

事実型質問応答における画像・地図を用いた回答提示

佐藤 充

横浜国立大学 大学院 環境情報学府
E-mail: {mitsuru,mori}

森 辰則

横浜国立大学 大学院 環境情報研究院
E-mail: forest.eis.ynu.ac.jp

1 はじめに

質問応答システムは、ユーザが自然言語で与えた質問文に対し、知識源となる大量の文書集合から回答語句そのもの(解)を見つけて出力するシステムである。

一般的な質問応答システムは文字による回答を返すが、質問によっては他のメディアを用いた方が分かりやすい場合がある。例えば、動植物や人物に関する質問の回答としてそれらの画像を提示することは、ユーザにとって有用と考えられる。場所を問う質問に対しては、場所を表すのに特化した画像ともいえる地図がある。そこで本稿では、Web上の画像検索エンジンや地図APIを質問応答システムと組み合わせ、画像や地図を用いた多面的な回答を提示するシステムを提案する。

2 関連研究

2.1 テキストと画像の対応付け

藤井らは、事典検索サイト Cyclone[1]において、見出し語の説明テキストに対して画像の対応付けを行なった[2]。見出し語が「ハブ(装置/蛇)」のような多義語であったとき、画像の多義性を解消する必要がある。そのため同手法ではまず、画像のリンク元のHTMLファイルにはその画像を特徴付ける語が多く存在すると考え、それらの語を画像の索引とする。次に、見出し語の、ある語義に対応する説明テキストを入力として、類似度の高い候補を上記のHTMLファイル群から検索、そこからリンクされている画像を説明テキストと対応付ける。

Cycloneは自然言語による検索も可能であるため、応用として「テキストによる質問に対して画像で回答すること」が可能であるとも述べられている。しかし、Cycloneに自然言語による質問文が入力された場合、入力文と適合する説明テキストが検索され、そこから質問文タイプに適合する見出し語が逆引きされるため、見出し語は必ずしも質問の答えにはならない¹。一方、我々の提案システムがベースとする質問応答システム[3]では、組織化されていないプレーンテキストから回答語句そのものを抽出するという点が異なる。

また、Cycloneでは予め収集した画像とそのリンク元HTMLファイルの対に対して事前に索引付けを行なっておくことを前提としている。システムの応答時間の面では有利であるが、この作業はコストが高い。提案システムでは既存の画像検索エンジンを利用することで、手元に画像データを持つことなく、質問が入力された時点で画像の検索・絞り込みを行なう。

2.2 住所探索

提案システムにおいて地図で回答する際、処理の途中で対象の住所を調べる必要がある。佐藤[6]は、以下の手順に従って与えられた名称に対する住所情報をWebから探し出すシステムを構築した。

まず、対象の名称に「住所」や「所在地」などの語を加えてクエリとし、既存の検索エンジンを用いて検索を行なう。得られたページの中から情報抽出を行なう領域を特定し、そこから住所情報を抽出する。住所の抽出に

¹説明テキスト中に回答が含まれていても、回答が見出し語になっていないとは限らない。

は独自に作成した辞書を用いている。最後に、収集した住所情報の同一性を判定し、同じ対象を表していると思われる住所情報を統合する。同手法は多くのHTMLファイルを解析するため、ある程度の時間(1~2分)を要するとされている。提案システムにおいては、実時間での応答が求められること、および住所探索の後のジオコーディングが細かい住所まで対応できないことなどを考慮して、同手法を簡略化して用いる。

3 提案システム

提案システムでは、質問応答の解を文字で出力することに加え、人物や動植物など物に関する質問の場合はその画像を表示し、場所に関する質問の場合は地図でその位置を示すようにする。こうすることで、視覚的に理解しやすい回答を提示でき、文字による解だけでは得られない情報も提供できると考えられる。

