の種類を常に選択した場合の値とする。すなわち、質問・回答間のタスクでは常に「直接回答」、回答・回答間のタスクでは常に「対立」を選択した場合の性能をベースラインとする。

要約生成に関しては、関係同定の結果を用いずに発言を選択した場合の実験を行い、性能を比較する。このときの選択基準として、全発言、ベストアンサー発言の 2 種類を採用する。ここで、全発言からなる要約は、関係同定におけるベースラインを用いて生成した場合の要約と同一である。また、発言間の関係を考えることの有効性を示すために、発言間の関係の正解を用いた場合の実験も行い、性能を比較する。

\*3 <http://chasen.org/~taku/software/TinySVM/>

表 4: 関係同定の結果

	質問・回答間	回答・回答間
one-vs-rest	0.906	<b>0.480</b>
pairwise	0.906	0.473
ベースライン	0.906	0.464

表 5: 回答・回答間の関係同定において、質問タイプに関する素性を抜いた場合の結果

抜いた素性	one-vs-rest	pairwise
なし(全素性を利用)	0.480	0.473
質問タイプ	0.480	0.472
質問タイプに対応する固有表現	0.480	0.473
核文の注目名詞	0.484	0.474

## 6. 結果

### 6.1 関係同定

関係同定の結果を表 4 に示す。回答・回答間は、提案手法がいずれもベースラインを上回った。また、one-vs-rest 法と pairwise 法を比較すると、前者のほうが性能が良かった。質問・回答間は、提案手法においてもすべて「直接回答」と分類され、ベースラインと同じ結果となった。

田村らの定義した核文・質問タイプを利用する素性を抜いたときの結果を、表 5 に示す。各列の右側の数字は、全素性を利用した場合の結果との差であり、この値が負であれば、当該素性が同定性能の向上に貢献していることになる。質問タイプ素性を抜いた場合、pairwise の正解率が下がり、この素性が関係同定の性能向上に貢献していることが示された。

一方、質問タイプに対応する固有表現の素性は、関係同定の性能向上に貢献していないことが示された。この原因としては、回答がテキストとなる質問タイプの場合、対応する固有表現の種類が考えられず、この素性が本質的に利用できないことが考えられる。質問タイプを限定して問題を扱えば、この素性も性能向上に貢献できるかもしれない。また、この素性には質問タイプ同定および固有表現抽出を自動で行った結果を用いているため、各段階の誤りが累積した可能性も考えられる。

### 6.2 回答要約

回答要約生成の結果を表 6 に示す。ここで、「提案手法(同定された関係)」は、すべての素性を用いて関係同定を行い、その結果を利用して回答要約を生成したときの値である。

提案手法による要約は、ベストアンサーによる要約に比べて被覆率・F 値がともに高く、全回答からなる要約に比べて冗長度が低かった。この結果から、発言間の関係を考慮するとより良い要約を生成できることが示された。また、関係同定の正解を用いて要約を生成すると、同定した関係を用いるよりも適合率・F 値が高く、かつ冗長度が低くなった。つまり、正解を用いた場合の方が要約の性能が良い。このことから、関係同定の性能が向上すれば要約の性能も向上することが期待できる。

## 7. 結論と今後の課題

本研究では、質問応答サイトの発言間の関係を定義したうえで、実際のデータ上でこれを同定する手法を提案した。さら

表 6: 各要約の被覆率・冗長度

要約手法	適合率	被覆率	F 値	冗長度
ベストアンサー	0.940	0.594	0.690	0.000
全回答	0.686	1.000	0.785	0.351
提案手法(同定された関係)	0.752	0.835	0.756	0.113
提案手法(正解の関係)	0.850	0.787	0.788	0.043

に、同定した発言間関係の結果を利用することにより、質問応答サイトの要約を精度良く生成できることを示した。

本研究での提案手法は、質問・回答間の関係同定の性能向上に貢献しなかった。この関係を同定する手法について、さらに検討したい。

本研究では、発言を単位として発言間の関係を考えた。しかし実際には、同一発言中で複数の内容を述べる場合もある。発言を内容ごとに部分に区切り、それらの部分を単位としてこれらの関係を考えれば、より綿密にスレッドの内容を構造化でき、より冗長性の低い要約が得られると考えられる。

本研究では、スレッドの内容をユーザにわかりやすく提示するために、抜粋による要約という方法をとった。スレッドの内容をユーザにわかりやすく提示する方法としては、ほかに発言間の関係そのものを可視化する方法もある。どのような方法で提示すればよりわかりやすくなるか、さらに検討したい。

## 参考文献

- [1] 田村晃裕, 高村大也, 奥村学. 複数文質問のタイプ同定. 情報処理学会論文誌, Vol.47 No.6, pp.1954-1962, 2006.
- [2] Yuanjie Liu, Shasha Li, Yunbo Cao, Chin-Yew Lin, Dingyi Han, Yong Yu. Understanding and Summarizing Answers in Community-Based Question Answering Services. In *Proceedings of the 22nd International Conference on Computational Linguistics*, pp.497-504, 2008.
- [3] Dragomir Radev. A Common Theory of Information Fusion from Multiple Text Sources, Step One: Cross-Document Structure. In *Proceedings of the 1st ACL SIGDIAL Workshop on Discourse and Dialogue*, pp.74-83, 2000.
- [4] 平尾努, 奥村学, 福島孝博, 難波英嗣, 野畑周, 磯崎秀樹. 抜粋による複数文書要約を評価するためのコーパスと評価指標. 情報処理学会論文誌:データベース, Vol.48 No.SIG 14, pp.60-68, 2007.
- [5] Corinna Cortes and Vladimir Vapnik. Support-Vector Networks. In *Machine Learning Journal*, 20, pp.273-297, 1995.
- [6] 国立情報学研究所. 「Yahoo!知恵袋」データの提供について. <http://research.nii.ac.jp/tdc/chiebukuro.html>
- [7] 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, pp1-8, 2009.
- [8] Shilin Ding, Gao Cong, Chin-Yew Lin, Xiaoyan Zhu. Using Conditional Random Fields to Extract Contexts and Answers of Questions from Online Forums. In *Proceedings of ACL-08: HLT*, pp.710-718, 2008.
- [9] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系. 岩波書店, 1997.
- [10] 大野晋, 浜西正人. 角川類語新辞典. 角川書店, 1981.