

## Non-factoid 型質問応答システムにおける質問タイプ判別法の改善

## Improving type identification method for non-factoid Question Answering

小田 貴博†  
Takahiro Oda秋葉 友良†  
Tomoyosi Akiba

## 1. まえがき

キーワードリストを入力として関連する「文書」を出力する従来の文書検索の発展システムとして、自然言語質問を入力としてその直接の「答」のみを出力するオープンドメイン質問応答が提案されている。本研究では、任意のタイプの質問を対象とする質問応答システムを実現することを目的として、主にベースライン手法における入力質問と回答候補のタイプの一致の判定法の改善を試み、評価実験により提案法の効果を確認した。

## 2. 質問応答システム

オープンドメイン質問応答システムで扱われる質問は factoid 型の質問と non-factoid 型の質問の 2 種類に分けられる。factoid 型の質問は回答が数量や物の名前などの名詞や名詞の連続程度の短い句や節を答えとして期待する質問で、non-factoid 型の質問はそれ以外の理由を問う質問や方法を問う質問等の比較的長い回答を期待する質問である。

non-factoid 型の質問応答システムは理由や方法等のある特定のタイプの質問に対する回答を出力するシステム [1][2] とタイプに依存しない質問応答システム [3][4] がある。タイプに依存しない質問応答システムの固有の問題として入力質問の回答候補のタイプの一致判定を行う必要がある。水野ら [3] の手法では入力質問と回答候補のタイプの一致判定を SVM にて作成した分類器で行っている。文書検索によって得られた回答候補を分類器でタイプの一致判定を行い、一致する回答候補を回答として出力している。佐藤ら [4] の手法はタイプの判定を QA コーパスを利用して、 $\chi^2$  統計量を求めることによって得られた特徴表現を利用して行っている。

本研究では佐藤らの手法(以下ではベースライン手法と呼ぶ)を参考にしてベースライン手法のシステムを実装し、性能改善を試みた。

## 3. ベースライン手法の実装

文献 [4] を参考にして実装したベースライン手法の全体図を図 1 に示す。内容表現と特徴表現を求め、文書検索によって抽出した回答候補が内容表現と特徴表現をどの程度含んでいるかを式(1)で評価することによって回答を求める。

式(1)中の  $m$  は回答候補  $S_i$  に含まれる内容表現  $\omega_{ij}$  の異な

$$score(S_i) = \frac{\left\{ \sum_{j=1}^m T(\omega_{ij}) \right\}^\alpha \cdot \left\{ \sum_{k=1}^n \sqrt{\chi^2(b_{ik})} \right\}^{1-\alpha}}{\log(1 + |S_i|)} \quad (1)$$

り数であり、 $T(\omega_{ij})$  は回答候補  $S_i$  に含まれる内容表現  $\omega_{ij}$  の内容関連度である。また、 $n$  は回答候補  $S_i$  に含まれる特徴表現  $b_{ik}$  の異なり数であり、 $\chi^2(b_{ik})$  は回答候補  $S_i$  に含まれる特徴表現  $b_{ik}$  の  $\chi^2$  統計量である。そして  $\alpha$  は内容表現と特徴表

現の重みを定めるパラメータである。以下では本稿に関連する特徴表現について説明する。

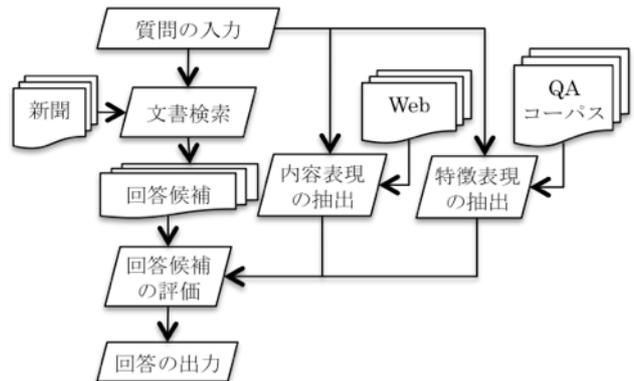


図 1 ベースライン手法の全体図

## 3.1 特徴表現

特徴表現とは回答に特徴的に出現する機能語を主とした n-gram である。文献 [4] では  $n=2$  としている。(例：理由を聞く質問に対する回答では「～だから」といった表現が頻出する) この特徴表現は QA コーパス (Yahoo 知恵袋<sup>1</sup>) を利用することによって求める。前処理にて QA コーパスの抽象化を行い、抽象化された QA コーパスから特徴表現を抽出する。

