質問の曖昧性を検出し複数の解答を提示する質問応答システム

松本　匡史　　白井　清昭
北陸先端科学技術大学院大学 情報科学研究科
{t-matsu, kshirai}@jaist.ac.jp

1 はじめに

オープンドメインな質問応答システムにおいて，ユーザの質問に対する答えは常に一つであるとは限らない．質問によっては複数の解答を返すべきものもある．複数の解答を返すべき質問には少なくとも以下の 2 種類がある．

1. 複数の解答が存在する質問
   例えば「日本三大祭りは何ですか?」や「日本の歴代の首相は誰ですか?」のような質問が該当する．
2. 質問が曖昧であるために解答を一つに絞ることができない質問
   例えば,「ワールドカップの優勝国はどこですか」という質問は，ワールドカップにはサッカー，ラグビーなどの種類があるという意味で曖昧であり，スポーツの種目に応じて解答が複数存在する．

1. のような質問について複数の解答を返す試みは，リスト型質問応答システムとして既に多くの研究が行われている [2, 3]．一方，2. のような質問を取り扱う研究としては徳江らによるものがある [5]．徳江らは，曖昧性を持つ質問に対し，ユーザとの対話を通じて質問の曖昧性を解消し，適切な解答を一意に絞り込む対話的質問応答システムを提案している．

本研究では，特に 2. のような曖昧な質問を対象とし，複数の解答をリストとして提示する質問応答システムを提案する [4]．曖昧な質問とは，ここでは，質問文中に含まれるキーワードの意味が曖昧であると定義する．解答のリストを表示する際，結果を分かりやすく伝えるために，解答を曖昧なキーワードの意味とともに表示する．本研究が考える理想的な複数解答の返答例を図 1 に示す．図 1 の例では，イギリス，ブラジル，ノルウェーという複数の解答が優先順位付きで出力されている．また，質問文中の「ワールドカップ」が曖昧なキーワードであると判断され，それぞれの解答候補に対する「ワールドカップ」の意味が括弧内に表示されている．ここで重要なのは，ユーザは自分の質問の曖昧性に気づいていない場合でも，図 1 のような出力を見ることで質問の曖昧性に気づき，適切な答えを選択できるという点である．もしキーワードの意味の情報を提示せず，解答リストのみを提示した場合，サッカーのワールドカップの優勝国を知りたかったユーザが (ラグビーワールドカップの優勝国である) イギリスを解答だと誤る可能性がある．

質問: ワールドカップの優勝国はどこですか?
解答: [1] イギリス (ラグビーのワールドカップ)
[2] ブラジル (サッカーのワールドカップ)
[3] ノルウェー (スキーのワールドカップ)
⋮

図 1: 提案システムの理想的な出力例

2 提案システム

提案する質問応答システムの処理の流れは以下に示す．

1. 質問文解析
   質問文の質問タイプを同定しキーワードを抽出する．
2. 文書検索
   キーワードをクエリとした検索を行う．
3. 解答候補抽出
   検索された文書の中から解答候補を抽出する．
4. 限定表現候補の抽出
5. 解答群の生成
6. 解答群の順位付け
7. 解答リスト作成・提示

処理 1〜3 は通常の質問応答システムとほぼ同じである．処理 3 が終了した時点で，次の情報が得られているものとする．

- 質問文のキーワード k_i $(1 \leq i \leq K)$
- 解答候補 a_j $(1 \leq j \leq A)$
- a_j に対して質問応答システムが与えるスコア $s(a_j)$
- a_j が抽出された文書
  文書は a_j と全てのキーワード k_i を含む．

残りの処理については次項以降で述べる．

2.1 限定表現候補の抽出

本研究では，質問文に含まれる曖昧なキーワードに着目して解答リストを作成する．キーワード k_1〜k_K の中から曖昧なキーワードを一つ選別するために，キーワー

ドを修飾しかつキーワードの意味を限定するような表現に着目する．例えば，文書中に

... サッカー ワールドカップ の優勝国・ブラジルで...

という文があったとき,「ワールドカップ」の直前にある「サッカー」はワールドカップの意味を限定する表現とみなせる．本研究ではこのような表現をキーワードの限定表現と呼ぶ．そして，同じキーワードが解答候補を含む文書別に異なる限定表現を持てば，そのキーワードは意味的に曖昧であるとみなす．

