

複数文質問のタイプ同定

田村 晃裕[†] 高村 大也^{††} 奥村 学^{††}

既存の質問応答システムは、複数文で構成される質問には答えられない。そこで、我々はそのような複数文質問にも対応できる質問応答システムの構築を目指す。その第1段階として、複数文質問の質問タイプを同定する手法を提案する。具体的には、まず最初に、入力として与えられた複数文質問から質問タイプを決める際に最も重要な1文を抽出する。そして、その抽出された1文を用いて質問タイプを同定するという手法をとる。また、本論文では、質問タイプを同定する際に有効な情報となる名詞を特定するルールも提案する。複数文質問を含んだ実験データに対して、これらの情報と手法を用いて質問タイプを同定することで、F値が8.8%、正解率が4.4%改善できた。

Classification of Multiple-sentence Questions

AKIHIRO TAMURA,[†] HIROYA TAKAMURA^{††} and MANABU OKUMURA^{††}

Conventional QA systems cannot answer to the questions composed of two or more sentences. Therefore, we aim to construct a QA system that can answer such multiple-sentence questions. As the first stage, we propose a method for classifying multiple-sentence questions into question types. Specifically, we first extract the core sentence from a given question text. Then, we use the core sentence in question classification. We also propose a rule for extracting the effective noun in question classification. The result of experiments with the dataset including multiple-sentence questions shows that the proposed method improves F-measure by 8.8% and accuracy by 4.4%.

1. はじめに

今日、Webに代表されるように電子化されたテキストが大量に存在している。この大量の電子化されたテキストから、効率良く情報を取得する技術が重要になっている。そのような状況で、多くの情報検索システムが答えを含む文書を出力とするのに対して、質問応答システムは答えそのものを出力するので、より効率良く情報を取得できる有用なシステムといえる。

既存の質問応答システムの評価は、海外では評価型ワークショップ TREC における QA-TRACK¹、国内では QAC (Question & Answering Challenge)²で行われている。これらのワークショップでシステムの入力として与えられる質問は、1文で構成される質問(以降「1文質問」と呼ぶ)のみである。

一方 Web には、インターネットユーザ同士が質問しあう Q&A サイト (Yahoo!知恵袋³, 教えて!goo⁴

など)があり、そこには複数文で構成される質問(以降「複数文質問」と呼ぶ)が数多くある。たとえば、「餃子の具が少し余ってしまいました。そのまま焼くほかに何か調理法を教えてください。」などである。

このような複数文質問は、既存のシステムでは想定外の入力で、既存の手法をそのまま適用できない。もしくは、適用できたとしても精度が極端に落ちてしまう。そこで、我々は複数文質問にも答えられる質問応答システムの構築を目指す。

加藤³も示しているとおり、一般的な質問応答システムは「質問解析部」、「文書検索部」、「解答抽出部」の3つから成り立っている。まず質問解析部で質問を解析し、その後の処理で必要な情報を得る。そして、文書検索部で解答が含まれている可能性のある文書を検索し、その後、解答抽出部でその文書から解答を抽出するという処理の流れが一般的である。ここで、質問解析部において、解答を抽出する際の手がかりとなる質問タイプが決定される。たとえば、「日本の首相は誰ですか?」の質問を分析し、その解答が「人名」で

[†] 東京工業大学大学院総合理工学研究科

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

^{††} 東京工業大学精密工学研究所

Precision and Intelligence Laboratory, Tokyo Institute of Technology

¹ <http://trec.nist.gov/tracks.htm>

² <http://www.nlp.is.ritsumei.ac.jp/qac/>

³ <http://knowledge.yahoo.co.jp/>

⁴ <http://oshiete.goo.ne.jp/>

あると判定する。この処理を質問タイプ同定と呼ぶ。これにより解答抽出部では、文書検索部により出力された、解答が含まれる可能性のある文書中のすべての部分に対して、解答になるかどうか判定する必要はなくなり、判定箇所を絞り込むことができ効率の良い解答抽出が可能になる。また、質問タイプに応じて解答抽出手法を変えることもできる。たとえば、質問タイプが「理由」であることが分かれば、理由を探すのに特化した手法（たとえば森本ら⁶⁾などの手法）を用いることができ、結果として解答抽出精度を上げることができる。このように、解答抽出の精度や効率性は質問タイプ同定の精度に依存するため、質問タイプ同定はシステム全体に関わる重要な処理である。そこで、本研究では、複数文質問にも対応したシステム構築の第1歩として、複数文質問のタイプ同定手法を提案する。ここで、複数文質問には「デスクトップ上のアイコンやファイルが消えてしまいました。どうやったら元の状態にもどせますか？」のように1つの事柄を尋ねる質問と「サッカーのワールドカップについて調べているので教えてください。最初のワールドカップはどこで開催されたのですか？ また、優勝国はどこですか？」のように2つ以上の事柄を尋ねる質問があるが、本研究では特に前者の1つの事柄を尋ねる質問を扱う。たとえば、前述した「デスクトップ上の～」という複数文質問が入力として与えられたとき、その質問タイプとして“方法”と同定することを目指す。そのために、我々は、核文抽出という処理を導入する。これは、複数文質問から質問タイプを同定する際に最も重要な1文を抽出する処理である。複数文質問には質問タイプを同定する際に不必要な文がある。それらの文を核文抽出という処理で取り除くことで、ノイズとなる情報を除去し、質問タイプ同定の精度の向上を期待したことが導入の動機である。