## 3.1.2 QA コーパスの抽象化

タイプの一致判定を行う前に抽象化処理を行う。この抽象化処理とは内容語を品詞に置き換える処理である(ただし一部の内容語はそのまま用いる)。表層表現をそのまま用いる内容語は、理由や目的など質問のタイプ推定の手掛かりになる語と、頻出する動詞と形容詞の 3 種類の語である。この処理によって内容を考慮せずタイプの一致のみを考慮した判定が可能になる。つまり、理由を聞いている質問同士ならば内容に関係なく、それらをよく似た質問として扱うことが出来るようになる。

## 3.1.2 特徴表現抽出

QA コーパスの質問で入力質問とタイプが一致する質問とそうでない質問を分けてタイプが一致する質問に対する回答集合  $A$  に頻出し、そうでない質問に対する回答集合  $A$  にはあまり出現しない機能語 2-gram が特徴表現になる。入力質問にタイプが一致する質問を抽出する手順を図 2 に示す。ベースライン手法では質問間類似度を距離 3 までを許すスキップ 2-gram を使って求めている。入力質問と QA コ

†豊橋技術科学大学情報工学系

<sup>1</sup> <http://chiebukuro.yahoo.co.jp>

一パスの質問文中に含まれている疑問詞を中心とした 7-gram からスキップ 2-gram を取り出して、どの程度同じ 2-gram を含んでいるかを尺度としている。入力質問として「インドとパキスタンが対立している理由は何ですか。」といった質問が入力されるとする。システムはこれを 7-gram に変換し、抽象化を行って”シ\_テ\_イル\_リ\_ユ\_ウ\_ハ\_ナ\_ニ\_デ\_ス”とする。そしてこの 7-gram に含まれるスキップ 2-gram をどれだけ多く含むかの尺度によって類似度計算し、入力質問にタイプが一致する質問として「日本が常任理事国入りしようとしている理由は何ですか。」といった質問を収集する。

次に、タイプが一致するとして収集した質問に対する回答に含まれる 2-gram  $b$  について  $\chi^2$  統計量を求める。特徴表現抽出の手順を図 3 に示す。この  $\chi^2$  統計量を文献[4]では式(3)を使って求める。式(3)で求められた  $\chi^2$  統計量が高い 2-gram を特徴表現とする。また、式中の  $n$  は QA コーパスに含まれている質問回答ペアの総数である。 $A$  はタイプが一致する質問に対する回答で、 $B$  は  $\chi^2$  統計量を求めたい 2-gram を含んでいる回答である。ただし本稿の実装では式(3)の計算にイェーツの補正を行った。タイプが一致しているとする質問は入力質問との類似度が高い上位 200 として、それらに対する回答文に含まれている 2-gram に対して  $\chi^2$  統計量を計算する。計算時間の短縮を図るために、得られた  $\chi^2$  統計量の大きい上位 500 の 2-gram を特徴表現とした。

$$\chi^2(b,A) = \frac{n \cdot (|A \cap B| \cdot |\bar{A} \cap \bar{B}| - |A \cap \bar{B}| \cdot |\bar{A} \cap B|)^2}{|A| \cdot |B| \cdot |\bar{A}| \cdot |\bar{B}|} \quad (3)$$