徳江らは，限定表現を抽出する際に，(1) キーワードに助詞「の」を介して連体修飾する名詞を抽出する，(2) キーワードの直前または直後にあり，キーワードともに複合名詞を構成する名詞を抽出する，など，いくつかの抽出パタンを用いて限定表現を抽出していた [5]．しかし，用いたパタンが少なかったため，限定表現の抽出に失敗することが多いという問題点があった．これに対し，本研究では限定表現を網羅的に獲得するために，徳江らが用いたパタンによって抽出される名詞に加え，キーワードと同一文書内に出現する全ての名詞を限定表現の候補とする．ただし，このままでは数多くの不適切な候補が抽出されるのは自明である．そこで，キーワードと名詞間の関連度を計算し，関連度の高い名詞のみを限定表現の候補とする．

名詞間の関連度はコーパスにおける文書内共起に基づいて定義する．具体的には式 (1) の Dice 係数を用いる．

$$D(x,y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

X, Y はそれぞれ名詞 x, y が出現する文書集合であり，$X \cap Y$ は x,y をともに含む文書集合である．$D(x,y)$ は毎日新聞の 13 年分の新聞記事から計算した．頻度が 100 以上かつ $D(x,y) \geq 0.0005$ となる全ての x と y の組について $D(x,y)$ の値を事前に計算し，テーブルとして格納した．限定表現の候補を抽出する際には，キーワードと記事中の名詞に対して $D(x,y)$ を求め，これがある一定の閾値 (本研究では 0.04) 以上のときに限定表現の候補とした．また，事前に作成した Dice 係数の表の中に見つからない x と y の組については，x, y を先頭から 1 文字ずつ順番に削り，$D(x,y)$ が見つかった時点で，そのダイス係数の値を x と y の関連度とした．

ここまでの処理の結果，(a_j, k_i, s_k) という 3 つ組が大量に取得される．ここで s_k は限定表現の候補，k_i は限定表現 s_k に対するキーワード，a_j は s_k, k_i が抽出された文書における解答候補である．抽出例を図 2 に示す．図 2 では，(豪州, ワールドカップ, ラグビー) といった適切な限定表現も抽出されれば，キーワードの意味を限定しているとは言えない限定表現も抽出されていることがわかる．

記事 (a_j=豪州, k_1=ワールドカップ, k_2=優勝国)

ＳＯＣＯＧのコーツ副会長は豪州が ワールドカップ（W杯）優勝国 であり、ラグビーに対する国民の関心が高いことを指摘しながらも、「（シドニー五輪の）選手総数をまず１万人に抑えるのが原則。１チーム２５人の編成が他の競技の選手数を圧迫することも考えなければならない」と述べた。

抽出された限定表現の候補 (a_j, k_i, s_k)

(豪州, ワールドカップ, W杯), (豪州, ワールドカップ, ラグビー), (豪州, 優勝国, 国民), (豪州, 優勝国, 1 万人), (豪州, 優勝国, 1 チーム 25 人), (豪州, ワールドカップ, 競技)

図 2: 限定表現の抽出例

2.2 解答群の作成

解答候補の集合 $A = \{a_1, \cdots, a_A\}$ の中から最終的にリストとして提示するべき解答群 (A の部分集合) を作成する．ここでの目的は，キーワードの曖昧性を的確に反映し，どのキーワードがどのような意味で曖昧であったのかをユーザが容易に理解できるような解答の部分集合を見つけることである．例えば，

- (イギリス, ワールドカップ, ラグビー)
- (ブラジル, ワールドカップ, サッカー)
- (ノルウェー, ワールドカップ, スキー)

という解答群を見れば，ワールドカップの種目が曖昧であるとユーザはすぐに理解できる．一方，

- (南ア, 優勝国, 筆頭)
- (フランス, 優勝国, 1998 年)
- (フィンランド, 優勝国, 連続)

という解答群を見ても,「優勝国」というキーワードにどのような曖昧性があるのか理解できない．したがって，このときは前者の解答群をユーザに提示すべきである．

本研究では，解答群のそれぞれの解答に対応するキーワードの限定表現がある程度似たような性質を持っていなければ，キーワードの意味の曖昧性を適切に表現できないと考える．そこで，2.1 項で抽出された (a_j, k_i, s_k) の集合から，キーワード k_i が共通で，かつその限定表現 s_k が共通の属性 $attr$ を持つものを選別し，これを解答群 $AG(k_i, attr)$ とする．ここで，属性 $attr$ とは以下の種類がある．

AG(ノーベル, $E2$: 学賞)= {
(江崎玲於奈, 物理学賞), (湯川秀樹, 物理学賞), (白川英樹, 化学賞), (福井謙一, 化学賞), (大江健三郎, 文学賞), (利根川進, 生理学賞)　}

AG(高校野球, $\langle NUM \rangle$ 回選抜)= {
(沖縄尚学, ７１回選抜), (東海大相模, ７２回選抜), (北野高, ２０回選抜)　}

