

単語属性 N -gram と統計的機械学習による質問タイプ同定

鈴木 潤[†] 佐々木 裕[†] 前田 英作[†]

質問タイプ同定とは、自然文で与えられた質問文の意図を同定する技術である。この技術は、ユーザからの要求意図を正しく理解するために必須であり、質問応答、対話、情報検索/抽出等関連する様々なアプリケーションに広く適用可能である。精度の良い質問タイプ同定を実現するためには、質問文中の多種多様な言語表現から質問タイプ同定に有効な特徴を抽出することと、抽出された様々な特徴から最適な質問タイプ同定器を構築することが必要となる。本論文では、質問タイプ同定の特徴を単語属性 N -gram により抽出する手法を提案し、この単語属性 N -gram と機械学習手法 Support Vector Machine (SVM) とを組み合わせた質問タイプ同定手法を提案する。質問文 10,000 問を用いた評価実験により既存手法との比較を行い、提案手法の有効性を実証した。さらに、単語属性 N -gram により抽出された素性を解析した結果についても報告する。

Question Type Classification Using Word Attribute N -gram and Statistical Machine Learning

JUN SUZUKI,[†] YUTAKA SASAKI[†] and EISAKU MAEDA[†]

Question type classification attempts to identify the intention of a given question. The approach to high performance question classification typically yields an extremely large number of features because question types are well featured by the structures of the word attributes inside the questions. We propose a technique for finding “word attribute N -gram” to extract effective features for question type classification, and then, we use these features with machine learning technique, Support Vector Machines (SVM), to create a high performance question type classifier. Results of question type classification experiments using 10,000 question samples showed much higher performance than the other conventional methods. Moreover, we clarify the performance of a feature extraction method and the effective features of each question type.

1. はじめに

本論文で扱う「質問タイプ同定」とは、自然文で与えられた質問文に対し、その質問の意図を正しく判定する技術を指す。質問の意図を正しく理解することは、ユーザからの要求意図を正しく遂行するうえで非常に重要なことである。たとえば、人間の場合でも、話し相手から投げかけられた質問に対し、「相手が何を訊いているのか」を正しく理解できなければ、どのように答えてよいのか分からない。これとまったく同様に、計算機でも、与えられた質問の意図を正しく解析できなければ、それ以降の処理を正しく行うことがきわめて困難になる。

この「質問タイプ同定」は、質問応答システムの研

究開発の中で考案された技術の 1 つである。本論文でも質問応答における質問タイプ同定を扱うが、質問タイプ同定技術は、質問応答システムに限らず、情報検索、情報抽出、対話システム等、ユーザからの要求を自然文による質問の形で受け付けるあらゆるシステムに広く適用可能な技術でもある。ここで、「質問タイプ」とは、あらかじめ定義されている質問文の種類であり、用意しておくべき質問タイプの種類は、それを実装するアプリケーションの目的、求められる性能に依存する。質問応答における「質問タイプ」は、人名 (PERSON)、地名 (LOCATION) といった質問の対象が属するカテゴリ名を指す。

質問応答は、自然文で与えられた質問に対し、大量文書から適切な解答を導き出す技術であり、情報検索を中心としたテキスト処理の評価型国際会議 TREC (Text REtrieval Conference) において、1999 年より質問

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Research Laboratories,
NTT Corporation

応答技術の評価を目的とした QA-Track (Question Answering Track) が新設される等, 新しい情報アクセス技術として世界的に注目を集めている. 日本語を対象とした質問応答技術の研究も関心が高まりつつあり, 第 3 回 NTCIR ワークショップの一部として, QAC (Question and Answering Challenge) が開催された. TREC, QAC では, 大量文書を対象とした対象分野を限定しない質問応答, いわゆるオープンメイン QA が基本的な枠組みであり, それらでは, 質問タイプ同定がきわめて重要な課題として扱われている.

TREC QA-Track 参加システムを含む英語質問応答システムでは, ルールベース, または, パターンマッチングにより質問タイプ同定を行っているものがほとんどである^{1)~3)}. 日本語質問応答システムでも同様に, 質問文の疑問詞等に注目した簡単な意味的制約や人手により経験的に作成したルールにより同定している^{4)~7)}. しかし, これらの方法では, 適用できる質問文の表現が限定される, あるいは, ルールや制約の作成, パラメータの最適化等に要するコストが大きいといった問題がある.

これらの問題を解決する 1 つの手段として, 質問タイプ同定器を学習サンプルから機械的に作成する統計的機械学習手法が有効であると考えられる. すでに, このような統計的機械学習手法を導入した質問タイプ同定法として, 決定木学習⁸⁾, 最大エントロピー法⁹⁾を用いた質問タイプ同定手法が提案されている. しかし, 質問タイプ同定問題は, 質問の意図を解析するという観点から考えても難しい課題であり, 多様な言語表現に対応し, かつ, 高精度な質問タイプ同定を実現するには, 多くの情報を考慮しなければならない. 特に, 質問タイプを特徴付けるのに有効な素性は, 質問文中に出現する単語の属性間の関係から得られるため, 考慮すべき素性数は膨大な数になり, 前述した統計的機械学習手法では, サンプル数に対して素性空間の次元数が高くなるため高精度な分類器を構成することが難しいという問題が発生する.

Support Vector Machine (SVM)^{10), 11)} は, サンプル数に対して素性空間の次元が高い場合でも高精度な分類が可能な学習器として知られおり, 質問タイプ同定に適した手法であると考えられる. そこで本論文では, 質問文中に出現する単語属性間の関係を抽出する単語属性 N -gram を提案し, そこで得られた素性集合を用いて SVM により質問タイプ同定を行う新しい

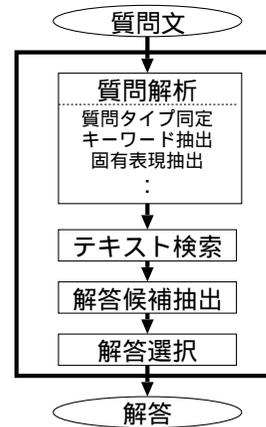


図 1 質問応答システムの一般的な処理の流れ

Fig. 1 The typical flow of a question answering system.

質問タイプ同定手法を提案する.