なお、本稿では質問文としては事実型の質問²のみを想定する。また、基盤となる質問応答システムは、Web文書を知識源とするように改訂した[3]である。同システムは質問応答の解を検索結果ページの要約(Snippet)から抽出するため、各ページに画像や位置情報が存在していたかどうかは分からない。そのため、質問応答の過程で解と同時にそれらを探すのではなく、質問応答の解をまず求め、次いで解を手がかりにしてそれらを探すという方針をとる。

3.1 回答方法の決定

画像と地図のいずれを用いて回答するかは、質問文タイプに基づいて決定する。質問文タイプとは、人名や地名など質問が問うている事物等の種類であって、疑問詞などに着目して表1のように分類される。

表1: 質問文タイプの種類

質問文タイプ	種別	質問例
person	人名	~は誰
place	地名	~はこの国
organization	組織名	~はこの会社
数量(lengthなど)	数値・数量	~は何メートル
none	不明	~は何

この中で、「person」は人物に関する質問であるから、その名前とともに画像を出力する。「place」は場所に関する質問であるから、地図を表示する。「organization」は組織に関する質問であるが、組織は場所の情報を持つことが多いので、この場合も地図でその位置を示すことにする。解が「~メートル」といった数量表現になる質問では、画像や地図を表示するのは適切ではないと考え、どちらも出力しないことにする。質問文タイプが判定できずに「none」となった場合、解が名詞句であれば物に関する質問と考え、画像を表示するようにする。

3.2 画像を用いた回答

3.2.1 情報源と画像検索エンジン

画像データの情報源としては、膨大な画像データが存在するWebを対象とする。現在の実装では、Web上の

²人名・地名・組織名等、短い名称が回答となるもの

画像データへのアクセスには Infoseek³の画像検索エンジンを用いている。予備実験として、Infoseek 画像検索で曖昧性のない語句 50 語に対して 10 件ずつ画像を検索した。その結果、0.8 以上の精度⁴で適切な画像を検索でき、十分実用に耐えるものであった。また、クエリには「AND」や「OR」などの論理演算子を組み合わせて用いることができる。

3.2.2 解の曖昧性解消

質問応答の解が曖昧性のない語句であれば、画像検索エンジンにクエリとして解を入力するだけで、高い精度で適切な画像が得られる。

しかし一般に語は複数の意味を持つ。検索エンジンで特定の語義に関する結果だけを絞り込みたいときは、検索語を追加することが有効である。しかし、「X AND Y」というクエリでは、X と Y のどちらの(またはその両方が含まれる)画像が出力されるかが分からない。すなわち、語義を絞り込む一方で解に対応しない画像が増えることになる。

藤井らの手法 [2] では、Cyclone の見出し語の説明テキストを使って画像の多義性を解消していた。一方で提案システムにおいては、質問文を使って同様のことができると考えられる。

[2] で使われていた「ハブ(蛇 / 装置)」の例を考える。蛇のハブが解になる質問文と、装置のハブが解になる質問文は以下のように内容が異なると考えられる。

質問 1 「沖縄に生息する毒蛇は何？」

質問 2 「スター型 LAN で使われる集線装置は何？」

しかしながら、質問文からキーワードを取り出して、質問 1 なら「ハブ (沖縄 OR 生息 OR 毒蛇)⁵」をクエリとして画像を検索すればよいかという、必ずしもそうではない。クエリにキーワードを追加することで、無関係の画像(例えば「沖縄」の画像)が混ざってしまう。

そこで、藤井らの手法 [2] を参考にして、次の手順を提案する。まず、解(上の例では「ハブ」)だけをクエリとして画像を N 件検索する。次に、それぞれの画像のリンク元の HTML ファイルを取得し、画像の周辺のテキストを切り出す(以下、「周辺テキスト」と呼ぶ)。周辺テキストは画像へのリンク(や<A> タグ)の前後 M 文の範囲とする。最後に、質問文と周辺テキストの類似度を計算し、類似度が高い候補は質問の回答として適切な画像であると考え、これを出力する。

藤井らの手法 [2] においては予めオフラインで画像データを蓄積・組織化するが、提案システムではユーザの質問が入力された時点で画像を検索する。そのため、最新の問題に関する質問に対応しやすいという利点がある。