図 3: 解答群の例

- 末尾 N 文字
  限定表現の末尾 N 文字を属性とみなす (1≤N≤3). 例えば,「60 キロ級」は「E3:キロ級」「E2:ロ級」「E1:級」という 3 つの属性を持つ.
- 意味クラス
  シソーラスによって得られる限定表現の意味クラスを属性とみなす. シソーラスは日本語語彙体系 [1] を用いた.
- 数量表現+接尾語
  数字と接尾語から構成される限定表現が持つ属性. 例えば,「60 キロ級」は「⟨NUM⟩ キロ級」という属性を持つ.
- 括弧
  限定表現が括弧で囲まれているときはこの属性を持つとみなす. 新聞記事中には製品名などの固有名詞は括弧で囲まれることが多いためにこの属性を設定した. 例えば,「スペースシャトル」の限定表現『「エンデバー」』はこの属性を持つ.

(a_j, k_i, s_k) の集合から, 可能なキーワードと属性の組み合わせについて, 解答群の候補 $AG(k_i, attr)$ を得る. 得られた解答群の例を図 3 に示す. $AG(k_i, attr)$ 中の要素は本来ならば (a_j, k_i, s_k) という 3 つ組だが, 解答群内ではキーワード k_i は全て等しいので, 図 3 中では (a_j, s_k) と省略表示している. AG(ノーベル, $E2$: 学賞) は「学賞」という末尾を持つ限定表現を共通に持つ解答を 1 つのグループにまとめたもので, ノーベル賞の種類に曖昧性があり, それによって解答 (ノーベル賞受賞者) が異なることを示している. 一方, AG(高校野球, $\langle NUM \rangle$ 回選抜) は数量表現に「回選別」という表現が続く限定表現を共通に持つ解答を 1 つにしたものであり, 選抜高校野球の開催回数によって解答 (優勝校) が異なることを示している.

2.3 解答群の順位付け

2.2 項までの処理によって, キーワードや限定表現の属性の種類に応じて多数の解答群が作成される. ここでは, これら解答群の中から, 最終的にユーザに提示する解答群をひとつ選択する.

解答群 $GA(k_i, attr)$ に対するスコアを式 (2) のように 4 つのサブスコア S_1,S_2,S_3,S_4 の重み付き和であると定義する.

$$Score(GA(k_i, attr)) = w_1 S_1 + w_2 S_2 + w_3 S_3 + w_4 S_4 \quad (2)$$

$$S_1 = \frac{|GA|}{A},\ S_2 = \frac{A_{type}}{|GA|},\ S_4 = \frac{\sum_{a \in GA} s(a)}{|GA|} \quad (3)$$

w_i はサブスコア S_i に対する重みである. 本研究では, いくつかの質問を入力したときの結果を参照し, 人手でこれらの重みを決定した. 具体的には $w_1 = 0.3$, $w_2 = 0.4$, $w_3 = 0.2$, $w_4 = 0.1$ とした.

次に各サブスコアの詳細について説明する. S_1 は解答群中の解答数を全解答候補数 (A) で割った値である. S_1 は解答群が多くの解答を含めば含むほど高いスコアを与える働きをする.

S_2 の定義式 (式 (3) の中央) において, A_{type} は解答群中の解答の異なり数であり, $|GA|$ は解答ののべ数である. もし, 解答の異なり数が少ない場合, 同じ解答候補が異なる限定表現を持つことが多いということを意味する. しかし, 適切に限定表現が抽出されているのなら, 一つの解答に対して得られる限定表現は 1 つのはずである. S_2 は, 異なる限定表現が同じ解答候補に対して出現する場合に低いスコアを与える働きをする.

S_3 は属性の種類によって与えられるスコアである. これは全て人手により設定し, $attr$ が「括弧」「数量表現+接尾語」「末尾 3 文字」のときは 1,「末尾 2 文字」のときは 0.5,「末尾 1 文字」のときは 0.2,「意味クラス」のときは 0.4 となる. スコアが高い属性ほど限定表現間の共通性が高く, ユーザに提示する解答リストとして適切であるとみなしている.

S_4 は解答群中の解答の信頼性を表わしている. 式 (3) の $s(a)$ は解答候補 a に対して我々の質問応答システムが与えるスコアである. すなわち, S_4 はこのスコアの平均であり, 信頼性の高い解答を多く含む解答群に高いスコアを与える働きをする.

全ての解答群について式 (2) のスコアを計算し, 最も高いスコアを持つ解答群をひとつ選択する.