以下, 2 章では質問応答における質問タイプ同定問題についての詳細を述べる. 3 章では, 質問タイプ同定に適した単語属性 N -gram による素性抽出手法について説明する. 4 章で統計的機械学習手法の 1 つである Support Vector Machine (SVM) を用いた質問タイプ同定について述べる. 5 章で本論文で行った質問タイプ同定の実験設定, 6 章で実験の結果およびその考察を述べ, 7 章で今後の課題について述べる. 最後に 8 章で結論を述べる.

2. 質問応答における質問タイプ同定問題

2.1 質問タイプ同定問題の概要

質問応答システムの一般的な処理の流れを図 1 に示す.

質問応答では, 一般的に質問文の解析を最初に行う. 人間同様に, 計算機による質問応答でも, この最初に行う質問解析を誤ると, 後の解析すべてに悪影響を及ぼすことになる. よって, 与えられた質問が何を訊いているのかを判断する「質問タイプ同定」は, 質問応答の性能を向上させるうえできわめて重要な要素の 1 つである.

しかし, 質問タイプ同定は難しい問題を含んでいる. 質問文「日本の首相は誰ですか?」の場合は人名を聞いているので, 質問タイプは人名 (PERSON) となる. この場合は「～は誰ですか?」の表現から質問タイプを人名 (PERSON) と容易に同定することができる. しかし, たとえば, 「～はどこですか?」という形の質問文であっても, 「旅行会社で, 一番人気があるのはどこですか?」と「旅行先で, 一番人気があるのはどこですか?」の質問タイプは, それぞれ組織名

(ORGANIZATION) と地名 (LOCATION) となり、簡単にパターン化はできない。

あるいは、「旅行先で、一番人気があるのはどこですか?」という質問に対して、「旅行先で、一番人気があるのは?」や、「人気が一番ある旅行先はどこ?」といったように、省略や言い換えにより、様々な表現方法でまったく同じ質問をすることが可能である。

このように、質問文に現れる多種多様な言語表現をカバーするルールやパターンを手で作成するのは大変コストのかかる作業であり、非常に困難であることは、容易に推察できる。

また、これまでに述べたように、本論文では、入力される質問文の表現を限定しない、より自然な質問タイプ同定を対象にして議論を進める。

2.2 質問タイプ

質問タイプの種類に関しては、TREC QA-Track に参加している英語質問応答システムにおいて、様々な質問タイプセットが提案されている。たとえば、MUC (Message Understanding Conference) で定義された固有表現のカテゴリを利用して定義したもの^{9),12)}、意味ネットワークである WordNet¹³⁾ を利用したもの^{1),2),14)}、独自に質問タイプを定義しているもの^{3),15)} 等があげられる。

本論文では、日本語固有表現抽出タスク IREX により、定義された固有表現タイプ 8 種を基礎として 35 種類の質問タイプを設定する。

3. 質問タイプ同定に有効な素性

3.1 質問タイプの特徴

質問タイプを同定するのに有効な素性を検討する。まず、質問文は単語の集合で構成されていることから、質問文に出現する各単語の情報を素性として利用することが考えられる。単語の情報には、単語自身、品詞、意味情報等があげられる。

しかし、2.1 節の例題の場合は、「～会社で～はどこ」、あるいは、「旅行先で～はどこ」といった部分が質問タイプを同定するのに重要な情報になると考えられるため、「会社」や「どこ」といった、ある単語が質問文中に出現するか、しないか、という情報だけでは質問タイプを同定することは難しいと考えられる。

また、表層の単語ではなく、たとえば「(会社名)で～はどこ」、「(地名)で～はどこ」のように単語を意味のクラスに置き換えて考えることで、より汎用的な

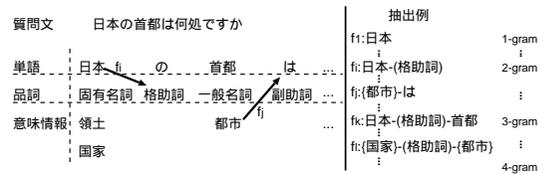


図 2 単語属性 N -gram の概略および抽出例

Fig. 2 An example of word attribute N -grams.

質問タイプ同定が可能であると考えられる。

これらの例から、質問タイプ同定では、質問文の構造、つまり、質問文中の単語がどのような関係で出現するかと、単語の表層と意味情報といった様々な種類の情報を組み合わせて考慮する必要があると考えられる。

英語質問応答システムでは、質問タイプを WordNet の意味カテゴリにマッピングし、質問文中に出現する意味カテゴリ間の関係を事前に学習した事例と照合して質問タイプ同定を行う手法が提案されている¹⁴⁾。同様に、構文情報および意味情報を付加する解析器を作成し、そこで付与される情報を利用して質問タイプ同定を行う手法も提案されている³⁾。これらの例からも、質問文に出現する単語の意味的な情報やその関係が、質問タイプを同定するのに有効であるといえる。

以上の議論から、質問タイプ同定では、質問文中に出現する各単語の持つ情報と、それらがどのような関係で質問文中に出現するかが重要であると考え、本論文では、質問タイプ同定に適した素性の抽出手法として、質問文中に出現する単語の属性間の関係を抽出する単語属性 N -gram を提案する。

3.2 単語属性 N -gram

質問文中に出現する個々の単語属性がどのような関係で質問文中に出現するかを素性として抽出する手法を述べる。

本論文では簡単のため、各単語から得られる情報、たとえば単語自身 (表記)、品詞、意味情報等をまとめて「単語属性」と呼ぶことにする。まず、質問文から各単語ごとに単語属性を抽出する。通常、各出現単語は複数の単語属性を持つ。本論文では、これら複数の単語属性の全組合せの連鎖を作成する。これは、単語属性の組合せ的 N -gram であるので、本論文では、「単語属性 N -gram」($N = 1, 2, 3, \dots$) と呼ぶ。このように単語、品詞、意味情報といった各種の属性をすべて用いることにより、質問タイプを特徴付ける表層的な定型表現や意味的な構造を網羅的に考慮した素性集合を作成することができる。

図 2 に、単語属性 N -gram の具体的な抽出例を示

<http://www.cs.nyu.edu/cs/faculty/grishman/muc6.html>

<http://cs.nyu.edu/cs/projects/proteus/irex/>

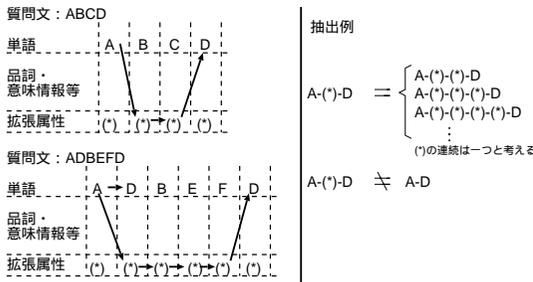


図3 拡張した単語属性 N -gram の抽出例

Fig. 3 An example of extended word attribute N -grams.

