

F-008

リスト型 QA タスクにおける複数回答の抽出 Extraction of multiple answers on the List type QA

糸井川 誠[†]

Makoto Itoigawa

福本 淳一[‡]

Jun'ichi Fukumoto

1. はじめに

情報化社会に伴うインターネットの普及によって、我々は容易に情報を得られるようになった。しかし、膨大なデータから本当に欲しい情報だけを取り出すのは難しく、時間がかかる。

これらの問題を解決する方法の一つに、質問を与えて欲しい情報のみを抽出する技術として QA (質問応答) がある。日本における QA の評価は、NTCIR[1] において、QAC[2] が設定されている。QAC は、日本語による QA 技術の向上を目的としており、3つのタスクによって構成されている。このうち、与えられた質問に対して、回答と思われるものを全てを1つのリストとして出力することで複数の正解を持つ質問に答えられるシステムを開発することを目的とした、Subtask2 (リスト型 QA タスク) が設定されている。このリスト型 QA タスクは、正解の数が限定されていない質問に適用するのが目的である。回答に正解でないものが含まれていたり、正解であるものが含まれていない場合は減点となる。

一般的に QA といえば Subtask1 用のシステムを指す。そのシステムが答えを出力するためには、キーワードとの距離などから与えられる評価値によって回答に優先順位をつけて5つを出力し、その中に一つでも正解が含まれていれば、その質問に対しては正解と見なされる。しかし、Subtask2 用のシステムを実現するには、優先順位をつけるだけでは不十分である。なぜなら、正解がない質問に対しては解がないと判断したり、複数の正解を持つ質問に対しては、その個数や、どれが正解であるかを判断するための情報が必要だからである。

複数の正解を抽出するには、正解の現れ方を考慮する必要がある。実際に、図1に対する質問として「爆笑問題のコンビは誰ですか」を想定する。この質問の場合、「誰」という表現から質問のタイプは人名であることがわかる。記事の構成を見ると、「太田光」と「田中裕二」という固有表現は、キーワードである「爆笑問題」に近く、「太田光と田中裕二が、…」と並列して記述されており、同時に互いの距離が非常に近いのを見て取れる。以上から正解は「太田光」と「田中裕二」の2つであるということがわかる。

このように、「キーワードとの距離が近い」、「同じ質問タイプで並列して記述されている」、「互いが比較的近い位置に存在する」、という条件のうち、より多くを満たしている回答候補が正解である可能性が高いと考えられる。さらに、質問文にある「コンビ」という表現からも、正解の個数が「2つ」とであると決定できる。

既存の Subtask2 用システムでは、一定の閾値以上の評価値を持つものを回答とするシステムが多いが、その評価値は Subtask1 用システムで得た結果に頼るものが

爆笑問題の太田光と田中裕二が、ぐいぐいと自分たちの世界に話を持っていく。その強引さはさすが、旬のコンビという感がある。また、同じく司会を務める家森幸子アナウンサーを二人が「口撃」するところも軽妙だ。飲み物の注文を取りにくる「ジュース君」や、ダイヤモンド指輪をじゃんけんで獲得するゲームなど、華やかな展開に起伏をつけようとする工夫もある。

_____ : キーワード、..... : 回答候補

図 1: 記事 1

ほとんどである。本研究では、一文中に複数存在する同じ質問タイプの固有表現や、固有表現間の距離を基に重み付けを行い、その重みに基づいたスコアリングの結果と質問文をもとに出力する回答を決定する。

2. 複数回答の選択

この節では、単一記事において複数回答の選択手法について述べる。まずは、回答候補にスコアリングによって評価値をつけることで、正解の可能性が高いものが高得点を獲得できるようにする必要がある。また、正解がいくつあるかということも分からないため、フィルタリングを行い、出力する回答候補を決定する必要がある。

2.1 スコアリング手法

以下では、どのような情報を用いてスコアリングを行うか述べる。

- 並列して記述されている固有表現

並列して記述されている回答候補については、いつも「〇〇と△△」のような書き方がされているわけではないため、「一文中に複数の回答候補が存在すれば、それらは並列して記述されている」と考えることとした。一文中に存在する回答候補の数を得点として加えれば、前節の例のように正解を上位にすることができると思う。

- 回答候補間の形態素距離

一文中に全ての答えが必ずしも存在するわけではなく、複数の文に渡って記述されている回答も考慮する必要がある。この場合、各回答候補の位置を特定し、回答候補間の距離も測ることで、相対的にどの回答候補が密集しているかを測ることができ、密集具合によって重みをつければ、スコアに反映することができると思う。

