

対話型質問応答システムにおける質問の曖昧性の検出

徳江 英範 白井 清昭

北陸先端科学技術大学院大学 情報科学研究科

{htokue, kshirai}@jaist.ac.jp

1 はじめに

本研究はオープンメインな対話型質問応答システムの構築を目的とする [5]。本研究における対話型質問応答システムとは、ユーザによって曖昧な質問が入力されたとき、ユーザとの対話を通じてその曖昧性を解消し、適切な解答を返すシステムである。ここでの曖昧な質問とは、質問中の単語の意味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を指す。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、これに対する解答を一意に決めることはできない。このような曖昧な質問が入力されたとき、「どの競技のワールドカップですか」といった問い合わせをシステムからユーザに行い、これに対するユーザからの返答によって適切な解答を決める。

従来のオープンメインな質問応答システムは、TREC や QAC [3] などの評価型ワークショップの参加システムに代表されるように、ユーザの質問に対してシステムが解答を返すだけで処理が終了する一問一答型のシステムが主流である。これに対し、本研究では、対話を通じてユーザの質問に対する適切な解答を提示するシステムの構築を目指す。対話型質問応答システムの先行研究としては、京都大学の学術情報メディアセンターのヘルプシステム [2]、ダイアログナビ [1]、ユーザとの対話を通じて情勢分析やレポート作成を支援する HITIQA [4] などがある。これらの多くは対象となるドメインを限定し、ユーザとの対話の内容に関してもある程度のシナリオが用意されているのに対し、本研究はドメインを限定しない質問応答システムを対象とする点が異なる。

2 提案システム

本研究で提案する質問応答システムの概要を図 1 に示す。本システムは、まずユーザによって入力された質問文の解析を行い、その結果得られるキーワードをクエリーとして文書検索を行う。次に、得られた文書集合から解答候補を抽出する。ここまでの処理は通常の質問応答システムとほぼ同じである。

解答候補が得られたら、ユーザの質問の曖昧性を検出する。具体的には、ユーザの質問に含まれる曖昧なキー

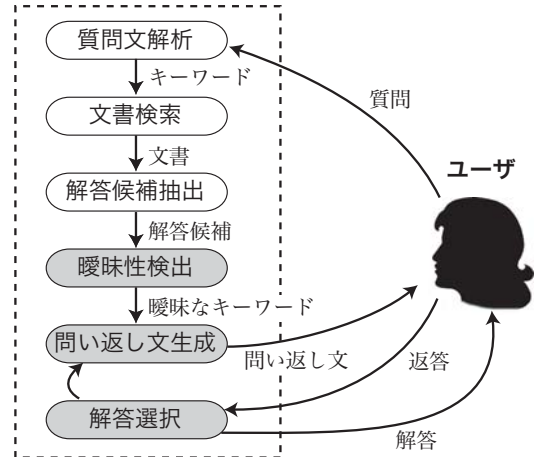


図 1: 提案システムの概要

ワードを検出する。例えば、1 節で挙げた「ワールドカップの優勝国はどこですか」という質問では、「ワールドカップ」の意味が曖昧であることを検出する。次に、曖昧なキーワードの意味が一意に決まるような問い合わせ文（例えば「どの競技のワールドカップですか」）を生成する。問い合わせ文に対するユーザの返答（例えば「サッカーのワールドカップです」）を受け取り、それにしたがって解答候補の中から適切な解答を絞り込む。このような対話を解答候補が一意に決まるまで繰り返し、最後に解答をユーザに提示する。

本論文では、図 1 に示した処理のうち、曖昧性検出、すなわちユーザの質問に含まれる曖昧なキーワードを検出する手法について述べる。

3 曖昧性検出

本節では、ユーザの質問に含まれる曖昧なキーワードを検出する手法について説明する。まず、提案手法の概要を例を挙げて説明する。「ワールドカップの優勝国は?」という質問が与えられたとき、以下の 2 つの文書からブラジル、イギリスという解答候補が得られたとする。

... サッカーのワールドカップの優勝国は ブラジル で...

... ラグビーのワールドカップ優勝国は イギリス で...