入力質問に対する特徴表現の例を表 2 のベースラインの行にて示す。

### 3.2 回答候補の抽出

入力質問から複合名詞に出来る名詞を複合語にした名詞の集合と、動詞、形容詞の集合、そして内容語の集合の 3 種類の集合を取り出す。そしてそれぞれをクエリとして文書検索を行う。文献[4]では Web を対象として文書検索を行っていたが本研究では新聞記事を対象として文書検索を行って、回答候補を抽出する。また、回答候補の単位も答えの部分のみを出力するのではなく、段落単位で出力する。回答候補を式(1)で評価する際も、段落単位で行った。

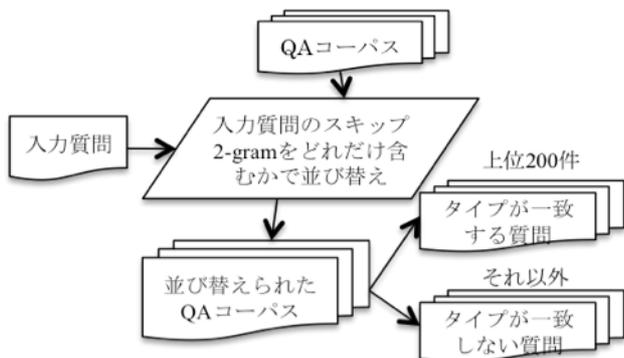


図 2 タイプが一致する質問の抽出手順

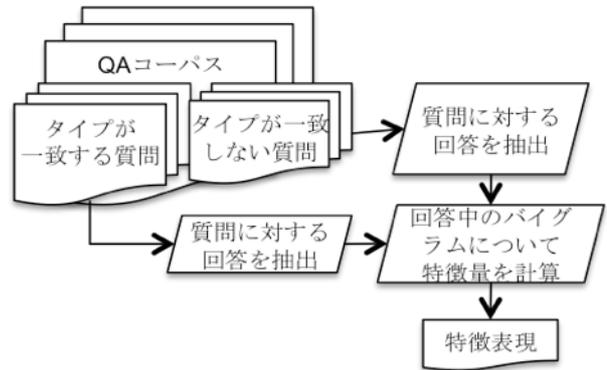


図 3 特徴表現の抽出手順

## 4. 提案法

名詞の主題化, 3-gram 特徴表現, 構文の特徴表現の利用といった手法をベースライン手法に適用した。

### 4.1 名詞の主題化

ベースライン手法は特徴表現を求める際に主に機能語を中心とした 2-gram を用いる。これに対して、提案法では QA コーパスの抽象化を工夫することによって内容語と機能語を同時に考慮する特徴表現を抽出する。QA コーパスを調査したところ、質問と回答の間に共通した名詞が出現する QA は約 80 万件のうち 60 万件とかなりの数が存在した。(例: **質問**. 東海地方とはどこ **回答**. 東海地方とは 愛知県全域 静岡県全域...) この傾向は特に定義型の質問に対して多くみられた。そこで質問文に含まれている名詞とその質問の主として知りたいこと、つまり「主題」と考えた。ベースライン手法ではすべて名詞は「<名詞>」と品詞化していた部分を、対応する質問文に現れる名詞については「<主題>」と品詞化し、それ以外の名詞を<名詞>と品詞化を行うように変更した。また、これ以外の部分はベースライン手法と同様に実装した。

質問が与えられたときに特徴表現で回答候補をどのように評価するかの手順を図 4 に示す。システムは予め抽象化された QA コーパスを用意しておく。回答文を抽象化する際、知恵袋内の各質問から一つずつ作成した主題リストを使用して、回答を名詞と主題を区別しながら抽象化を行う点がベースライン手法とは異なっている。質問が入力されるとシステムは入力質問から主題リストを作成し、作成した主題リストを利用して主題と名詞を区別して回答候補の抽象化を行う。そして、図 2 と図 3 で示した手順で抽象化された QA コーパスから特徴表現を抽出し、回答候補に含まれている特徴表現だけを取り出し、それらの特徴表現の  $\chi^2$  統計量の合計を回答候補が入力質問とタイプが一致しているかのスコアとする。ベースライン手法と提案法での QA コーパスの抽象化の違いを以下に示す。以下の例の場合「鳥」「水」「体内」が質問の主題になっている。