テキスト間の類似度の計算には種々の方法が考えられるが、提案システムではベクトル空間モデルを用いる。質問文と周辺テキストは内容語(名詞、動詞、形容詞)の重みを要素とするベクトルで表現され、2つのベクトル(それぞれ、「質問文ベクトル」「周辺テキストベクトル」とする)のなす角の cosine 値で類似度が計算される。重みは出現頻度を基本とするが、内容語の品詞の種類に応じて重みを変えることも考えられる。また、質問文に含まれるキーワードの異なり数は少ないため、解が抽出された文脈中の語の重みを質問文ベクトルに加えることも考えられる。予備実験による調整により、現在の実装では第 4 章の表 3 のようになっている。

さてこの手法では、最初に画像検索エンジンから得られた画像の中に適切な画像が含まれていなかった場合は

正解に辿り着けない。このような場合はどの画像の周辺テキストも質問文との類似度が低くなることが予想されるため、類似度の大きさを見れば適切な画像が見つからないと判断できる。その際はキーワードを追加して改めて画像検索エンジンで画像を検索するというように、類似度を見て手法を切り替えるハイブリット手法が有効と考えられる。この方法についても第 4 章で実験を行ない、その有効性を検証する。

3.2.3 システム動作例(画像)

図 1 は「ハブ」が解になる質問での動作例である。蛇のハブ、装置のハブが区別され、「ハブ」という解に対して曖昧性の解消が上手く働いていることが分かる。

また、正解が複数存在するリスト型質問の場合、何位までを正解とするかは、システムが解のスコア分布から自動で判定する [4]。この場合は正解の画像をリストとしてまとめて見ることができ、ユーザは画像を見ることでそれぞれの解を比較できて便利である。これは、「自然言語による画像検索」という見方をすれば、「~のメンバー」といった複数の対象を表す語句に対して画像検索が行なえることになる。これは従来の画像検索にはなかった機能であり、提案システムの利点の一つである。



図 1: 画像を表示する例

3.3 地図を用いた回答

3.3.1 Google Maps API を用いた地図表示

地図の表示には Google Maps API⁶を用い、解が示す位置を地図上に示す。位置の指定には経緯度が用いられるため、対象の経緯度を求める必要がある。

3.3.2 地図の表示レベル

国や都道府県のようにある程度の広がりをもった領域を示す場合は、その代表点が分かればよい。一方、「東京ディズニーランド」などランドマーク(建物や施設などの地理オブジェクト)の位置を示す場合は、詳細な位置の情報が必要になる。この場合、直接ランドマークの経緯度を調べるのは難しいので、まずはランドマークの住所を調べ、得られた住所を経緯度に変換する(ジオコーディング)という二段階を踏むことにする。

3.3.3 ランドマークの住所検索

日本のランドマークの住所は、佐藤の手法 [6] を簡略化したものを用いて求める。まず、検索エンジン(Goo⁷)に「ans (“住所” OR “所在地”）」というクエリを入力して Snippet を取得する。ここで、ans は質問応答の解を表すものとする(以下同様)。

ans がランドマーク名ではなく住所の一部であり、詳細な住所を求めたいときは問われているランドマーク名およびキーワードをクエリに追加する(3.3.6 節参照)。

次に、得られた Snippet を整形し、パターンマッチで「県××市 x-y」といった住所表現を抽出する。そして、頻度が最も高かったものを出力する⁸。

⁶ <http://www.google.com/apis/maps/>

⁷ <http://www.goo.ne.jp/>

⁸ 現状では候補が複数あった場合に対応できていない。後に述べる Wikipedia を用いた経緯度検索はこれに対応している。

³ <http://search.www.infoseek.co.jp/>

⁴ 精度は第 4 章の式 (1) で定義

⁵ 「AND」は省略し、スペースで代用している。

3.3.4 ジオコーディング

ジオコーディングとは、住所を経緯度に変換することをいう。東京大学空間情報科学研究センターが提供する「CSIS シンプルジオコーディング実験」^[5]を利用し、日本の住所を経緯度に変換する。このとき同時に返される信頼度 *iconf* が3以上のときは、変換が成功したとみなせる。よって、入力した文字列が日本の住所かどうかの判定にも使える。また、「神奈川県横浜市」というように住所の一部が入力された場合は、代表点としてその区域の重心の点が得られる。