以上の2つの情報を利用してスコアリングを行えば、正解である回答候補をより上位にすることが可能であると思う。

[†]立命館大学大学院理工学研究科 (m.itoigawa@nlp.is.ritsumei.ac.jp)

[‡]立命館大学情報理工学部 (fukumoto@media.ritsumei.ac.jp)

2.2 フィルタリング手法

2.1のようなスコアリングを行い、正解である回答候補が上位にランキングしている考えるなら、どこまでが正解であるかを特定する必要がある。以下では2つの手法で出力する回答を選ぶ。

- スコアの差による出力個数の限定

前述したような条件を元にスコアリングを行えば、条たしているものとあまり満たしていないものの差がつくはずである。スコアの差が最大のところまでを出力とすれば、より正確に回答を得られると考える。

- 質問分析による出力個数の特定

質問文中に、質問対象の個数を限定するような表現を見つけることができれば、スコアの差を見て回答を出力するよりも正確な個数の回答を出力することができる。"コンビ"は二人の組み合わせを表現する単語であることから、出力する回答は2つであると推測できる。表1に、現時点で数値表現である考えられる、形態素の品詞の組み合わせを挙げる。

表 1: 品詞コードを元にした数値表現のパターン

	品詞の組み合わせ	例
パターン 1	(名詞-数)-(名詞-接尾-助数詞)	2人、三個
パターン 2	(名詞-数)-(名詞-接尾-助数詞以外)	3つ(半角数字)
パターン 3	(名詞-一般)	コンビ
パターン 4	(名詞-代名詞-一般)	誰と誰

表2に、今回実装されているパターン3に含まれる数値表現を示す。

表 2: パターン3に含まれる数値表現

回答個数	数値表現
1	世界一、日本一、ひとり、ひとつ、一つ、1つ
2	ふたり、双子、兄弟、ペア、コンビ ふたつ、二つ、2つ、カップル、夫婦、夫妻
3	トリオ、三つ、3つ、三つ子
4	四つ、4つ、四つ子

3. 複数回答抽出処理

この節では、前節で提案した手法を元に実装した複数回答を抽出する処理について説明する。

3.1 複数回答抽出処理の構成

複数回答抽出処理は以下の処理で構成されている。

- 情報抽出部

コア QA システムとして RitsQA type-E[3] の処理データを受け取り、指定された記事から、回答候補と同じ固有表現を特定する。

- スコアリング部

回答候補とマッチした固有表現の位置関係からスコアリングを行う。

- フィルタリング部

各回答候補のスコアと質問文から出力するものを選び、回答とする。

以下では、複数回答抽出処理の各構成部の説明をする。

3.2 スコアリング部

スコアリングを行うにあたって、まずコア QA システムから「10位までの回答候補」、「各回答候補のスコア」、「検索に利用した記事番号」を受け取る。

次に受け取った回答候補と記事番号をもとに、先に述べた2つの手法を用いて2種類のスコアを出し、コア QA システムでのスコアを含めた3つの評価値の基準を平等にして最終的なスコアを算出する。以下にその手法を述べる。

- 手法1:並列して記述されている回答候補の処理

一文中に出現する回答候補の個数を、それぞれの回答候補の得点とする。ただし、一文中に回答候補が一つしかなかった場合、スコアは0となる。

- 手法2:互いに距離の近い回答候補の処理

一記事中で回答候補の距離をもとに回答候補群をつくる。回答候補群は比較的近い距離にある回答候補のまとまりであり、以下の手順で特定する。

1. 記事中の隣あう回答候補間の距離を測る
2. 距離の長い順に見て、上位1/3の距離を持つ回答候補間を境として回答候補のまとまりを作る
3. 現在注目している回答候補と一番近い回答候補の距離の1.5倍の距離以内にある回答候補を最終的な回答候補群とする

3.3 フィルタリング部

スコアリングによって順位を付けられた回答候補をどこまで出力するかを評価する部である。ここでは、スコアに注目して出力する個数を限定する方法と、質問を解析して出力する個数を特定する方法を述べる。