す。ここで、 $N = 1$ の場合は、個々の単語属性のみで構成される素性となる。また、素性は、質問文中にその素性が質問文中に出現した (1) か、出現しない (0) かの二値をとる。

次に、単語属性 N -gram を任意の距離離れた単語間の連鎖を考慮できるモデルに拡張する。これは、2章の例題で示したように、質問タイプを特徴付ける表現は「会社で～はどこ」のように非連続の部分にも存在すると考えられるからである。ただし、各単語が連続で出現するか、あるいは、非連続で出現するかの情報は質問タイプを特徴付けるうえで重要な情報であると考え、連続で出現する場合と非連続で出現する場合は別の素性として抽出できるようにする。

このような条件を満たすために、任意の属性を表す単語属性を導入する。図3に、拡張した単語属性 N -gram の抽出例を示す。

すべての単語は、図3中の(*)のような単語属性を持っているとする。ある1単語を挟んで出現した2つの単語AとBの関係抽出する場合、2単語間に出現する任意の単語の属性が(*)であれば、2単語間にどのような単語が現れてもA-(*)-Bという同一の素性として抽出できる。さらに、A, B間に任意の数の単語が存在したとしても、抽出される素性はA-(*)-Bであるとするにより、任意の長さ離れた単語間の関係を考慮することができる。つまり、単語属性の連鎖A-(*)-(*)-BやA-(*)-(*)-(*)-Bが得られたとしても、A-(*)-Bが得られたと考える。ただし、A-(*)-BとA-Bは別の素性として考慮される。前者は単語A, B間に1単語以上の単語が含まれることを意味し、後者は単語A, Bは連続で出現することを意味する。本論文では、このような単語属性 N -gram を用いて素性の抽出を行う。

実際の問題に適用する場合、このような組合せの N -gram では、サンプル数が多くなると素性数が爆発的に多くなることが予想される。このようなときには、

素性の出現頻度 t を閾値として扱い、出現頻度 t 未満の素性は抽出しないようにする等の方法で抽出される素性数を制限すればよい。

実際には、PrefixSpan¹⁶⁾ や、String Kernel¹⁷⁾ (Kernel法による学習手法を用いる場合) といったアルゴリズムを多少変更して適用することで、単語属性 N -gram を効率的に求めることができる。

以下の項では、単語属性 N -gram の各要素となる単語属性について述べる。

3.2.1 単語および品詞

質問文を単語(形態素)単位に分割し、品詞を付与するのに形態素解析器を利用する。本論文では、形態素解析器としてALTJAWSを用いる。ALTJAWSとは、日英翻訳システムALT-J/E¹⁸⁾ で用いられている形態素解析器である。

3.2.2 意味情報

前述の形態素解析器ALTJAWSでは意味情報として意味カテゴリが同時に付与される。意味カテゴリとは、日本語語彙大系¹⁹⁾ 内に記述されている各単語の意味を表記するためのカテゴリであり、全2715カテゴリ存在する。各カテゴリは木構造の各ノードに対応付けられて記述されており、任意のカテゴリの親カテゴリは、そのカテゴリおよび兄弟カテゴリを包括した意味を持つカテゴリとして定義されている。

各単語が、ある意味カテゴリに含まれるか含まれないかを判定する際に以下のような仮定をする。

- (1) 単語がある意味カテゴリに属する場合、対象カテゴリより上位のカテゴリにもその単語は属する。
- (2) 単語が複数の意味カテゴリに属する場合は、そのすべてのカテゴリに属する。

図4に、各単語がどの意味カテゴリに属するかを判定する例を示す。図4中の下の表で、ある単語に対して、意味カテゴリの値が1であるカテゴリには、その単語が属していることを意味し、値が0の場合は属していないことを意味する。

4. SVMを用いた質問タイプ同定手法

4.1 Support Vector Machine

Support Vector Machine (SVM) は Vapnik らによって提唱されたノンパラメトリックな機械学習手法であり、正例および負例の2クラス間の距離(マージン)を最大にするという基準で識別平面(分離平面)

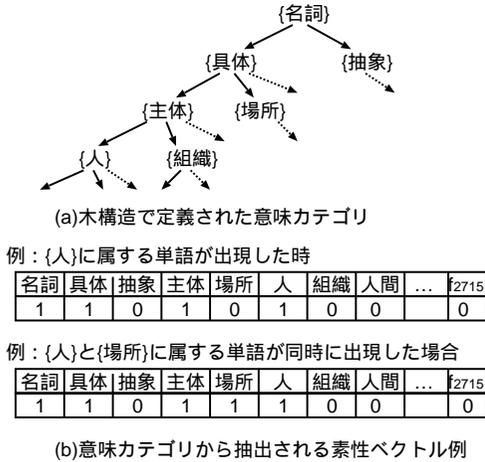


図 4 意味カテゴリに関する素性ベクトル作成方法の概略図
Fig. 4 An example of making feature vectors for semantic categories.

を決定する Large Margin Classifier の一種である . SVM に関してはすでにいくつかの解説論文^{20)~22)} があるので、ここでは簡単に紹介する .

正例、負例のラベル付きデータが m 事例与えられたとすると、これらの事例とそのラベルを以下のようなベクトルで表現する .

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m) \quad (1)$$

ここで、 $\mathbf{x}_i \in \mathbf{R}^n$ は、 i 番目のデータの素性を表す n 次元ベクトルであり、 $y_i \in \{\pm 1\}$ は i 番目データのクラスラベルである .