提案手法は、まず核文抽出で質問タイプ同定の際に重要な1文を抽出し、その後、その1文の情報のみを用いて質問タイプを同定する。

まず、2章で関連研究を述べる。そして、3章で提案する質問タイプ同定手法を述べ、4章で提案手法の有効性を確認する実験とその結果の評価・考察を行う。最後に5章で、本研究のまとめと今後の課題について述べる。

2. 関連研究

本章では、いくつかの既存の質問タイプ同定手法を紹介する。手法は大きく2つに分けられる。1つはルールベースによる手法、もう1つは機械学習による手法

である。質問応答システム SAIQA⁸⁾ や、Xuら¹⁰⁾ は人手で作成したルールにより質問タイプを同定している。しかし、この手法はパターンやルールを作成するコストが大きい点や、ルールやパターンにマッチした限られた質問しか扱えない点が欠点としてあげられる。

これらの欠点を解消できるものが機械学習を用いた手法である。Liら⁴⁾ は SNoW を用いて質問タイプを同定している。SNoW とは、素性数が多い場合にもうまく学習できる複数クラスの分類器である。Zukermanら¹²⁾ は決定木学習によって、また、Ittycheriahら²⁾ 最大エントロピー法によって質問タイプを同定している。Suzuki⁹⁾ は Support Vector Machines (SVMs) を用いて質問タイプを同定している。また、Suzuki⁹⁾ はいくつかの機械学習手法（決定木学習、最大エントロピー法、SVM）とルールベースによる手法で質問タイプ同定の結果を比較している。これによると、SVM が最も結果が良い。Suzuki⁹⁾ は、質問タイプ同定の精度を上げるには、より多くの情報を考慮する必要があり、SVM はサンプル数に対して素性空間の次元数が高くなるときにも精度の高い分類が可能であるため適しているとしている。さらに、Zhangら¹¹⁾ も質問タイプを5つの機械学習手法で同定し、その結果を比較しており、Suzuki 同様、SVM が最も良かったと結論付けている。これらのことより、3.2節で述べるが、本研究でも質問タイプを同定する際に SVM を用いる。

ここで、本研究と従来研究の最大の違いは、従来研究は1文質問のみを対象としているのに対し、本研究は複数文質問も対象としている点であることを再確認しておく。また、本研究の実験データは、従来研究が実験データとして用いている QA タスクで用意された人工的な質問ではなく、インターネット上にある実際になされた様々な質問を用いている。そのため、本研究の結果と従来研究の結果は単純には比べられないことに注意してほしい。

3. 複数文質問のタイプ同定手法（2段階手法）

本章では、複数文質問のタイプを同定するための提案手法を説明する。まず最初に質問タイプ同定手法の流れを図1に示す。

まず、入力として質問が与えられたら、前処理として括弧部分を除き「」?」?」、空白などで文分割する。括弧部分を除く理由としては、構文解析を行う際に括弧部分が悪影響を及ぼす可能性があるためである。

前処理後は、質問が1文か複数文かで処理が分れる。1文質問の場合、質問がそのまま質問タイプ同定

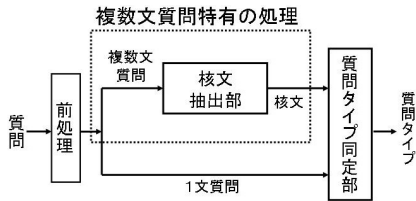


図1 質問タイプ同定の流れ

Fig. 1 The entire flow of question classification.

部の入力となり、タイプ同定される。複数文質問の場合、まず核文抽出部により処理される。核文抽出部ではタイプ同定の際に最も重要な1文(以降「核文」と呼ぶ)が決定される。その後、この核文が質問タイプ同定部の入力となり、タイプ同定される。

核文抽出部が複数文質問特有の処理であり、先行研究にはない部分である。以降、核文抽出部を3.1節で、質問タイプ同定部を3.2節で説明する。

3.1 核文抽出部

複数文質問が入力として与えられたときに、その質問から核文を抽出する部分である。たとえば「今、アメリカの歴史について調べています。独立記念日はいつでしたっけ?」という質問が入力されたときに「独立記念日はいつでしたっけ?」という1文を抽出する。

複数文質問の場合、質問を構成するすべての文を対象に分類器を適用してもノイズが多く、分類精度が落ちてしまう。上の例では「今、アメリカの歴史について調べています。」という文は質問タイプを同定する際に考慮しないでよい。それどころか逆に悪影響を及ぼす可能性がある。そのため「複数文質問の質問タイプは核文となる1文で同定できる」という仮定を導入し、質問の核を決め、ノイズを除去するという目的でこの処理を導入した。この仮定を導入するにあたり、本研究では、複数文質問の中でも1つの事柄を尋ねる質問のみを扱っている点に注意してほしい。