曖昧なキーワードを検出するために、本研究では、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、「ワールドカップ」に連体修飾

する「サッカーの」や「ラグビーの」という句は、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、同じキーワードが記事毎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。

3.1 限定表現の抽出

ユーザの質問の曖昧性を検出するため、まず最初にキーワードの限定表現を抽出する。具体的には以下の5種類の限定表現を考える。

- 連体修飾句 (s_{no})
助詞「の」を介してキーワードに連体修飾する句。例えば、「48キログ級の金メダリスト」という表現があったとき、「金メダリスト」の限定表現 s_{no} は「48キログ級」となる。
- 直前の複合名詞 (s_{prev})
キーワードの直前にあり、キーワードとともに複合名詞を構成する名詞。例えば、「女子柔道」という表現があったとき、「柔道」の限定表現 s_{prev} は「女子」となる。
- 直後の複合名詞 (s_{succ})
キーワードの直後にあり、キーワードとともに複合名詞を構成する名詞。例えば、「柔道60キログ級」という表現があったとき、「柔道」の限定表現 s_{succ} は「60キログ級」となる。
- デ格的格要素 (s_{de})
キーワードがある用言の格要素であるとき、同じ用言を主辞とするデ格的格要素。例えば、「81キログ級で金メダルを獲得した」という表現があったとき、「金メダル」の限定表現 s_{de} は「81キログ級」となる。
- 日付表現 (s_{date})
キーワードの近辺に現われる日付表現。具体的には、固有表現タグとして“DATE”が付与され、キーワードの近辺にあり、かつ動詞の格要素となる表現を抽出する。日付表現が月日を含む場合は年のみを取り出す。また、「今年」という表現は記事の掲載日を参照して西暦年に正規化する。さらに、日付表現の主辞となる動詞も同時に抽出する。例えば、「2000年7月に開催された世界大会」という表現があったとき、限定表現 S_{date} は [2000年; 開催する] となる。
限定表現を抽出する手続きは以下の通りである。まず、解答候補 a_i に対して、もっとも近くにあるキーワード k_j を見つける。次に、キーワードを含む文に対して形態素解析¹、文節の係り受け解析²、固有表現抽出²を行

表 1: 限定表現の抽出例

		S_{succ} (直後の名詞)
$k_1 = \text{金メ}$	$a_1 = \text{田村}$	
	$a_2 = \text{野村}$	
	$a_3 = \text{滝本}$	・(1)
$k_2 = \text{五輪}$	$a_1 = \text{田村}$	女子 (3), 代表選手・役員 (1)
	$a_2 = \text{野村}$	
	$a_3 = \text{滝本}$	競泳 (1)
$k_3 = \text{柔道}$	$a_1 = \text{田村}$	女子48キログ級 (2), 部 (1)
	$a_2 = \text{野村}$	60キログ級 (1)
	$a_3 = \text{滝本}$	81キログ級 (1)

($k_1 = \text{金メダリスト}$, $k_2 = \text{シドニー五輪}$, $k_3 = \text{柔道}$
 $a_1 = \text{田村亮子}$, $a_2 = \text{野村忠宏}$, $a_3 = \text{滝本誠}$)

表 2: 日付の限定表現の抽出例

	S_{date} (日付表現)
$a_1 = \text{白川英樹}$	[1937年; 生まれる](1), [2000年; 受賞する](4)
$a_2 = \text{野依良治}$	[2001年; 受賞する](2), [2001年; 決まる](1)
$a_3 = \text{福井謙一}$	[1981年; 受賞する](1)

($k_1 = \text{受賞}$, $k_2 = \text{ノーベル}$, $k_3 = \text{化学賞}$)

い、各キーワードについて上記5種類の限定表現を抽出する。ただし、抽出した限定表現の中にキーワードが含まれる場合は、キーワードの意味を限定する働きをしないことがほとんどであったため、限定表現からキーワードに相当する文字列を除去した。抽出された限定表現の例を表1に示す。括弧内の数値は限定表現の出現頻度を表わす。表1では直後の複合名詞 s_{succ} のみを示しているが、他の種類の限定表現も同様に抽出される。

5種類の限定表現のうち、日付表現 s_{date} だけは特別な取り扱いをする。日付表現は、他の限定表現のようにキーワードの意味を限定するのではなく、キーワードを含む文によって表わされるイベントの意味を限定するとみなす。日付表現と同時にその主辞となる動詞も取り出したのは、抽出した動詞が日付表現が限定するイベントを表わすとみなせるからである。また、日付表現はキーワードの意味を限定しているわけではないので、表1のようにキーワードの区別せず、解答候補 a_i 毎に抽出する。 s_{date} の抽出例を表2に示す。この例では、3人のノーベル化学賞の受賞者の解答候補について、それぞれの受賞年が「受賞する」というイベントの限定表現として抽出されている。

¹茶筌を使用した。

²南瓜を使用した。

3.2 曖昧なキーワードの検出

表 1 のような限定表現の集合から、曖昧なキーワード k_x と $S(k_x, type, attr)$ を決める。 $S(k_x, type, attr)$ とは、種類が $type(no, prev, succ, de, date)$ のいずれか) であるようなキーワード k_x の限定表現のうち、共通の属性 $attr$ を持つものの集合である。ここでは、限定表現の属性として以下の 5 種類を考える。