3.3.5 経緯度の検索

海外の地名の場合は前述の住所検索やジオコーディングが行えない。そのため、Webを情報源として対象の経緯度を直接調べる。

フリー百科事典 Wikipedia⁹には多くの経緯度情報が記載されているため、まずはこれを利用する。「カイロ (エジプト/アメリカ)」というような多義語の場合を考慮して、最初にGooでWikipedia内を検索する。クエリは「site:ja.wikipedia.org intitle:ans keywords」とし、得られた結果の上位から順に調べていく。ここで、*keywords*はキーワードを「OR」で結合したものとする(以下同様)。例えば、「エジプトの首都はどこですか」の解「カイロ」の経緯度を検索するときのクエリは

「site:ja.wikipedia.org intitle:カイロ (エジプト OR 首都)」となる。質問文中のキーワードをクエリに追加するため、曖昧性を排除してページにアクセスできる。経緯度情報は英語版のページに記載されていることが多いため、日本語版のページからリンクを辿ってアクセスし、パターンマッチで「30° 03' N 31° 22' E」のような経緯度情報を抽出する。

Wikipediaで経緯度情報が得られなかった場合は、住所検索と同様にSnippetからの抽出を試みる。

「ans (“北緯” OR “南緯”) (“東経” OR “西経”)」というクエリをGooに送り、経緯度情報をパターンマッチで抽出する。得られた経緯度が日本の領域外であれば、これを採用する。

3.3.6 地図表示までの流れ

以上より、地図表示までの流れは以下のようになる。まず、質問応答の解でジオコーディングを行ない、日本の住所 (*iConf* が3以上) かどうかを判定する。

解が日本の住所の一部だった場合、質問文タイプが「place.prefecture」(「place」のサブタイプで、都道府県を表す) ならば、得られた経緯度をそのまま採用する。質問文タイプがそれ以外の場合は、「landmarkはどこにあるか」といった質問であると考え、ランドマーク名 *landmark* を抽出する。そして「ans landmark keywords (“住所” OR “所在地”)」をクエリとして3.3.3節の方法で詳細な住所を求める。これをジオコーディングし、経緯度を得る。

解が日本の住所の一部ではなかった場合、3.3.5節の方法で経緯度情報を得る。経緯度が得られなかった場合、解は日本のランドマークであると考え、「ans (“住所” OR “所在地”)」をクエリとして3.3.3節の方法で住所を求める。これをジオコーディングし、経緯度を得る。

最後に、Google Maps APIを用いて地図を表示し、得られた経緯度の点にマーカーを示す。

3.3.7 システム動作例 (地図)

図2は「世界三大夜景といったらどこのことですか」という、複数の正解を持つ質問を入力した場合の動作例

⁹<http://ja.wikipedia.org/wiki/>

である。解のスコア分布から正解は3位までと判定され、「函館」「ナポリ」「香港」が解として出力される。このような場合は、1つの地図にまとめて解の位置を表示することができる。こうすることで、それぞれの位置関係が一目で把握できるようになり、これは複数の場所を解に持つ質問に対する効果的な回答の提示の仕方であるといえる。また、図のように解に関連する画像を表示することもできる¹⁰。

世界三大夜景といったらどこのことですか。



図2: 地図を表示する例

4 評価実験

質問応答システムが正しく解を出力できた場合に、どの程度適切な画像や地図を出力できるかを評価する。実験に用いた質問文は全部で100問で、その内訳は表2の通りである。この質問セットは内容が偏らないように4人¹¹で作成した。いずれも質問応答システムが正解を出力できた質問であり、正解は1つだけであるとする¹²。