核文抽出の手順を説明する。

step1 1文ずつ核文であるか判定

1文ごとに分類器を用いて核文であるか判定する。ここで、分類器は分類結果とともに $S_i \in Question$ に対して $Score(S_i)$ という値を出力するものとする。 $Question$ は質問を構成する文の集合、 S_i は i 番目の文、 $Score(S_i)$ は S_i の核文としてのもっともらしさを表す。

step2 核文の決定

step1の結果を用いて、 $Score(S_i)$ が最大となる文を核文に決定する。

$$\text{核文} = \operatorname{argmax}_{S_i \in Question} Score(S_i). \quad (1)$$

本節の最後に詳しく述べるが、step1で核文かどうかを判定する際に用いた文の性質(素性)は、単語の情報である。ただし、素性を組み込む際は次の点を考慮する必要がある。ある文を核文かどうか判定する際は、判定する文の情報だけではその文が核文かどうか決められない、という点である。具体例で説明する。

質問1: よく効く花粉症の薬を教えてください。頭痛と鼻づまりを治したいです。よろしくおねがいします。

質問2: 頭痛と鼻づまりを治したいです。よろしくおねがいします。

上の質問1と質問2の2つの質問を考える。それぞれの核文は、下線がひかれた文である。

ここで、質問1と質問2に太字で書かれた「頭痛と鼻づまりを治したいです」という文がともに存在している。この太字の文は、質問1では核文としてふさわしくないが、質問2では核文になる。質問1では「よく効く花粉症の薬を教えてください」という文が前にあるため、太字の文が核文として適当でないのである。このことから、太字の文を核文かどうか判定する際、太字の文の情報だけでは核文かどうかを決めることができない。つまり、核文かどうかを判定する際には、判定する文の前にどのような文があるのか、後にどのような文があるのかを考慮し、相対的に決定しなければならない。この点をふまえ、素性に判定対象文の前後の文の情報を含める。

ここで、「窓サイズ」という概念を導入した。「窓サイズが n 」というのは、「素性に判定対象文のほかに前 n 文と後 n 文の情報を加える」という意味である。たとえば、窓サイズが0ならば、判定対象文のみの情報を使用することであり、窓サイズが ∞ ならば、質問を構成するすべての文の情報をを用いるということである。

本研究では、分類器としてSVMを使用した。したがって、 $Score(S_i)$ はSVMが出力する分離平面からの距離を使用した。また、素性を作る際の単語の情報は、出現した形態素の unigram と bigram を使用した。よって、素性ベクトルは窓内にある文の形態素の unigram, bigram で構成される。ただし、判定対象文、前の文、後ろの文の素性は別々に扱った。つまり、前の文に判定対象文と同じ形態素が含まれていても、別々の素性として扱っている。

3.2 質問タイプ同定部

1文質問や核文抽出処理により抽出された核文に対してタイプ同定を行う部分である。2章で述べたように、SVMを用いてタイプ同定する。素性は、形態素の

unigram, bigram, 名詞の意味カテゴリ, 注目名詞, 注目名詞の意味カテゴリの5種類を用いる. 意味カテゴリはシソーラスより取得する. たとえば「店」「駅」「市」などである. 注目名詞とは, Moldovanら⁵⁾の質問対象語(質問文中で解答が何であるかを限定している語)にあたるものである. たとえば「今までの日本の最高気温は何度ですか?」という質問では「最高気温」が質問対象語になる. 従来研究では, 質問対象語の特定に人手で記述したパターンを用いている. しかし, 本研究では「~を探しています」や「~がわかりません」のように疑問形をしていない質問などを含め, あらゆる形式の質問を扱うため, 人手でパターンを作成するのは困難である. そこで本研究では, 以下の手順でタイプ同定の際に重要な名詞を特定し, その名詞を“注目名詞”として質問対象語の代わりに用いた.

step1 最後の動詞を含む文節・「?」で終る文節をみつける.

step2 step1 でみつけた文節にかかる文節をみつける.

step3 step2 でみつけた文節内の名詞・未知語を注目名詞とする.

4. 評価実験・結果と考察

本章では, 3章で提案した手法の有効性を確認する実験を行う.

学習には SVM 学習パッケージ TinySVM¹⁾を使用した. また, 形態素は日本語形態素解析器 ChaSen²⁾を, 係り受け関係は日本語係り受け解析器 cabocha³⁾を使用して解析した. 意味カテゴリを取得するシソーラスは日本語語彙大系の意味体系¹⁾を利用した.

4.1 実験設定

本研究の目的より, 実験データとなる質問は, 多くの複数文質問や様々な形式の質問を含んでいる Q&A サイトから収集した. Q&A サイトとは, 質問をのせておくと他のユーザが答えてくれるサイトである. Q&A サイトから, はてな⁴⁾と Yahoo!知恵袋⁵⁾の2つを選び, 各サイトから 2,000 個ずつ, 合計 4,000 個の質問を抽出した. この 4,000 個の質問から今回の実験では用いない3つの種類の質問を人手で取り除いた. 除い

表 1 質問タイプと質問の分布

Table 1 The question types and the distribution of 2,376 questions.