- 末尾 N 文字
限定表現の末尾 N 文字を属性とみなす ($1 \leq N \leq 3$)。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を用いた。例えば、「女子」は「0046³」という属性を持つ。
- 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。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。
- 日付表現に係る動詞
限定表現が日付 (s_{date}) のとき、その日付に係る動詞。具体的には、3.1 項で述べたように、日付と同時に抽出する動詞を指す。また、日付の限定表現はこれ以外の属性は持たないとする。

一般に $S(k_x, type, attr)$ は複数得られる。表 1 で示した解答候補とキーワードに対して得られる $S(k_x, type, attr)$ の例を以下に挙げる。

$$S(k_3, succ, (NUM)+キロ級) = \{48 \text{ キロ級}, 60 \text{ キロ級}, 81 \text{ キロ級}\} \quad (1)$$

$$S(k_1, prev, E3:キロ級) = \{48 \text{ キロ級}, \text{女子 } 48 \text{ キロ級}, 60 \text{ キロ級}\} \quad (2)$$

$$S(k_3, prev, 0046) = \{\text{女子}, \text{男子}, \text{男子}\} \quad (3)$$

$S(k_x, type, attr)$ は問い合わせ文の内容を間接的に表わしていることに注意していただきたい。例えば、式 (1) は柔道 (k_3) の階級を、式 (2) の S は金メダリスト (k_1) の階級を、式 (3) は柔道選手の性別を問う質問文を生成することに対応するとみなせる。

³日本語語彙体系における「人間<生物学的特徴>」という意味クラスを表わす。

ここで、属性 $attr$ を導入した理由は以下の通りである。限定表現を抽出する際、解析誤りを始めとする様々な要因のため、キーワードの意味を限定しているとはいえない表現も抽出される。例えば、表 1 の k_3 の s_{succ} として抽出された限定表現の多くは「～キロ級」のような階級を表わす表現であるが、中には「部」のように階級とは無関係な限定表現も含まれる。問い返し文を生成することを考えれば、このような無関係な限定表現はあらかじめ削除したい。本研究では、これを共通の属性を持たない限定表現を排除することで実現している。

3.3 スコア付け

次に、複数の $S(k_x, type, attr)$ の候補の中から、最適な問い返し文を生成できそうなものを 1 つ選ぶ。具体的には、 $S(k_x, type, attr)$ のスコアを図 2 のように定義し、最大の値を持つ $S(k_x, type, attr)$ を選択する。各項①～⑤の詳細を以下に述べる⁴。

- ① $|A|$ は解答候補の数、 $|A_{attr}|$ は属性 $attr$ を持つ限定表現を持つ解答候補の数である。この項は、 $S(\cdot)$ が多くの解答候補に対する限定表現を含めば含むほど高いスコアを与える働きをする。例えば、式 (1) に対しては 3/3、式 (2) に対しては 2/3 となり、前者に高いスコアを与える。
- ② 分母は $S(\cdot)$ の要素数、分子の $NT(S(\cdot))$ は $S(\cdot)$ の要素の異なり数である。この項は、ユーザへの問い返しによって解答候補をどれだけ効率良く絞り込むことができるかを測っている。例えば、式 (1) については 3/3、式 (3) については 2/3 となる。後者の場合、ユーザの返答が「男子」のときは a_2, a_3 のいずれの解答が適切かを判断することができないので、スコアは前者よりも低い。
- ③ $O(s)$ は限定表現 s の出現頻度を表わす。この項の分母は、種類が $type$ であるような k_x の限定表現の頻度の和であり、分子は上記のうち属性 $attr$ を持つ限定表現の頻度の和である。直観的に言えば、同じキーワードに対する限定表現のうち、同じ属性を持つ限定表現の割合が大きい $S(\cdot)$ ほど高いスコアを与える働きをする。例えば、式 (1) に対しては 5/6 となる。
- ④ $S(attr)$ は属性の種類によって与えられるスコアである。例えば、 $attr$ が「末尾 3 文字」のときは「末尾 1 文字」のときよりも高いスコアを与える。これらは全て人手で設定した。

⁴図 2 ならびに以下の説明において、 $S(\cdot)$ は $S(k_x, type, attr)$ の簡略表記である。

$$Score(S(k_x, type, attr)) = \left\{ w_1 \cdot \frac{|A_{attr}|}{|A|} \textcircled{1} + w_2 \cdot \frac{NT(S(\cdot))}{|S(\cdot)|} \textcircled{2} + w_3 \cdot \frac{\sum_{s \in S(\cdot)} O(s)}{\sum_{attr} \sum_{s \in S(\cdot)} O(s)} \textcircled{3} \right\} \\ \times S(attr) \textcircled{4} \times \frac{\sum_{s \in S(\cdot)} O(s)}{|S(\cdot)|} \textcircled{5}$$