表2: 質問セットの内訳

	種類	数
画像	解が曖昧性のない語句になる質問 (例) 日本の国鳥は何ですか。	25
	解が曖昧性のある語句になる質問 (例) 1月の誕生日は何ですか。	25
地図	国に関する質問 (例) 世界最小の国はどこですか。	10
	解が海外の地名になる質問 (例) ブラジルの首都はどこですか。	10
	都道府県・市区町村に関する質問 (例) リンゴの出荷量が多い県はどこですか。	10
	解がランドマークになる質問 (例) 大相撲の聖地はどこですか。	10
	ランドマークの位置を問う質問 (例) 富士急ハイランドはどこにありますか。	10
合計		100

1つの解に対して、画像は上位3件までを、地図は1件のみを評価対象とする。それぞれの画像や地図が質問の回答として適切であったかどうかを以下の3段階で評価し、精度は式(1)で計算する。

- 「」: 適切。その画像や地図に1点を与える。
- 「」: 適切とはいえないが、間違ってもいけないもの。画像の場合、対象以外の余計なものが目立つ場合や、対象が小さくて分かりにくい場合など。地図の場合、正しい位置よりも少しずれている、住所は合っているが位置は間違っているといった場合。この場合はその画像や地図に0.5点を与える。
- 「×」: 不適。この場合の得点は0である。

$$\text{精度} = \frac{(\text{の数}) + 0.5 * (\text{の数})}{(\text{の数}) + (\text{の数}) + (\times \text{の数})} \quad (1)$$

¹⁰処理時間の都合より、3.2.2節の後半で述べた手法はとらず、「ans keywords」をクエリとして画像検索を行なった結果を出力。

¹¹4人は情報工学を専攻する大学院生

¹²正解が複数存在する質問では、そのうちの1つのみを評価対象とする。

4.1 画像を用いて回答する場合の評価

4.1.1 パラメタ設定

予備実験によりパラメタを調整し、表3の条件の下で本実験を行なった。

表3: パラメタ設定

パラメタ	値
画像の周辺テキストを解析する件数 N	40
周辺テキストを切り出す範囲: 画像へのリンク (や<A> タグ) の前後 M 文	30
質問文ベクトルと周辺テキストベクトルの要素の重みを、内容語の品詞の種類に応じて変えるかどうか	変えない
解を抽出した文脈中の語も質問文ベクトルに加えるかどうか	加えない
類似度が非常に低くなる場合、適切な画像が得られないと判断して手法を切り替える。その際の類似度の閾値 TH	0.03

4.1.2 手法

手法1 「ans」をクエリとして画像検索を行なった結果をそのまま出力。(ベースライン1)

手法2 「ans keywords」をクエリとして画像検索を行なった結果をそのまま出力。(ベースライン2)

手法3 手法1で得られた画像の周辺テキストを解析し、質問文との類似度が高いものを出力。

手法4 手法3で類似度が $TH = 0.03$ 以下になったら、手法2へ切り替える¹³。(提案手法)

4.1.3 結果と考察

結果を表4に示す。ここで、「 \square 」「 \square 」「 \times 」の各列の数字は、その評価が与えられた画像の総数を表す。「失敗数」とは、3件の画像の評価が全て \times であった質問の総数を表す。

表4: 画像を表示する場合の結果

手法				\times	精度	失敗数
1	曖昧性なし	60	9	6	0.86	0
	曖昧性あり	23	2	50	0.32	11
	合計	83	11	56	0.59	11
2	曖昧性なし	32	16	27	0.53	3
	曖昧性あり	33	6	36	0.48	4
	合計	65	22	63	0.51	7
3	曖昧性なし	56	10	9	0.81	0
	曖昧性あり	31	5	39	0.45	7
	合計	87	15	48	0.63	7
4	曖昧性なし	57	9	9	0.82	0
	曖昧性あり	40	6	29	0.57	1
	合計	97	15	38	0.70	1