【質問】鳥はどんな方法で水を飲み、どうやって体内に吸収されていくのですか？

【回答】ほとんどの鳥は嚙下が出来ずくちばしに含んだ水を上を向いて流し込みます。一度に少量しか口に出来ないため何度もくりかえし上を向いて飲みます。ハトは特別で、くちばしを水につけゴクゴク飲むことが出来ます。

【ベースライン手法】〈副詞〉〈ノ〉〈名詞〉〈ハ〉〈名詞〉〈ガ〉〈デキズ〉〈名詞〉〈ニ〉〈動詞〉〈ダ〉〈名詞〉〈ヲ〉〈名詞〉〈ヲ〉〈動詞〉〈テ〉〈動詞〉〈マス〉〈記号〉〈名詞〉〈ノ〉〈ニ〉〈名詞〉〈シカ〉〈名詞〉〈ニ〉〈デキ〉〈ナイ〉〈ノデ〉〈ナン〉〈モ〉〈動詞〉〈シ〉〈名詞〉〈ヲ〉〈動詞〉〈テ〉〈動詞〉〈マス〉〈記号〉〈名詞〉〈ハ〉〈名詞〉〈デ〉〈記号〉〈名詞〉〈ヲ〉〈名詞〉〈ニツケ〉〈名詞〉〈動詞〉〈名詞〉〈ガ〉〈デキ〉〈マス〉〈記号〉

【提案法】〈副詞〉〈ノ〉〈主題〉〈ハ〉〈名詞〉〈ガ〉〈デキズ〉〈名詞〉〈ニ〉〈動詞〉〈ダ〉〈主題〉〈ヲ〉〈名詞〉〈ヲ〉〈動詞〉〈テ〉〈動詞〉〈マス〉〈記号〉〈名詞〉〈ニ〉〈名詞〉〈シカ〉〈名詞〉〈ニ〉〈デキ〉〈ナイ〉〈ノデ〉〈ナン〉〈モ〉〈動詞〉〈シ〉〈名詞〉〈ヲ〉〈動詞〉〈テ〉〈動詞〉〈マス〉〈記号〉〈名詞〉〈ハ〉〈名詞〉〈デ〉〈記号〉〈名詞〉〈ヲ〉〈主題〉〈ニツケ〉〈名詞〉〈動詞〉〈名詞〉〈ガ〉〈デキ〉〈マス〉〈記号〉

また、以上のような QA コーパスに対する抽象化と同様に、入力に対する回答候補についても主題化を適用する。入力質問を「スケルトンとはどういった競技ですか。」とする。回答候補に含まれる「～は」といった表現のうち「スケルトンは」や「競技は」といった 2-gram は「<主題>\_ハ」に置き換えられ、それ以外は「<名詞>\_ハ」に置き換えられる。それ以外の「<名詞>」を含む表現も、含まれる名詞が入力質問から作成した主題のリストに存在するならば「<主題>」と置き換えられる。ベースライン手法の特徴表現は機能語のみを考えていたが、名詞の主題化を導入することによって内容語と機能語の両方を考慮した特徴表現を考えることが出来る。

### 4.2 3-gram 特徴表現の利用

QA コーパス内の回答文には「～とは」といった表現が頻出する。特に定義型の質問に対して、「～とは...です」といった何かの定義を説明する形で頻出する。しかし、文献[4]では特徴表現は回答文中の 2-gram から抽出しているの、「<名詞>\_ハ」や「ト\_ハ」といった特徴表現は抽出できても「<名詞>\_ト\_ハ」は抽出出来ない。そこでベースライン手法では特徴表現は 2-gram についてのみ考えていたが 2-gram と 3-gram の両方について特徴表現を考えることを行った。また、より長い表現を使うことでより  $\chi^2$  統計量が高い特徴表現を抽出出来ると推測した。

### 4.3 構文的特徴表現の利用

2-gram や 3-gram では単純な語の並び方でしか、表現出来ない。そこで特徴表現をさらに拡張して、回答文に含まれている部分木に対して特徴表現を求めることを行った。回答文を CaboCha<sup>2</sup> にて構文解析し、その出力として得られる構文木から部分木の頻度を求める。3-gram 特徴表現を利用した場合との比較のために、部分木の大きさを 2~3 に制限した。部分木の例を表 2 の構文的特徴表現の行に示す。