解答が名詞		解答がテキスト	
質問タイプ	質問数	質問タイプ	質問数
人名	64	理由	132
製品名	238	方法	500
施設名	139	定義	73
地名	393	叙述	228
時間表現	108	意見	173
数値表現	53	その他(テキスト)	131
その他(名詞)	144		
	1,139		1,237
合計 2,376			

た質問は「答えが Yes か No のみとなる質問」「2つ以上の事柄を尋ねる質問」「質問ではないもの」である. これらを除いた後に残った 2,376 個の質問を実験データとする. 実験前に, あらかじめこれらの質問には人手で質問タイプと核文をタグ付けしておいた. タグ付けは1人で行った. 核文をタグ付けする際の基準は「質問を構成する文の中で, 答えを要求する最も重要だと思われる文を核文とする」とした. 表 1⁶⁾に本研究で定めた質問タイプの種類とデータの分布を示す. 解答が名詞になるものについては, 佐々木らの研究⁷⁾を参考にした. 本研究の主張である核文に関して一致度を調べるため, データの一部について2人の被験者により核文を特定してもらい, その一致率を示す Kappa 係数を算出したところ, 0.97 となり高い一致度を示した.

2,376 個の質問のうち, 818 個が1文質問で 1,558 個が複数文質問であった. また, 複数文質問の平均文数は 3.49 文であった. よって核文抽出というタスクは, 複数文質問 1,558 個に対して, 平均 3.49 文の中から核文を1つ決定することになる.

核文抽出実験の評価指標は, すべての複数文質問のうち, システムが出力した核文がどの程度正しかったかを表す正解率とする. 質問タイプ同定実験の評価指標は, 質問タイプごとの F 値と, すべての質問の中で最終的に定まった質問タイプがどの程度正しかったかを表す正解率とする. F 値は, $2 \times \text{再現率} \times \text{精度} / (\text{再現率} + \text{精度})$ で表される. また, すべての実験は2分割交差検定により評価する.

4.2 核文抽出部

核文抽出に対して 3.1 節で説明した素性の有効性を検証するため, 窓サイズと素性を変えて核文抽出を行った. 結果を表 2 に示す. 表 2 より, 90%以上の高

¹⁾ <http://chasen.org/~taku/software/TinySVM/>

²⁾ <http://chasen.naist.jp/hiki/ChaSen/>

³⁾ <http://chasen.org/~taku/software/cabocha/>

⁴⁾ <http://www.hatena.ne.jp/>

⁵⁾ <http://knowledge.yahoo.co.jp/>

⁶⁾ 佐々木ら⁷⁾の質問タイプには「組織名」が含まれていたが, 本研究では「組織名」に属する質問の数が少なかったため, 「その他」に統合した.

表 2 異なる窓サイズと素性による核文抽出正解率

Table 2 Accuracy of core sentence extraction with different window sizes and features.

窓サイズ \ 素性	Unigram	Bigram	Unigram+Bigram
0	1,350/1,558= 0.866	1,378/1,558= 0.884	1,385/1,558= 0.889
1	1,357/1,558= 0.871	1,386/1,558= 0.890	1,396/1,558= 0.896
2	1,364/1,558= 0.875	1,397/1,558= 0.897	1,405/1,558= 0.902
∞	1,376/1,558= 0.883	1,407/1,558= 0.903	1,416/1,558= 0.909

表 3 核文抽出時に影響力のある素性

Table 3 Influential features for core sentence extraction.

順位	正の重み	負の重み
1	を教え (bigram, 判)	?-記号 (unigram, 前)
2	?-記号 (unigram, 判)	を-助詞 (unigram, 後)
3	サイト-名詞 (unigram, 判)	か? (bigram, 前)
4	方法-名詞 (unigram, 判)	サイト-名詞 (unigram, 後)
5	か (bigram, 判)	か-助詞 (unigram, 前)
6	ですか (bigram, 判)	方法-名詞 (unigram, 後)
7	を探し (bigram, 判)	を教え (bigram, 後)
8	って (bigram, 判)	を教え (bigram, 前)
9	知る-動詞 (unigram, 判)	.-句点 (unigram, 判)
10	か-助詞 (unigram, 判)	探して (bigram, 前)

判: 判定対象文の素性, 前(後): 出現位置が判定対象文より前(後)の文の素性

い正解率で核文を抽出できたことが分かる。この結果は、核文抽出処理結果をこの後に続く質問タイプ同定の入力として十分使えることを示していると考えられる。また表 2 より、窓サイズを大きくすればするほど結果が良くなっていることが分かる。これは、核文抽出において前後の文の情報を考慮することが重要であることを示しているとともに、核文抽出に有効な手がかりは質問内に散在していることを示している。素性に関しては、すべての窓サイズにおいて unigram と bigram の両方を用いたものが最も有効であった。

ここで、最も結果の良いモデル(窓サイズ ∞ , 素性 unigram と bigram)を用いて、核文抽出に影響した素性や誤りの分析を行う。SVM で学習した結果、各素性に付与された重みの絶対値が大きい素性を 20 個(符号が正の素性を 10 個、負の素性を 10 個)表 3 に示す。重みが正のものは判定対象文を核文にするように働き、重みが負のものは核文にしないように働く素性である。この表の「正の重み」の列には「~か?」、「~を教えてください」、「~を探してください」などの、人間が質問するときを使う典型的な表現に含まれていそうな unigram, bigram が多い。このような unigram, bigram があると核文になりやすくなっている。また「負の重み」の列より、判定対象文以外の文に、人間が質問するときを使う典型的な表現に含まれていそうな unigram, bigram があると、判定対象文が核文になりにくくなっていることが分かる。このことから、核文は、判定対象文の核文としてのふさわしさと判定対象文以外の文が核文になりうるかを考慮