2.4 解答リスト提示

選択した解答群を解答リストとして出力する. 解答の出力順序は, 質問応答システムが出力する解答候補のス

コアの順とする．さらに，図 1 の ( ) 内のように，曖昧なキーワードの意味の情報も出力する．現在，各解答候補には共通のキーワードと限定表現が取得されているので，これを単に並べて出力するだけでもユーザに質問の曖昧性を認識させる効果があると考えられる．とはいえ，キーワードと限定表現から図 1 のようなわかりやすい表現を生成することも重要である．本研究は解答リストの生成に焦点を当て，キーワードの意味の曖昧性をユーザにわかりやすく伝える文の生成までは行っていないが，今後の重要な課題と考えている．

3 予備実験

2 節で述べた手法によって得られた解答群がどれだけ適切であるかを調べる実験を行った．31 個の曖昧な質問を用意し，2.1〜2.3 項の手法によって得られた解答群が適切であるかを人手でチェックした．ここで解答群が適切であるとは，正しい解答を複数個含み，かつ得られた限定表現がキーワードの意味の曖昧性を正しく反映していることを指す．実験結果を表 1 に示す．

表 1: 実験結果

適切な解答群が得られない (a)	8 (26%)
適切な解答群が得られた	
適切な解答群が最上位 (b)	7 (22%)
適切な解答群が最上位以外 (c)	16 (52%)

適切な解答群が得られた質問 (表 1(b)+(c)) の割合は約 74%であった．この結果から，提案手法によって質問の曖昧性を検出し，またキーワードの意味によって異なる解答を抽出することがある程度可能であるとわかった．適切な解答群が得られなかった場合 (表 1(a)) の主な原因は，解答そのものの抽出に失敗したこと，想定していた限定表現が新聞記事に現われていなかったこと，などである．後者の例を挙げると,「小泉内閣の官房長官は誰ですか」という質問に対し,「小泉内閣」というキーワードに第 1 次，第 2 次，第 3 次といった曖昧性があることが検出されると期待したのだが，少なくとも官房長官の解答候補が抽出された記事には「第○次小泉内閣」という表現はなかった．また，2.1 項で述べたように，限定表現の候補としてキーワードとの関連度がある閾値以上の名詞を抽出しているが，実装上の制約から，関連度計算できる名詞は新聞記事に 100 回以上出現した名詞のみとしている．それ以外の名詞については関連度を計算できないため，限定表現として抽出されることはない．ところが，そのような名詞の中にも限定表現としてふさわしいものがいくつか見つかっている．したがって，今後は限定表現の抽出方法を改善する必要がある．

一方，表 1 から，適切な解答群が抽出されれば，式 (2) のスコアによって適切な解答群が 1 位になる割合は約 30%であることがわかる．1 位にならなかった質問 (表 1(c)) について，その誤りの原因を調査したところ，主に以下の要因があることがわかった．まず，限定表現の異表記の問題である．例えば,「シドニー五輪の柔道の金メダリストは誰ですか?」という質問に対し，柔道の階級の曖昧性を表わす限定表現として「48 キロ級」と「女子 48 キロ級」が得られた．現在，これらは別々の限定表現とみなされるが，本来は同じ限定表現とするべきである．また，限定表現の出現位置によってスコアを変えることも有望であるとわかった．例えば,「48 キロ級 金メダリスト」や「48 キロ級 の金メダリスト」のように，キーワードと隣接していたり，キーワードに助詞「の」を介して連体修飾する限定表現は，単にキーワードの近傍に出現する限定表現よりも限定表現として適切である場合が多い．したがって，キーワードの限定表現になりやすい位置に出現する限定表現を多く持つ解答群に高いスコアを与えるようにすれば，適切な解答群が 1 位になる可能性が高くなると期待できる．

4 おわりに

本研究では，曖昧な質問が入力されたために複数の解答が正解に該当するとき，質問文中の曖昧なキーワードを検出し，複数の解答をそのキーワードの意味の情報とともにリストとして出力する質問応答システムについて述べた．今後は，限定表現の抽出方法や解答群に与えるスコアの見直し，限定表現の異表記の取り扱い，ユーザに理解しやすいキーワードの意味情報の提示方法の検証，などの課題に取り組む予定である．

参考文献

[1] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩己, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙体系. 岩波書店, 1997.

[2] 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. 情報処理学会自然言語処理研究会, Vol. 2005, No. 94, pp. 41–47, 2005.

[3] 加藤恒昭, 桝井文人, 福本淳一, 神門典子. リスト型質問応答の特徴付けと評価指標. 情報処理学会自然言語処理研究会, Vol. 2004, No. 93, pp. 115–122, 2004.

[4] 松本匡史. 質問の曖昧性を考慮した質問応答システムに関する研究. Master's thesis, 北陸先端科学技術大学院大学, 3 2006.

[5] 徳江英範, 白井清昭. 対話型質問応答システムにおける質問の曖昧性の検出. 第 11 回言語処理学会年次大会, pp. 1092–1095, 2005.