### 5.1 評価実験

各手法の評価は non-factoid 型質問の QA テストコレクションである NTCIR-6 QAC4 100 問を入力質問として行った。回答候補として NTCIR-6 QAC4 で対象になっている毎日新聞 1998 年-2001 年分から文書検索によって段落を抽出し、

スコアが 1 位から 5 位までの回答を出力する<sup>3</sup>。正解の判定方法は MRR(平均逆順位)を求めることで評価した。

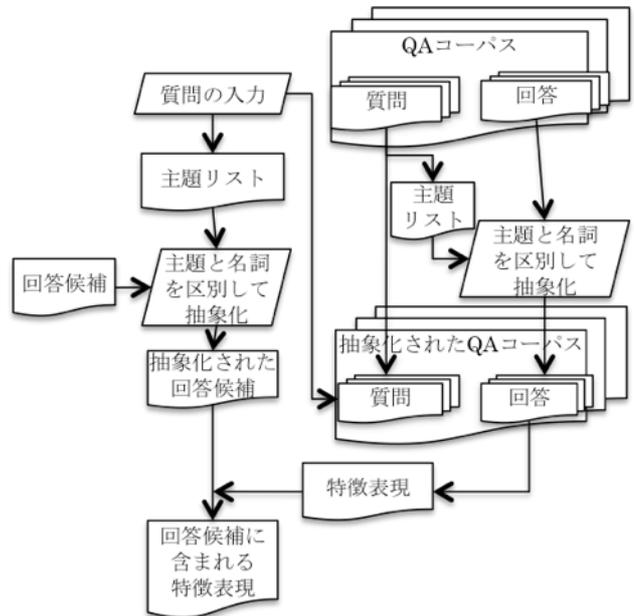


図 4 主題を考慮した特徴表現による回答候補の評価

表 1 各手法を適用した場合の MMR(平均逆順位)

|                 | 全体    | 定義型   | 非定義型  |
|-----------------|-------|-------|-------|
| ベースライン          | 0.339 | 0.350 | 0.305 |
| 主題の考慮           | 0.392 | 0.470 | 0.273 |
| 3-gram 特徴表現     | 0.361 | 0.457 | 0.300 |
| 構文的特徴表現         | 0.157 | 0.169 | 0.151 |
| 主題& 3-gram 特徴表現 | 0.392 | 0.470 | 0.273 |

### 5.2 結果

評価実験の結果を表 1 に示す。主題の考慮と 3-gram 特徴表現を使用した場合には性能を改善することができ、両方を併用した場合に最も性能を改善することが出来た。システムの性能を質問の定義型とそれ以外の非定義型の 2 種類に分けて評価を行うと、定義型の質問では主題の考慮と 3-gram 特徴表現を併用した場合が最も性能が高くなり、ベースライン手法に対して 24% 近く改善することが出来た。それぞれの手法のみを利用した場合でも定義型に対する性能を改善することが出来た。

<sup>2</sup> <http://chasen.org/~taku/software/cabocho/>

<sup>3</sup> 構文的特徴表現を求める手法以外では約 80 万件の QA ペアを利用して特徴表現を求めているが、計算量を削減するために構文的特徴表現では 30 万件の QA ペアで特徴表現を求めることを行った。

5.3 考察

各手法によって抽出出来る特徴表現を表2に示す。表2の左側の質問、「NPO法とは何ですか」は定義型の質問で、右側の質問の「インドの核実験強行の目的は何ですか。」は非定義型の理由をたずねる質問である。定義型の質問の場合、ベースライン手法で抽出出来る特徴表現とそれ以外の提案法を考慮する場合に抽出出来る特徴表現はかなり異なったものになった。また、非定義型の質問では提案手法でもベースライン手法にて抽出できた”モクテキ”を含んだ特徴表現を主に抽出した。この結果から提案法は主に定義型質問に対して性能を改善することが確認出来る。逆に非定義型の質問に対しては性能が下がってしまったが、これは本来”<名詞>”として抽出しなければならない表現を”<主題>”として抽出としていることが原因の一つであると考えられる。