して選ばれていることが分かり、このような考え方で核文を抽出することで 90%以上の高い精度を実現できている。

核文抽出で誤った質問の考察をする。前述したように、核文は表 3 の「正の重み」の列にあるような典型的な質問表現を含む文が選ばれやすい。そのため、正しい核文が SVM により学習された典型的な質問表現を含まない質問に関しては核文抽出に失敗していた。たとえば「魚を左側にして盛り付ける理由があればお願いします。見栄えの問題だと思っているのですが、教えてくださいませんか?」のような質問である。この質問では、最初の文が核文なのだが、「か?」を含む 2 文目が核文として選ばれてしまっている。ここで「お願いします」も核文の表現としてふさわしいとも考えられるが、この表現は、挨拶としての「よろしく申し上げます」や条件を指定する際の文「100 円以下でお願いします」など、核文以外の文でも多くみられる表現であり、核文の表現としてうまく学習できなかったと考えられる。また、正しい核文に表 3 にあるような表現が含まれていても、同質問内に表 3 のような表現を含んだ核文以外の文がある場合には、間違える確率も上がっていると考えられる。このような問題は、今回素性として用いた単語の表層的な情報だけでは解決しにくく、文の位置情報などの単語以外の情報を考慮する必要があると考えられる。この問題については、今後の課題とする。

次に、3 つの簡単なルールをベースライン手法とし、核文抽出を行う。その結果と 3.1 節で提案した SVM を用いた手法による結果を比較することで、提案手法の有効性を確認する。ベースライン手法の 1 つ目は、最初の文を核文とする手法。2 つ目は、最後の文を核文とする手法。3 つ目は、最後の疑問文(疑問文がない場合は最初の文)を核文とする手法である。結果を表 4 に示す。表 4 より、ベースライン手法のような単純な手法では核文抽出はうまくいかないことが分かり、提案手法の有効性が確認できる。

4.3 質問タイプ同定: 核文抽出の有効性

質問タイプを同定する際に核文抽出部がどの程度有効に働くかを調べる。質問タイプ同定部の入力として

表 4 3つのベースライン手法による核文抽出結果

Table 4 Accuracy of core sentence extraction with baseline methods.

手法	正解率
最初の文	743/1,558 = 0.477
最後の文	471/1,558 = 0.302
最後の疑問文	1,077/1,558 = 0.691

表 5 質問タイプ同定部において入力異なる3つのモデルでのタイプ同定結果

Table 5 F-measure and accuracy of the three models for question classification.

タイプ同定の入力	質問全体	核文抽出部の結果	正しい核文
核文抽出精度	-	0.909	1.000
人名	0.462	0.434	0.505
製品名	0.381	0.467	0.480
施設名	0.584	0.569	0.586
地名	0.758	0.780	0.824
時間表現	0.340	0.508	0.524
数値表現	0.262	0.442	0.421
その他(名詞)	0.049	0.144	0.145
理由	0.280	0.539	0.579
方法	0.756	0.778	0.798
定義	0.643	0.624	0.656
叙述	0.296	0.315	0.317
意見	0.591	0.675	0.659
その他(テキスト)	0.090	0.179	0.186
平均 F 値	0.423	0.496	0.514
正解率	0.617	0.621	0.652

以下の3つのモデルを考え、タイプ同定実験を行い、比較する。

質問全体 核文抽出処理をせずに、質問全体をそのまま質問タイプ同定部の入力とする。

核文抽出部の結果 3.1節で提案した手法で抽出した核文を質問タイプ同定部の入力とする。核文抽出処理の精度は90.9%である(表2参照)。

正しい核文 人手でタグ付けしておいた正しい核文を質問タイプ同定部の入力とする。これは、核文抽出処理の精度が100%であるときに等しい。

素性は、形態素の unigram, bigram, 名詞の意味カテゴリを使用した。注目名詞については、特定する際に入力が1文であることを前提としているため、“質問全体”モデルでは特定できない。そのため、条件を揃えて厳密な比較をするために、3つのモデルすべてに注目名詞に関する素性を使用しなかった。結果を表5に示す。表5より、核文抽出を導入することでほとんどの質問タイプにおいてF値が改善され、平均F値や正解率も向上していることが分かる。このことから、質問タイプ同定において核文抽出という処理が有効であるといえる。また、“正しい核文”モデルが最も結果が良くなったことから、核文抽出精度をさらに改善することで、質問タイプ同定の精度の向上が期

待できる。

さらに詳しく、核文抽出が質問タイプ同定に与える影響を考察する。核文抽出処理で間違った文が抽出された質問に対して、最終的に決定された質問タイプがどの程度正解していたかを調べた。結果は、核文抽出処理で間違った質問142個中54個が質問タイプを正しく同定できた。つまり、正解率は38%であり、表5の数値と比べてもかなり低い。したがって、核文抽出の時点で間違えてしまうと正しい質問タイプを同定することは難しいといえる。

4.4 質問タイプ同定部

4.4.1項で、質問タイプ同定時に有効な素性を調べる。また、4.4.2項で核文以外の文の情報の必要性を検討する。これら2つの実験には、質問タイプ同定部の入力に人手でタグ付けした正しい核文を用いる。最後に4.4.3項で、2段階(核文抽出、質問タイプ同定)に提案手法を用いて質問タイプを同定することで、提案手法の有効性を結論づける。