手法1と手法2の結果より、画像検索のクエリを工夫するだけでは全体としての精度向上はあまり望めないことが分かる。

手法3は曖昧性なしの場合の精度をあまり下げることなく、曖昧性ありの場合の精度を上げることに成功している。これより、画像の周辺テキストと質問文の類似度を計算する手法は、解の曖昧性の解消に有効といえる。

手法4では、手法3から曖昧性なしの場合の精度を下げることなく、曖昧性ありの場合の精度だけを大きく向上させている。さらに、失敗数はわずか1となり、残りの49問においては何らかの関連のある画像を表示でき

¹³手法3と手法1は同じ母集合から画像を選択している。手法2へ切り替えるのは、手法2がある程度曖昧性を解消でき、さらに処理時間がそれほどかからないことによる。

たことになる。これより、提案手法の有効性を示すことができたといえる。

4.2 地図を用いて回答する場合の評価

4.2.1 パラメタ設定

検索エンジン(Goo)を利用する際、Snippetの取得件数は50件とした。

4.2.2 結果と考察

表5: 地図を表示する場合の結果

			\times	精度
国名	10	0	0	1.00
海外の地名	8	1	1	0.85
都道府県・市区町村	9	1	0	0.95
解がランドマーク名	7	2	1	0.80
ランドマークの位置	5	3	2	0.65
合計	39	7	4	0.85

結果を表5に示す。上手いかなかったケースとしては、海外の地名で「サハラ」のような大きな領域を持つ場合、住所検索は成功したがジオコーディングの精度の問題で位置がずれてしまった場合、「landmarkはどこですか」という質問でランドマーク名の抽出に失敗した場合、「ムンク美術館」といった海外のランドマークの場合などがあつた。経緯度を直接求めることができなかつた海外のランドマークは現状では失敗になってしまう。

上記の失敗以外は概ね意図した通りに動作した。位置がわずかにずれた場合でも、表示された地図はユーザが自由に動かせるため、それほど問題はないと思われる。

5 まとめと今後の課題

本稿では、質問応答の回答として、文字による解だけでなく画像や地図も表示できるシステムを提案した。評価実験の結果、画像は0.70、地図は0.85の精度で適切なものを表示できることが分かった。また、複数の正解を持つ質問に対して対応する画像をリストにして表示したり、一つの地図上に複数の解の位置をまとめて表示するといった、質問応答システムにおける新しい回答の提示方法を示した。

今後の課題としては、各処理の精度の向上だけでなく、インタフェースとしての評価や、他のメディアへの拡張が挙げられる。また、画像や地図を使って質問できるシステムも考えられる。

参考文献

- [1] Atsushi Fujii, Katunobu Itou, and Tetsuya Ishikawa. Cyclone: An Encyclopedic Web Search Site. In *The 14th International World Wide Web Conference (WWW2005)*, pp. 1184-1185, 2005.
- [2] Atsushi Fujii, Katunobu Itou, and Tetsuya Ishikawa. Image Retrieval and Disambiguation for Encyclopedic Web Search. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pp. 1598-1599, 8 2005.
- [3] Tatsunori Mori. Japanese Question-answering System Using A* Search and Its Improvement. *Transactions on Asian Language Information Processing (TALIP)* Vol. 4, No. 3, pp.280-304, ACM, 2005.
- [4] 石下円香, 森辰則. 優先順位型質問応答の解スコア分布に基づくリスト型質問応答. 自然言語処理研究会報告 2005-NL-169, pp.41-47, 情報処理学会, 9 2005.
- [5] 東京大学空間情報科学研究センター. CSIS シンプルジオコーディング実験 (街区レベル位置参照情報) <http://pc035.tkl.iis.u-tokyo.ac.jp/~sagara/geocode/modules/simple-geocode1/>.
- [6] 佐藤理史. ワールドワイドウェブを利用した住所探索. 情報処理学会論文誌 Vol.42 No.1, pp.59-67, 情報処理学会, 1 2001.