関数 $\phi(\mathbf{x})$ により、高次の素性空間へ事例 x を写像し、その空間で線形識別関数を求めることで、元空間の非線形識別関数を求めることができる . 識別関数 $f(\mathbf{x})$ を以下のように定義する .

$$f(\mathbf{x}) = \text{sgn}(g(\mathbf{x})) \quad (2)$$

$$g(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b \quad (3)$$

ここで、 $g(\mathbf{x}) = 0$ により定まる平面を識別平面と呼ぶ .

一般の問題に SVM を適用する場合、学習データは線形分離不能であることが多い . そこで、正の Slack 変数 $\xi_i \geq 0, (i = 1, \dots, m)$ を導入し、線形分離不可能なデータ x_i を $\xi_i / \|\mathbf{w}\|$ だけ識別平面の重みベクトル \mathbf{w} に沿って移動できるようにする .

$g(\mathbf{x}) = 1$ と $g(\mathbf{x}) = -1$ 間の距離 $d = 2 / \|\mathbf{w}\|$ はマージンと呼ばれる (図 5). Slack 変数を含めて最適なマージンを持つ識別平面の条件は以下になる .

$$\forall i, y_i [\mathbf{w} \cdot \phi(\mathbf{x}_i) + b] \geq 1 - \xi_i \quad (4)$$

上記の条件の下で、Slack 変数によるペナルティ項を含めてマージン d を最大化することは、以下の

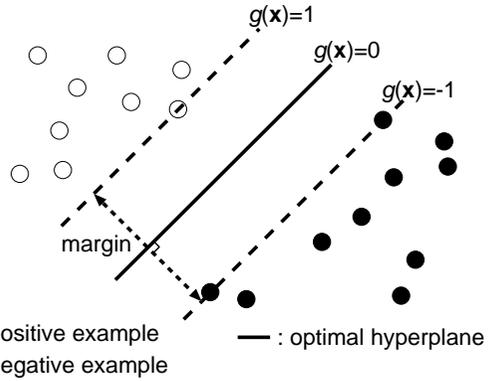


図 5 SVM により決定される識別平面
Fig. 5 The decision surface of a linear Support Vector Machine.

制約条件付き最小化問題で解くことができる .

$$\begin{aligned} \text{Minimize : } & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \quad (5) \\ \text{s.t. : } & \forall i \quad y_i [(\mathbf{w} \cdot \phi(\mathbf{x}_i)) + b] \geq 1 - \xi_i \end{aligned}$$

ただし、変数 C は、マージンとペナルティ項のパラメータを調整するパラメータである .

式 (5) はラグランジュ乗数 α_i を用いて以下の式で解くことができる .

$$\begin{aligned} \text{Maximize : } & \\ F(\alpha) = & \sum_{i=1}^m \alpha_i + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t. : } & \sum_{i=1}^m \alpha_i y_i = 0 \quad (0 \leq \alpha_i \leq C) \quad (6) \end{aligned}$$

ここで、 $K(\mathbf{x}, \mathbf{y})$ はカーネル関数と呼ばれ、 ϕ を使って次のように定義される .

$$K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{y})$$

本論文では、多項式カーネル $K^{poly}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d$ を用いる .

α_i, b の最適値 α_i^*, b^* は 2 次計画法により決定できる . このとき $\alpha_i > 0$ となる x_i をサポートベクトルと呼ぶ . すべてのサポートベクトルの集合を SV とすると、新しい事例 \mathbf{x} が与えられたときの $g(\mathbf{x})$ は、

$$g(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \quad (7)$$

で計算できる .

4.2 SVM と質問タイプ同定問題の親和性

3 章で示したように、質問タイプを同定する素性には、単語属性の関係が有効であると考えられる . しかし、これまでの機械学習手法では、大量の素性を用いる場合、非常に大量の学習サンプルを用意しない限り、

過学習により学習が有効に行われないう問題が考えられる。

SVM は、サンプル数に対して素性空間の次元数が大きくなる問題でも良好な汎化性能を持つことが、理論的にも実験的にも示されている。質問タイプ同定問題の類似問題にテキスト分類問題があげられ、SVM を用いることで非常に高性能な分類が可能であることがすでに示されている^{23)~25)}。

このように、取り扱う必要がある素性が大量に存在する問題には、サンプル数に対して高次元素性空間を用いて学習を行っても、高性能な識別器を作成可能であるという SVM の特徴は非常に有効である。

以上の理由により、本論文では、SVM を用いた質問タイプ同定手法を提案する。

5. 実験

本論文で提案した手法が、実際に質問タイプ同定に有効であるかを実験により検証する。

本論文で提案する手法は、2 つの段階に分けて考えることができる。

- (1) 質問タイプを特徴付ける素性として単語属性 N -gram を使用。
- (2) 質問タイプ同定器を学習する際に SVM を使用。

そこで、本論文では、大きくわけて、単語属性 N -gram が質問タイプ同定に有効な素性であるか検証することと、単語属性 N -gram を素性として用いた場合に、質問タイプ同定器を学習する手法として SVM を用いることの有効性を検証することを目的として実験を設定する。

また、2 章で述べたように、本論文で扱う質問タイプ同定の前提条件である、「質問文の表現を限定しない質問タイプ同定」を満たすため、実験で扱うデータは、多種多様な表現を含んだ質問文から構成されるデータを用いることとする。

5.1 実験データ

佐々木ら⁷⁾ が使用した 2,000 問を拡張して作成された質問応答テストコレクション 10,000 問を用いて実験を行った。このテストコレクションでは、各質問に対し質問タイプのラベルが 1 つ付与されている。本論文では、この与えられたラベルを「正解ラベル」と呼ぶ。また、分類器によって最終的に選ばれた質問タイプのラベルを「出力ラベル」と呼ぶ。

このテストコレクションは 35 種類の質問タイプからなるが、このうち質問数が 100 問に満たない質問タイプを 1 つにまとめ、新たな質問タイプ「その他 (OTHER)」を作成した。これにより、質問数 100 問

表 1 実験に用いた質問タイプおよび各質問タイプごとのサンプル数
Table 1 The number of samples used in the experiments.