4.4.1 有効な素性

質問タイプ同定時に有効な素性を調べるため、3.2節で提案した5種類の素性(形態素の unigram, bigram, 名詞の意味カテゴリ, 注目名詞, 注目名詞の意味カテゴリ)から、1種類ずつ素性を抜いてタイプ同定実験を行った。F値の変化によって、それぞれの素性の有効性を確認する。結果を表6に示す。括弧内の値は、提案素性をすべて使用したときとの差である。この値が負のものが有効な素性、正のものが有効でない素性である。表6より、質問タイプによって有効な素性やF値が異なることが分かる。そこで、これらの違いを“製品名”と“方法”を例にとり考察する。まずF値の違いについて述べる。これは、質問タイプを特徴付ける表現があるかどうかによると考えられる。“方法”の質問をみると、「どうすれば〜?」「どうしたら〜?」などの“方法”特有の質問形式があった。また、「方法」「やり方」といった質問タイプ“方法”を特徴付ける名詞もあった。このためタイプ同定が比較的容易で、F値が高かったと考えられる。一方、“製品名”の質問を見ると、“製品名”特有の質問形式がなかった。“製品名”の質問形式として多かったものは、「〜を教えてください」や「〜は何ですか?」であった。しかし、これらは“地名”や“施設名”といった“製品名”以外の質問タイプでも多く使われているため、“製品名”特有のものではない。“製品名”を分類するときには、質問形式ではなく名詞が重要な働きをしていると考えられる。たとえば「花粉症の薬を教えてください」と「何故電子レンジで金属を温めては

表 6 素性変化時のタイプ同定結果

Table 6 Experiments with each feature set being excluded.

	すべて	除いた素性				
		Unigram	Bigram	意味(名)	注目名詞	意味(注)
人名	0.574	0.571 (-0.003)	0.620 (+0.046)	0.536 (-0.038)	0.505 (-0.069)	0.505 (-0.069)
製品名	0.506	0.489 (-0.017)	0.579 (+0.073)	0.483 (-0.023)	0.512 (+0.006)	0.502 (-0.004)
施設名	0.612	0.599 (-0.013)	0.642 (+0.03)	0.549 (-0.063)	0.615 (+0.003)	0.576 (-0.036)
地名	0.832	0.826 (-0.006)	0.841 (+0.009)	0.844 (+0.012)	0.825 (-0.007)	0.833 (+0.001)
時間表現	0.475	0.506 (+0.031)	0.548 (+0.073)	0.420 (-0.055)	0.502 (+0.027)	0.517 (+0.042)
数値表現	0.442	0.362 (-0.080)	0.475 (+0.033)	0.440 (-0.002)	0.466 (+0.024)	0.413 (-0.029)
その他(名詞)	0.210	0.182 (-0.028)	0.267 (+0.057)	0.204 (-0.006)	0.198 (-0.012)	0.156 (-0.054)
理由	0.564	0.349 (-0.215)	0.622 (+0.058)	0.603 (+0.039)	0.576 (+0.012)	0.582 (+0.018)
方法	0.817	0.803 (-0.014)	0.787 (-0.030)	0.820 (+0.003)	0.817 (±0.000)	0.807 (-0.010)
定義	0.652	0.659 (+0.007)	0.603 (-0.049)	0.640 (-0.012)	0.647 (-0.005)	0.633 (-0.019)
叙述	0.355	0.308 (-0.047)	0.355 (±0.000)	0.363 (+0.008)	0.357 (+0.002)	0.334 (-0.021)
意見	0.696	0.670 (-0.026)	0.650 (-0.046)	0.703 (+0.007)	0.676 (-0.020)	0.685 (-0.011)
その他(テキスト)	0.183	0.176 (-0.007)	0.179 (-0.004)	0.154 (-0.029)	0.190 (+0.007)	0.198 (+0.015)
平均 F 値	0.532	0.500 (-0.032)	0.551 (+0.019)	0.520 (-0.012)	0.530 (-0.002)	0.518 (-0.014)
正解率	0.674	0.632	0.638	0.668	0.661	0.661

意味(名): 名詞の意味カテゴリ, 意味(注): 注目名詞の意味カテゴリ

いけないのか理由を教えてください。」という質問を考える。両方とも「～を教えてください。」という同じ質問形式であるが、前者の質問タイプは“製品名”、後者の質問タイプは“理由”であり質問タイプは異なる。前者の質問タイプは、「薬」という名詞で質問タイプが“製品名”と同定されると考えられる。しかし、“製品名”を特徴付ける名詞は多様であり、その名詞を学習することは難しい。そのため、タイプ同定が難しくなり、F 値も低くなったと考えられる。

次に有効な素性の違いについて述べる。最も有効な素性は、“製品名”に関しては名詞の意味カテゴリ、“方法”に関しては bigram である。前述したとおり、“製品名”は名詞がタイプ同定の際に重要な働きをしているため、名詞の情報である意味カテゴリが最も有効であったと考えられる。一方、“方法”については、前述したように「どのように～？」などの質問形式が重要な働きをしている。そのため、形態素のつながりが重要で、形態素の隣接関係の情報である bigram が有効であったと考えられる。