図 2: $S(k_x, type, attr)$ のスコア

- ⑤ S 中の限定表現の平均出現頻度である。出現頻度が大きい限定表現を多く含む $S(\cdot)$ に高いスコアを与える。式 (1) については 6/5 となる。

w_i は項①～③の重みである。本論文ではこれらを人手で設定し、 $w_1 = 5$, $w_2 = 1$, $w_3 = 4$ とした。また、例外処理として、限定表現を持つ解答候補が 1 のとき ($|A_{attr}| = 1$), また $S(\cdot)$ の異なり数が 1 のとき ($NT(S(\cdot)) = 1$) は、解答候補を絞り込むための問い返し文を生成できないことは明らかなので、スコアは 0 とした。

限定表現が日付 (s_{date}) のときは、表 2 に例示したようにキーワードの区別をしないで限定表現を抽出する。したがって、限定表現の集合は $S(k_x, \mathbf{date}, attr)$ ではなく、 $S(\mathbf{date}, attr)$ となる。スコアの計算は他の限定表現と同様に図 2 の式で求める。

抽出された限定表現の集合から得られる全ての $S(\cdot)$ のうち、スコアが最大となるものを選択する。表 1 については式 (1) の $S(k_3, succ, (NUM)+キロ級)$ が最大となる。これは、問い合わせ文の内容が決定されると同時に、ユーザの最初の質問に含まれる曖昧なキーワード k_x (この場合は「柔道」) が検出されたことも意味する。

4 予備実験

曖昧性検出手法を評価する簡単な予備実験を行った。曖昧な質問を 33 個用意し、それぞれに対し解答候補を 5 つ抽出した。さらに、図 2 のスコアに基づいて $S(k_x, type, attr)$ をひとつ選択し、それがユーザへの問い合わせ文の内容として適切かどうかを人手で判断した。結果を表 3 に示す。

表 3 (a) は問い合わせ文を生成するために必要な限定表現を抽出することができなかった質問の数であり、全体の約 21% に相当する。これは限定表現を抽出する手法に不備があることを示唆している。本論文では 3.1 項に挙げた 5 種類の限定表現しか考慮していないが、キーワードの意味を限定する表現はもっと多様であると考えられる。したがって、限定表現を網羅的に獲得する手法を確立する必要がある。一方、問い返し文を生成するための限定表現が抽出されている状況 (表 3 の (b), (c)) においては、おおむね適切な $S(\cdot)$ を選択することができた。具体的には、ノーベル賞の受賞者を問う質問に対し

表 3: 予備実験結果

(a) 適切な $S(k_x, type, attr)$ が存在しない	7
(b) 選択した $S(k_x, type, attr)$ が不適切	5
(c) 選択した $S(k_x, type, attr)$ が適切	21

て「物理学賞」「化学賞」といった賞の種類に曖昧性があることや、ノーベル化学賞の受賞者を問う質問に対して受賞した年に曖昧性があることを検出することができた。一方、適切な $S(\cdot)$ の選択に失敗する主な原因は、複数の限定表現の共通の属性をうまく認識できないことである。例えば、キーワード「殺人事件」に対して、 s_{prev} として「保険金(殺人事件)」や「愛犬家(殺人事件)」は抽出されたが、これらに共通の属性が存在しないため、キーワードの曖昧性の検出に失敗している。

今後は、上記の考察を考慮して提案手法を改良するとともに、より大規模な評価実験を行う予定である。

5 おわりに

曖昧な質問が入力されたとき、ユーザとの対話を通して曖昧性を解消し、適切な解答を選択する対話型質問応答システムの構想について述べた。また、その要素技術として、ユーザの質問の曖昧性を検出する一手法を提案した。今後は、曖昧性検出手法を改良するとともに、問い返し文の生成や、ユーザの返答に基づく解答候補の絞り込みなども行い、提案システムの完成を目指す。

参考文献

- [1] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答-ダイアログナビ-. 自然言語処理, Vol. 10, No. 4, pp. 145-175, 2003.
- [2] Sadao Kurohashi and Wataru Higasa. Dialogue help-system based on flexible matching of user query with natural language knowledge base. In *Proceedings of the 1st ACL SIGdial Workshop on Discourse and Dialogue*, pp. 141-149, 2000.
- [3] National Institute of Informatics. *Proceedings of the Third NTCIR Workshop*, 2002.
- [4] Sharon Small et al. HITIQA: Towards analytical question answering. In *Proceedings of the COLING*, pp. 1291-1297, 2004.
- [5] 白井清昭, 徳江英範. 対話型質問応答システムにおける問い返し文生成に関する基礎研究. 情報処理学会情報処理学会自然言語処理研究会, pp. 53-58, 2005. NL-165-8.