質問タイプ	略称	サンプル数	質問対象
AGE	AGE	130	年齢
DATE	DATE	1,885	日付
EVENT	EVEN	165	事柄
LOCATION	LOCT	1,530	場所
MONEY	MONY	250	値段
NORGANIZATION	NORG	140	組織名数
NPERSON	NPER	365	人数
ORGANIZATION	ORGN	1,605	組織名
PERCENT	PCEN	190	割合
PERIOD	PERI	260	期間
PERSON	PERS	1,615	人名
PRODUCT	PROD	135	製品名
PTITLE	PTIT	270	役職名
SUBSTANCE	SUBS	130	物質名
TIME	TIME	125	時間
TITLE	TITL	150	作品名
OTHER	OTHR	1,055	その他
total		10,000	

以上の 16 質問タイプと「OTHER」を合わせて 17 質問タイプを用いて実験を行った。

表 1 に使用した質問タイプとそこに含まれるサンプル数を示す。

5.2 評価方法

質問タイプ同定は、与えられた質問に対して、「その他 (OTHER)」を含め必ず 1 つ以上の質問タイプを付与する必要がある。質問タイプ「その他 (OTHER)」も評価に用いるのは、質問タイプ同定問題では、入力される質問文が必ず事前に設定された質問タイプに属する質問とは限らないという設定の下で使用されるからである。また、必ず 1 つ以上の質問タイプが必要なものは、質問タイプ同定が、質問応答システムで解答を抽出するための手掛かりとして使用されるためである。

本論文では与えられた 1 つの質問に対して、同定器により最良と判定された質問タイプを 1 つ出力ラベルとして選出するタスクと定める。これは、実験で使用するテストコレクションでは、各質問に対して正解ラベルが必ず 1 つだけ付与されているためである。

評価には、テストコレクションを 5 分割し、4 つを学習サンプル、残り 1 つを評価サンプルとする 5fold の交差検定により分類性能を求めた。また、実験の評価指標として、質問タイプ同定器としての性能の評価には各質問の正解率 (Accuracy) と、各質問タイプごとの分類性能の評価には F 値 (F-measure) を用いた。これは、質問タイプ同定器全体としての性能を評価する指標としては正解率でよいが、各質問タイプごとの分類器の性能を評価する際には、それぞれの分類器がどの程度正確に対象となる質問タイプを分類でき

ているか評価する F 値による評価が必要となるため、2種類の指標を用いて評価を行うこととした。

5.3 実験の設定

本論文で行った実験のすべてを以下にまとめる。

- (1) SVMの有効性を検証する実験
 - ルールベース手法, 決定木学習, 最大エントロピー法と SVM の分類性能比較
 - (a) 多クラス分類手法間の性能比較
 - one vs. rest 法, pairwise 法の分類性能比較
 - (b) 多項式カーネルの次数の効果
 - 多項式カーネル 1 次, 多項式カーネル 2 次の分類性能比較
 - (c) サンプル数と性能の学習曲線
- (2) 単語属性 N -gram の有効性を検証する実験
 - 単語と意味情報, 単語の連鎖, 意味情報の連鎖, 単語の連鎖と意味情報の連鎖, 単語と意味情報の組合せによる連鎖, 単語と意味情報と品詞の組合せによる連鎖のそれぞれを素性として用いた場合の分類性能比較
 - (a) 単語属性 N -gram の閾値 t の効果
 - (b) 質問タイプを有効に特徴付ける素性の解析

5.3.1 関連研究との比較実験

SVM を用いた質問タイプ同定手法の有効性を検証する実験の方法について述べる。

本実験では, これまでに報告されている質問タイプ同定手法との性能比較を行った。比較対象として, 人手により作成されたルールによって質問タイプ同定を行うルールベース手法⁷⁾ (RULE) と他の統計的機械学習手法を用いた。

他の統計的機械学習手法を用いた研究として, Zukerman ら⁸⁾ の決定木学習を用いた研究, Ittycheriah^{9),12),15)} らの最大エントロピー法を用いた研究があげられる。そこで, 他の統計的機械学習手法の比較対象として, 決定木学習²⁶⁾ (C4.5) と最大エントロピー法 (MEM) を用いて実験を行った。なお, 決定木学習および最大エントロピー法で使用した素性は, SVM で用いた素性と同等に, 単語属性 N -gram により抽出される素性である。

ルールベース手法では, ルールベース手法の仕様で, 出力ラベルを複数個出力することを許している。また, 複数の出力ラベル間の優先順序は付かないように設計されているので, 本手法との比較では, 出力ラベル中に正解ラベルが存在した場合は正解とするという, 緩い制約とした。

各パラメータとして, 式 (5) の C を 1, 単語属性 N -gram の出現頻度の閾値は t を 30 として実験を行った。

SVM は 2 クラスの分類器なので, これを多クラス分類に拡張する必要があるが, 多クラス分類手法として one vs. rest (one-against-others 等) 法^{27),28)} を用いた。これは, 付属の実験として行った, one vs. rest 法と pairwise 法による多クラス手法間の分類性能比較の実験結果より, one vs. rest 法の方が分類性能が高かったからである (6.1.1 項参照)。

5.3.2 単語属性 N -gram の有効性を検証する実験

質問タイプ同定において, 単語属性 N -gram による素性の有効性を検証する実験の方法について述べる。

本実験では, 接続を考慮しない「S + W ($N = 1$)」, 意味情報の接続のみ「S」, 単語の接続のみ「W」, 意味情報の接続と単語の接続「S + W」, 意味と単語の組合せの接続「S * W」, 意味と単語と品詞の組合せによる接続「S * W * P」を用いて, 素性の有効性を検証した。ここで, 「+」はそれぞれの素性を合わせて用いることを意味し, 「*」は用いる単語属性の組合せを考慮することを意味することとする。また, 品詞の接続のみを用いた場合では, 正解率で 0.3 程度しか得られなかったため, 品詞のみでは質問タイプを同定するために必要な情報が得られないと判断し, 本実験では取り扱わないこととした。

本実験では, はじめに, $N = 1$ のみの素性, つまり単語属性の連鎖を考慮しない素性を用いた実験結果と, 単語属性 N -gram を考慮した場合の実験結果を比較することにより, 質問タイプを同定するうえで, 単語属性の連鎖が有効であるかを検証する。さらに, 意味情報の連鎖, 単語の連鎖, または, 意味情報の連鎖と単語の連鎖といった, 単独の単語属性で単語属性 N -gram を構成した場合の実験の結果と, 単語属性の組合せを用いた場合の実験結果の比較を行うことにより, 複数種の単語属性の組合せを考慮することが有効であるか検証を行う。