最後に、本研究で提案した注目名詞の有効性について考察する。提案素性には注目名詞に関して、注目名詞自体とその意味カテゴリの両方を入れたため、片方

を素性からはずしても、もう片方が情報を補完してしまう可能性がある。そのため、表 6 からは注目名詞の有効性は分かりにくい。そこで、表 5 に注目名詞の 2 つの情報を抜いた結果があるのでその部分と比較すると、平均 F 値は 0.514 から 0.532 に、正解率は 0.652 から 0.674 に上昇し、注目名詞の有効性が確認できる。ここで、表 5 は 3 種類の素性(形態素の unigram, bigram, 名詞の意味カテゴリ)を使用していることを確認しておく。

4.4.2 核文以外の文の情報

核文以外の文の情報の必要性について考察する。ここで 3.1 節同様、窓サイズを導入する。「窓サイズが n 」というのは、「質問タイプ同定時の素性ベクトルに核文のほかに前 n 文と後 n 文の情報を加える」という意味である。たとえば、窓サイズが 0 ならば、核文のみの情報を用いることであり、窓サイズが ∞ ならば、質問を構成するすべての文の情報を用いるということである。この窓サイズを 0, 1, 2, ∞ に変えたときのタイプ同定実験結果を比較することで、核文以外の情報の必要性を検討する。この実験の素性は、核文については提案した 5 種類を使用し、前後の文については形態素の unigram, bigram, 名詞の意味カテゴリ

表 7 窓サイズ変化時のタイプ同定結果

Table 7 Experiments with different window sizes.

	窓サイズ			
	0	1	2	∞
人名	0.574	0.558	0.565	0.570
製品名	0.506	0.449	0.441	0.419
施設名	0.612	0.607	0.596	0.578
地名	0.832	0.827	0.817	0.815
時間表現	0.475	0.312	0.288	0.302
数値表現	0.442	0.322	0.296	0.311
その他(名詞)	0.210	0.123	0.120	0.050
理由	0.564	0.486	0.472	0.439
方法	0.817	0.808	0.809	0.792
定義	0.652	0.658	0.658	0.641
叙述	0.355	0.358	0.357	0.340
意見	0.696	0.670	0.658	0.635
その他(テキスト)	0.183	0.140	0.129	0.133
平均 F 値	0.532	0.486	0.477	0.463
正解率	0.674	0.656	0.658	0.653

りの3種類を使用した。結果を表7に示す。表7より、核文のみ(窓サイズが0)のときが最も結果が良い。この結果より、核文以外の文の情報は質問タイプを同定する際に有効な情報を含んでおらず、むしろ、ノイズとなっていると考えられる。したがって、3.1節で導入した仮定「複数文質問の質問タイプは核文となる1文で同定できる」が正しかったといえる。

4.4.3 提案手法の有効性のまとめ

これまでの実験で、核文抽出と注目名詞がタイプ同定に有効であることを示した。しかし、これまでは実験ごとに入力や素性などの設定が異なっていたので、本項では提案手法を用いてタイプ同定することで最終的にどの程度結果が向上したのかを示す。結果を表8に示す。ここで“従来手法”とは、複数文質問を1文質問と見なし、従来のように1文質問に対して行っていたSVMによる手法ですべての質問をタイプ同定した結果である。この部分は、表5の“質問全体”に等しい。また、“提案手法”の部分は素性を5種類用いている点で、表5の“核文抽出部の結果”とは異なっていることに注意してほしい。表5の“核文抽出部の結果”の素性は3種類である。

表8より、提案手法を用いることで、“施設名”、“定義”を除いたすべての質問タイプでF値が向上している。“施設名”、“定義”が悪くなった原因は、表6の素性をすべて入れたときの値(核文抽出100%、素性が5種類)と表8の値を比較すると分かる。核文抽出が100%のときは、2つの質問タイプともに提案手法を用いることで結果が向上している。つまり、この2つの質問タイプは核文抽出精度が低かったために提案手法の有効性が確認できなかったと考えられる。

表8より、結論として、提案手法により平均F値が8.8%、正解率が4.4%向上した。この結果が有意な

表 8 提案手法の有効性をまとめる実験結果

Table 8 The result of concluding experiments.

核文抽出	従来手法	提案手法
	なし	あり
素性	Unigram, Bigram 意味(名)	Unigram, Bigram 意味(名), 意味(注) 注目名詞
人名	0.462	0.492
製品名	0.381	0.504
施設名	0.584	0.575
地名	0.758	0.792
時間表現	0.340	0.495
数値表現	0.262	0.456
その他(名詞)	0.049	0.189
理由	0.280	0.537
方法	0.756	0.789
定義	0.643	0.626
叙述	0.296	0.321
意見	0.591	0.677
その他(テキスト)	0.090	0.189
平均 F 値	0.423	0.511
正解率	0.617	0.661

差であるかどうかを符号検定を用いて検定すると、有意水準1%で有意差が認められた。

最後に、提案手法とは別に複数文質問をタイプ同定する際に考えられる手法について述べる。核文抽出を行わない1段階手法であり、以下のような手順で質問タイプを決めるものである。

step1 文分割

step2 1文ずつ質問タイプを同定

1文ごとにSVMを用いて質問タイプを決める。

step3 複数文質問のタイプを同定

step2を行った際にSVMが出力する分離平面からの距離を、その質問タイプとしてのふさわしさととらえ、最大となる質問タイプを最終的な質問タイプとする。

この1段階手法によりタイプ同定を行うと、その正解率は61.7%(2,376個中1,467個が正解)であった。提案手法の正解率は、66.1%(2,376個中1,570個が正解)であるので、本研究で提案した手法の方がより効果的であるといえる。