- 手法1:スコアの落差による回答個数の限定

隣り合う回答候補のスコアの差を取り、その差が一番大きな回答候補のペアのうち、高い方の順位までを出力とする。2位と3位の差が一番大きい時、1位と2位が回答なる。

- 手法2:質問解析による回答個数の特定

茶筌[4]によって質問文を形態素解析し、表1を元に、出力する個数を決定する。

本研究では、上記の2つの手法を順番に実行することにしてている。これにより、双方で異なる回答個数が算出された場合、後者の結果が優先される。

4. 実験

4.1 実験方法

本手法によって複数回答をどの程度抽出できるかの確認するための実験を行う。質問文は QAC-2 Subtask2 の Formal Run で出題された200問、検索の対象となるテキストは、毎日新聞の1998年と1999年の2年分を使用し、スコアリング手法1,2とフィルタリング手法1,2をすべて用いて実行した。本研究では、スコアリングとフィルタリングの精度を測ることが目的であるため、評価の対象はコア QA システムの処理で上位10位以内に1つ以上正解が含まれていた質問とし、コア QA システムで抽出された正解の数をその質問の正解の数とした。

4.2 実験結果

使用した 200 問のうち、コア QA システムが上位 10 位以内に 1 つ以上正解を抽出できた質問は 109 問であった。そのうち、本システムが 1 つ以上正解を出力できた質問は 58 問であった。平均再現率、平均適合率、平均 F 値は表 3 のようになった。

表 3: 実験の結果

平均再現率	平均適合率	平均 F 値
0.505	0.657	0.510

次にフィルタリングの効果を分析する。回答の個数が正解の個数と一致した質問数が 109 問中 39 問であった。また、質問文から正解の個数を推測できると判断された 200 問中 13 問については、すべてにおいて正解の個数を特定できた。

5. 考察

平均再現率は 0.505 という数字から、出力された回答リストのうち、約半分は正解を抽出できたと考えられることから、本手法で用いたスコアリングの有効性が伺える。しかし、うまく正解を抽出できなかった質問に関しては、「木星の衛星にはどんな名前がついていますか。」という質問では、コア QA システムの処理では多くの正解が出ていたにもかかわらず、一つしか出力できなかった。これは、各回答の根拠となっている記事に統一性がないためであることが分かった。記事を見ると、同一記事に出現する回答同士も記事の各所に点在していたため、スコアリングがうまく機能しなかったものと考えられる。実際に回答が含まれている記事は一つではないため、今後は記事間の類似性を考慮する必要があると考える。平均適合率においては、109 問中 39 問と 1/3 しか正解の個数と一致しなかったにも関わらず、0.657 という数字であったことから、多くの質問で正解の個数に近い個数の回答を出力できているのがわかる。今回はコア QA システムの出力において正解が含まれていた質問のみを評価したが、それらの質問についてもすべての正解が含まれているわけではないため、コア QA システムの精度の向上によって、本手法の精度も向上すると思われる。

次にフィルタリングにおいて、200 問中 13 問においては、質問文から正確に個数を特定できていることから、質問文に数値表現が含まれている場合はフィルタリング 2 が非常に有効であると考えられる。今後は更にパターンを増やすことで精度の向上が期待できる。しかし、「15ヶ国のうち何ヶ国」という表現のように、質問対象の個数でない数値を誤認識しているものが見られたため、このような誤認識を減らすことが今後のフィルタリングの課題である。

6. おわりに

本稿では、質問応答システムにおける、複数の解答を持つ質問に答えるための手法について述べた。提案した手法は、複数回答の新聞記事内の出現位置に注目したスコアリングと、スコアの差や質問文に含まれる数値表現を基に、出力する回答を決定するフィルタリングを行った。結果は、前処理が正解を出すことできた質問のうち、

約半数の質問において一つ以上の正解を得ることができた。結果はコア QA システムの精度にも大きく依存するためコア QA システムの精度向上が求められる。

参考文献

- [1] NII-Test Collection for IR Home Page
<http://research.nii.ac.jp/ntcir/index-ja.html>
- [2] Question Answering Challenge Home Page
<http://www.nlp.is.ritsumei.ac.jp/qac/index-j.html>
- [3] 遠藤哲哉, 福本淳一 “詳細化された質問タイプによる質問応答システム” 情報処理学会研究報告 2004-NL-159, pp.25-30, 2004
- [4] ChaSen Home Page <http://chasen.aist-nara.ac.jp/>
- [5] J. Fukumoto, T. Kato, and F. Masui, Question Answering Challenge for Five ranked answers and List answers - Overview of NTCIR4 QAC2 Subtask1 and 2: NTCIR Workshop 4 Meeting, pp.283-290, 2004
- [6] J. Fukumoto, T. Niwa, M. Itoigawa, and M. Matsuda, Rits-QA: List answer detection and Context task with ellipses handling: NTCIR Workshop 4 Meeting, pp.310-314, 2004