パラメータ t には, 15 を用いて実験を行った。これは, $t = 15$ より値を小さくすると, 最大エントロピー法と 2 次の多項式カーネルによる SVM で分類性能が低下するため, すべての手法で性能の良かった値を用いることとしたためである。また, 式 (5) の C は, 前述の実験同様 1 として実験を行った。

6. 実験結果および考察

6.1 SVM の有効性を検証する実験

表 2 に各質問タイプごとの F 値とその平均 (ave.F),

表 2 提案手法と他の既存手法との性能比較

Table 2 A comparison of the proposed method with the conventional methods.

type	m	F measure				
		RULE	C4.5	MEM	SVM1	SVM2
AGE	130	.784	.878	.710	.904	.873
DATE	1885	.832	.924	.931	.965	.962
EVEN	165	.545	.296	.517	.585	.574
LOCT	1530	.616	.575	.738	.744	.784
MONY	250	.734	.810	.645	.808	.829
NORG	140	.746	.654	.637	.727	.722
NPER	365	.853	.834	.836	.863	.858
ORGN	1605	.618	.541	.739	.734	.751
PCEN	190	.817	.765	.759	.812	.800
PERI	260	.439	.734	.627	.774	.745
PERS	1615	.816	.707	.873	.894	.888
PROD	135	.402	.185	.348	.587	.521
PTIT	270	.790	.751	.881	.886	.874
SUBS	130	.498	.387	.373	.646	.647
TIME	125	.718	.778	.758	.812	.823
TITL	150	.316	.299	.335	.478	.404
OTHR	1055	.434	.575	.649	.665	.666
ave.F		.645	.629	.668	.758	.748
acc.		.683	.670	.779	.807	.813

表 3 正解数による提案手法と他の既存手法との性能比較

Table 3 A comparison of the proposed method with the conventional methods based on the number of correct questions.

type	m	#. of correct question type				
		RULE	C4.5	MEM	SVM1	SVM2
AGE	130	87	111	77	118	107
DATE	1885	1391	1700	1839	1837	1832
EVEN	165	81	46	61	83	78
LOCT	1530	893	913	1196	1184	1259
MONY	250	164	199	127	198	201
NORG	140	86	82	72	97	95
NPER	365	285	297	288	318	322
ORGN	1605	1048	870	1293	1177	1217
PCEN	190	133	135	123	147	144
PERI	260	160	181	128	190	177
PERS	1615	1246	1149	1496	1490	1509
PROD	135	39	19	31	66	56
PTIT	270	182	203	223	237	229
SUBS	130	57	41	31	72	66
TIME	125	86	98	83	104	100
TITL	150	30	43	31	60	44
OTHR	1055	855	615	692	695	694
total	10000	6828	6702	7791	8073	8130

および、正解率 (acc.) を示す。また、参考資料として、表 3 に各質問タイプごとの正解数とその合計を示す。表中の SVM1 および SVM2 は、それぞれ、多項式カーネル 1 次および 2 次の SVM を表す。

提案手法である SVM を用いた結果は、F 値の平均、正解率ともに他の手法に比べて非常に良い結果であった。特に、F 値の平均が高いことから、各質問タイプに偏りが少なく、学習サンプル数の少ない質問タイプ

に対しても比較的精度の高い分類が行えていることが分かった。

ルールベース手法の傾向として、PERSON に対する「～は誰ですか」のような定型的な表現の質問に対しては、ほぼ正解することができていた。しかし、全体として良い結果が得られなかった理由は、本論文で使用したテストコレクションには、非常に多くの表現を含むことが原因である。つまり、質問に現れる表現を網羅的にルール化することができず、作成漏れしていたルールの質問に対応できなかったためである。このことから、提案手法が多様な表現を許した質問文に対しても有効であるといえる。また、質問タイプ「OTHER」の F 値が低く正解数が多いのは、ルールに適合しない質問はすべて「OTHER」に分類されたためである。

決定木学習を用いた場合は、ルールベース手法よりも分類性能が低い結果となった。原因は、使用した素性空間の次元が高いため、汎化能力の高い分類器を学習することができなかったためと考えられる。しかし、高性能な質問タイプ同定に、単語属性 N -gram のような属性間の関係の情報が必要であるとすると、大量の素性を取り扱うのが困難である決定木学習は、質問タイプ同定に適していないといえる。

最大エントロピー法の結果は、各質問タイプごとの結果を比較すると、学習サンプル数の偏りの影響を受けたと考えられる結果となっている。サンプル数の少ない質問タイプでは大幅に精度が低くなっており、このことは、正解率が比較的高い割に F 値の平均が低いことから明らかである。これは、サンプル数の少ない質問タイプでは十分な学習ができなかったことを示している。

以上の結果より、SVM を用いた質問タイプ同定の有効性を示すことができた。

6.1.1 多クラス分類手法に対する比較実験

多クラス分類手法として one vs. rest 法と pairwise 法^{29)~31)} を用いて比較実験を行った。

それぞれの多クラス分類手法において、出力ラベルは以下の基準で決定した。

- one vs. rest 法
各分類器の $g(x)$ (式 (3)) の値で最大の値を出力した分類器の正例ラベルを、与えられた質問に対する出力ラベルとする。
- pairwise 法
各分類器で判定された質問タイプに投票数 1 を与え、投票された数が最も多かった質問タイプを、与えられた質問に対する出力ラベルとする。

表 4 多クラス分類手法間の性能比較

Table 4 Classification performance versus multi-class classification method.