提案法を適用することによって、定義型の質問に対する特徴表現は主に”<主題>”を含んだものになるが、非定義型の質問に対する特徴表現はベースライン手法と提案法の間には大きな違いはない。このことから非定義型の質問のタイプを判定するのに、回答候補が”<主題>”を含んでいるかどうかはあまり手掛かりにはならないと考えられる。

QAコーパスを用いる本手法を新聞記事に適用する場合の問題としてQAコーパスと新聞記事との間での表現の仕方の違いと回答候補の不足が原因であったのではないかと考えられる。新聞記事の特徴としては「～である」や「～だ」といった表現がよく使われているのに対して、QAコーパスの回答文では「～です」といった表現がよく使われる。QAコーパスから抽出している特徴表現からも表2のようにその傾向がある。

6. まとめ

実験結果から質問文中の主題を考えることによって定義型の質問に対しては性能が改善出来ることを示すことが出来た。今後の予定としては回答候補を毎日新聞から取り出したのでWebから回答候補を取り出した場合の評価も行うことを考えている。

参考文献

[1] 三原英理, 藤井敦, 石川徹也, “Webを用いたヘルプデスク思指向の質問応答システム”, 言語処理学会 11 回年次発表論文集, pp. 1009-1012, 2005.  
 [2] 磯崎秀樹, 東中竜一郎, “パターンマイニングを用いて「なぜ」に答えるシステム”, 言語処理学会 14 回年次発表論文集, pp. 1009-1028, 2008.  
 [3] 水野淳太, 秋葉友良, “任意の回答を対象とする質問応答のための実世界質問の分析と回答タイプ判定法の検討”, 言語処理学会 13 回年次発表論文集, pp. 1002-1005, 2007.  
 [4] 佐藤充, 石下円香, 森辰則, “Web 文書を情報源とする記述的回答が可能な質問応答システム”, 言語処理学会 14 回年次発表論文集, pp. 1009-1012, 2008.

表2 各手法で抽出出来る特徴表現

| 質問         | NPO法とは何ですか.  | インドの核実験強行の目的は何ですか.   |
|------------|--|--|
| ベースライン     | “チガイ_ハ”, “ノチガイ”, “グライ_ナ”, “<名詞>_ト”,  | “モクテキ_ハ”, “ノモクテキ”, “モクテキ_ヲ”, “モクテキ_モ”, “モクテキ_デス”, ...                                    |
| 主題の考慮      | ..., “< 主 題 >_ト”, “<記号>_<主題>”, “<主題>_ハ”, “<主題>_ノ”, ...   | “モクテキ_ハ”, “ノモクテキ”, “モクテキ_ヲ”, “モクテキ_モ”, “モクテキ_デス”, ...                                    |
| 3-gram特徴表現 | “マシ_タ_トイウ”, “ニタイシテ<記号>_<形容詞>”, “ダ_<記号>_ナド”, “ガ_<名詞>_トシテ”, “<名詞>_ト_ハ”, ...                              | “モクテキ_ハ_<記号>”, “<名詞>_ノ_モクテキ”, “ノ_モクテキ_デス”, “モクテキ_ヲ_<動詞>”, “<記号>_モクテキ_ヲ”, “ノ_モクテキ_ヲ”, ... |
| 構文的特徴表現    | “<記号>(マス(イイ))”, “(ト(<記号>(クレル)))”, “(ノッ (< 記 号>(ニ)))”, “(ホウホウ(レル(<動詞>)))”, “(マス(イイ))”, “マス(イイ(ヲ))”, ... | “モクテキ(ノ(<名詞>))”, “(モクテキ(ノ))”, “(ニヨッテ(<名詞>(スル)))”, “(イウ(デ(<名詞>)))”, “(イウ(デ))”, ...        |