5. まとめ

本研究では、複数文質問のタイプ同定手法を提案した。まず最初に、入力として与えられた複数文質問から核文を抽出し、その核文の情報を用いて質問タイプを同定する手法である。また本研究では、複数文質問の質問タイプを同定する際に有効な情報となる注目名詞の特定方法を提案した。実験を通じて、核文抽出と注目名詞が質問タイプ同定に有効であることを示した。提案した注目名詞は、1文から抽出することを仮定し

ているため、核文抽出をしないと使用できない。この意味でも、核文抽出は有効であると考えられる。結論として、提案手法を用いることで正解率が 8.8%、質問タイプの平均 F 値が 4.4% 向上した。

核文抽出に関しては、判定文の情報の他に前後の文の情報を考慮することが重要であることが分かった。そうすることで、90.9%という高い正解率を得ることができた。また、前後の情報を考慮する際には窓サイズを大きくするほど核文抽出結果が良くなることが分かった。

今後の課題は、本研究では 1 つの質問に対して核文は 1 つであると限定したが、これを変えることで「2 つ以上の事柄をきく質問」にも対応した質問タイプ同定手法に拡張したい。また、複数文質問に対応した質問応答システムを構築するために、タイプ同定以外の処理についても今後考える必要がある。

参 考 文 献

- 1) 池原 悟, 宮崎正弘, 白井 諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林 良彦: 日本語語彙大系 (1 意味体系), 岩波書店 (1997).
- 2) Ittycheriah, A., Franz, M., Zhu, W.J. and Ratnaparkhi, A.: Question Answering Using Maximum Entropy Components, *Proc. NAACL-2001*, pp.33-39 (2001).
- 3) 加藤恒昭: 自然言語による情報検索と類似性, 日本ファジィ学会誌, Vol.13, No.5, pp.436-444 (2001).
- 4) Li, X. and Roth, D.: Learning Question Classifiers, *Proc. COLING-2002*, Taipei, Taiwan, pp.556-562 (2002).
- 5) Moldovan, D., Harabagiu, S., Pasca, M., Mihalcea, R., Goodrum, R., Girju, R. and Rus, V.: Lasso: A Tool for Surfing the Answer Net, *Proc. TREC-8*, pp.175-184 (1999).
- 6) 森本格行, 福本淳一: Why 型質問に対する解答抽出, 情報処理学会論文誌, Vol.42, No.7, pp.293-296 (2004).
- 7) 佐々木裕, 磯崎秀樹, 鈴木 潤, 国領弘治, 平尾 努, 賀沢秀人, 前田英作: SVM を用いた学習型質問応答システム SAIQA-II, 情報処理学会論文誌, Vol.45, No.2, pp.635-646 (2004).
- 8) 佐々木裕, 磯崎秀樹, 平 博順, 平尾 努, 賀沢秀人, 鈴木 潤, 国領弘治, 前田英作: SAIQA: 大量文書に基づく質問応答システム, 情報処理学会情報学基礎研究会 FI-64, pp.77-82 (2001).
- 9) Suzuki, J.: Kernels for Structured Data in Natural Language Processing, Ph.D. Thesis, Nara Institute of Science and Technology (2005).
- 10) Xu, J., Licuanan, A. and M.Weischedel, R.:

TREC 2003 QA at BBN: Answering Definitional Questions, *Proc. TREC 2003*, pp. 98-106 (2003).

- 11) Zhang, D. and Lee, W.S.: Question Classification using Support Vector Machines, *Proc. SIGIR-2003*, Toronto, Canada, pp.26-32 (2003).
- 12) Zukerman, I. and Horvitz, E.: Using Machine Learning Techniques to Interpret WH-questions, *Proc. ACL-2001*, Toulouse, France, pp.547-554 (2001).

(平成 17 年 9 月 22 日受付)

(平成 18 年 3 月 2 日採録)



田村 晃裕

1981 年生。現在、東京工業大学大学院総合理工学研究科知能システム科学専攻修士課程在学。自然言語処理に関する研究に従事。



高村 大也 (正会員)

1974 年生。1997 年東京大学工学部計数工学科卒業。2000 年同大学院工学系研究科計数工学専攻修了 (1999 年はオーストリアウィーン工科大学にて研究)。2003 年奈良先端科学技術大学院大学情報科学研究科博士課程修了。博士 (工学)。2003 年より東京工業大学精密工学研究所助手。自然言語処理, 特に学習理論等の応用に興味を持つ。ACL 会員。



奥村 学 (正会員)

1962 年生。1989 年東京工業大学大学院情報理工学研究科計算工学専攻博士後期課程修了。1989 年より東京工業大学大学院情報理工学研究科助手。1992~2000 年北陸先端科学技術大学院大学助教授。1997~1998 年トロント大学客員助教授。2000 年より東京工業大学精密工学研究所助教授。自然言語処理, 自動テキスト要約, コンピュータによる語学学習支援, テキストデータマイニングに関する研究に従事。工学博士。AAAI, ACL, JSAI, JCSS 各会員。