		one vs. rest		pairwise	
		SVM1	SVM2	SVM1	SVM2
type	m	F-measure			
AGE	130	.904	.873	.857	.791
DATE	1885	.965	.962	.961	.957
EVEN	165	.585	.574	.608	.496
LOCT	1530	.744	.784	.763	.769
MONY	250	.808	.829	.781	.746
NORG	140	.727	.722	.737	.672
NPER	365	.863	.858	.856	.843
ORGN	1605	.734	.751	.720	.717
PCEN	190	.812	.800	.780	.743
PERI	260	.774	.745	.759	.688
PERS	1615	.894	.888	.880	.878
PROD	135	.587	.521	.584	.426
PTIT	270	.886	.874	.890	.841
SUBS	130	.646	.647	.590	.460
TIME	125	.812	.823	.825	.760
TITL	150	.478	.404	.464	.301
OTHR	1055	.665	.666	.663	.610
ave.F		.758	.748	.748	.688
acc.		.807	.813	.802	.783

one vs. rest 法と pairwise 法の比較実験の結果を表 4 に示す。

表 4 より、本実験では one vs. rest 法が良い結果を得ていたことが分かる。one vs. rest 法では全サンプルに対して 1 つの分類器を作成するが、pairwise 法では分類対象の質問タイプに対するサンプルしか用いずに分類器を作成する。よって、pairwise 法では素性空間の次元数に対して学習サンプル数が非常に少なくなり、適切な学習が行えず、結果として one vs. rest 法より精度が低くなったと考えられる。

one vs. rest 法と pairwise 法の比較実験に関しては、これまでに pairwise 法を用いた方が性能が良いと報告されている³¹⁾が、本実験では、one vs. rest 法を用いた方が性能が良いことが分かった。

6.1.2 多項式カーネルの次数の効果

表 2, 4 の多項式カーネル 1 次と 2 次の結果を比較すると、多項式カーネル 2 次の結果は、サンプル数の少ないいくつかの質問タイプ (AGE, EVENT, PERCENT, PERIOD, PRODUCT, PTITLE, TITLE) に対して大幅に精度が低下していた。つまり、学習サンプル数が少ない質問タイプの場合、多項式カーネル 2 次で非線形の識別平面を考慮することが逆に悪影響を及ぼしている結果となった。これは、多項式カーネル 2 次で写像される高次元素性空間では、サンプル数が少ない影響は増加され、識別境界決定時に、サンプル数が少ない質問タイプの空間が過小に評

価されて識別平面が決定されたと考えられる。

本実験においては、単語属性 N -gram により抽出される素性空間の次元が高次元であるので、複雑な非線形識別境界を求めなくても、十分な性能を得られることが分かった。

6.1.3 サンプル数と性能の学習曲線

質問タイプごとに学習サンプル数を変化させ、サンプル数が与える分類性能への影響を検証する。

図 6 に質問タイプ DATE, LOCATION, AGE, TITLE の正解率の結果を例として示す。これらの質問タイプは、各質問タイプごとの質問数と分類性能を基準にして代表的な質問タイプを選出した。つまり、質問数が多く分類性能も比較的良かった例として DATE, 質問数が多いが分類性能があまり良くなかった例として LOCATION, 質問数は少ないが、分類性能が良かった例として AGE, 質問数が少なく分類性能も良くなかった例として TITLE を選んだ。

これらの結果から、質問タイプ DATE は学習サンプルが 100 問程度でも F 値で 9 割近い分類が可能であることから、比較的分類が簡単な質問タイプであることが分かる。また、学習サンプル 400 問程度で十分な性能が得られており、テストコレクションとして質問タイプ DATE の学習サンプル数は十分であったといえる。

逆に、質問タイプ LOCATION に関しては、学習サンプル 1200 問では、質問タイプ LOCATION を分類するのに必要な情報を十分に得ることができなかった。これは、質問タイプ LOCATION の分類が難しいことを意味する。図 6 から、学習サンプル数を増加させることで、多少性能を向上させられる可能性はあるが、9 割以上の分類性能を得るためには、学習サンプル数を増やすだけでは難しいことが読み取れる。

同様に、質問タイプ AGE では、学習サンプル 100 問程度で 9 割近い性能が得られ、質問タイプ TITLE では、学習サンプル 100 問では 5 割に満たない性能しか得られなかった。

学習サンプルの少ない質問タイプの場合は、現状で比較的良い精度を出している質問タイプでも、学習サンプルを増やすことで、性能を向上させることが可能であると期待できる。特に、質問タイプ TITLE のように、学習サンプルが少なく、かつ、精度が良くなかった質問タイプに関しては、今後学習サンプルを増加させることで、大幅に性能を向上させられる余地があることが確認できた。

6.2 素性の有効性に関する実験

単語属性 N -gram による素性の有効性を検証した

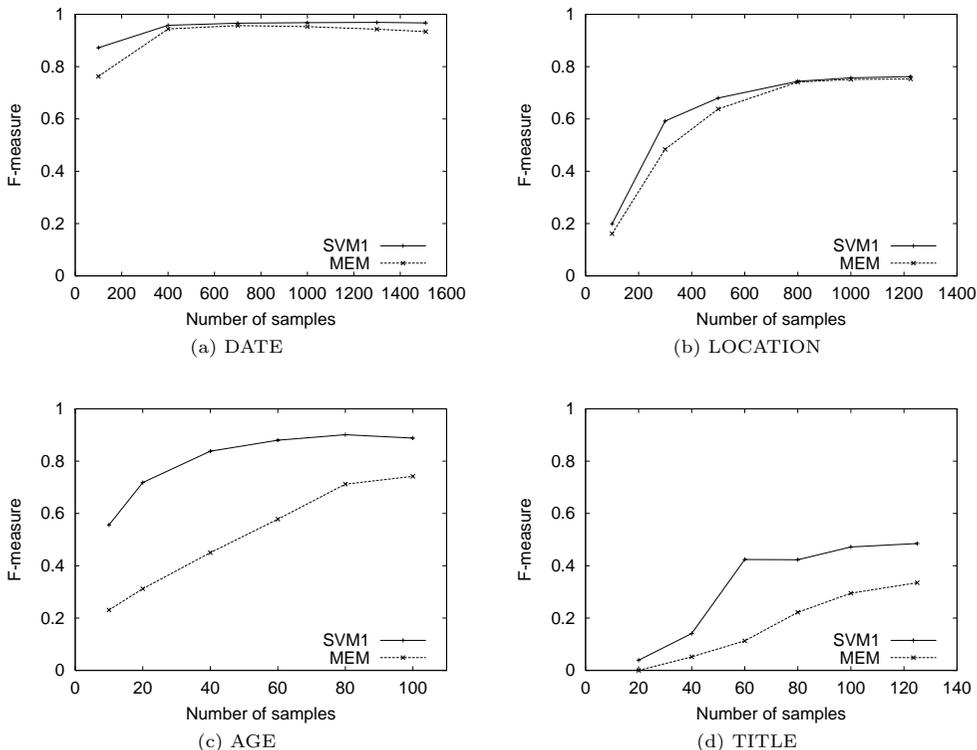


図 6 各質問タイプごとのサンプル数による分類性能の変化
Fig. 6 Learning curves of each question type.

表 5 素性の違いによる性能比較
Table 5 Effects of features.

単語属性	素性数	MEM		SVM1		SVM2	
		ave.F	acc.	ave.F	acc.	ave.F	acc.
S + W (N = 1)	4090	.466	.644	.645	.729	.643	.758
S	7565	.559	.665	.636	.724	.668	.744
W	10114	.592	.733	.673	.763	.651	.750
S + W	17681	.659	.767	.741	.798	.747	.810
S * W	45525	.684	.792	.773	.824	.754	.818
S * W * P	118509	.670	.777	.780	.824	.733	.807

実験の結果を表 5 に示す。

まず、S + W (N = 1) と S + W の結果から、連鎖を考えない場合 (N = 1) より、連鎖を考えた方が、質問タイプ同定に有効な素性が抽出できることが分かった。

次に、S + W と S * W を比較すると、意味のみ、または、単語のみといった単独の単語属性による連鎖を考えるより、単語属性の組合せを考慮することにより性能を大幅に向上させられることができた。

以上の結果より、本論文で提案した単語属性の組合せを考慮する N-gram を用いることで、質問タイプ同定に有効な素性を抽出することができることが分

かった。

ただし、本論文で提案した単語属性 N-gram の場合でも品詞を考慮した場合に、素性数が大幅に増加した割には、それほど性能を向上させることができなかった。この理由として、品詞を含んだ単語属性 N-gram には、質問タイプを有効に特徴付ける素性がほとんど存在しなかったと考えられる。また、最大エントロピー法では、品詞を考慮することで、むしろ性能が低下した。これは、品詞を考慮することで得られる効果よりも、素性数が多くなることで過学習が引き起こされる負の効果の方が大きかったと考えられる。つまり、質問タイプ同定には品詞の情報はそれほど重要ではないと考えられる。

6.2.1 単語属性 N-gram の閾値 t の効果

単語属性 N-gram のパラメータである出現頻度の閾値 t による素性抽出の制限の効果を実験により検証した。出現頻度の閾値 t を変化させたときの正解率の推移を図 7 に示す。

多項式カーネル 1 次の SVM を用いる場合は、素性数が大幅に増加しても、出現頻度の閾値 t を下げることで性能は向上した。このことから、出現頻度が低い

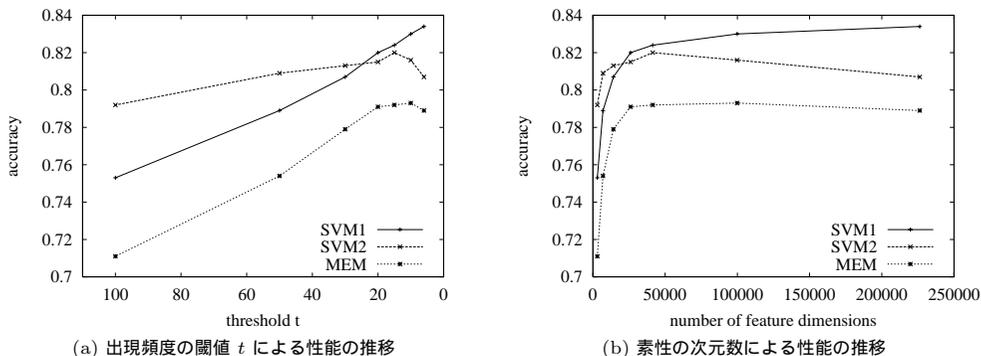


図 7 単語属性 N -gram の出現頻度の閾値 t による素性抽出の制限の効果
 Fig.7 Effects of the threshold t on the word attribute N -gram.

単語属性 N -gram の中にも質問タイプを同定するのに有効な素性が存在することが分かった。このように、低出現頻度の単語属性 N -gram も考慮できる枠組みが必要であり、素性数が多くても高性能な分類が行える SVM を用いる意味は大きいといえる。また、理想的には、考えられるすべての単語属性 N -gram から、質問タイプを同定するのに有効な素性のみを選択することができれば、さらに性能を向上させることができると考えられる。

逆に、最大エントロピー法の場合は t を小さくして素性数を増やしすぎると性能が悪化した。これは、考慮する素性が増えることで、分類に有効な素性が抽出される半面、不必要な素性も大幅に増加することから、サンプル数に対して素性空間の次元が高くなりすぎて過学習を引き起こし、分類性能が低くなったと考えられる。

同様に、多項式カーネル 2 次を用いた場合は、素性の次元がある程度小さい部分では、多項式カーネル 1 次よりも分類精度が高いが、素性空間を高次元にすると精度が悪くなった。これは、多項式カーネルの次数の効果を検証した実験でも触れたように、素性空間が高次元になることで、複雑な識別平面を考えなくても、十分に分離可能であるので、複雑な識別平面を考慮することが汎化性能を劣化させる原因になったといえる。

6.2.2 質問タイプを有効に特徴付ける素性の解析

本論文の素性抽出手法では、SVM が高次元素性空間を用いても過学習しにくいという前提から大量の素性を使用した。この素性集合は、SVM を用いることにより質問タイプ同定に有効であることは実験により確かめられたが、実際にどの素性が有効であったか明らかにするのは難しい。

しかし、近年、SVM の枠組みを利用して有効な素性

を抽出する手法が提案されている。本論文では、有効な素性を抽出する手法の 1 つとして Guyon らが提案している SVM Recursive Feature Elimination (SVM RFE)³²⁾ を取り上げた。本来この手法は、素性選択を行うことを目的としているが、同時に分類に有効な素性も抽出することができる。そこで、本論文ではこの手法を利用して質問タイプを特徴付ける素性を抽出し、その考察を行う。

SVM RFE では、式 (3) の w の値を利用して各素性が質問タイプを特徴付けるのに、どのくらい寄与しているか計算する。

各質問タイプを特徴付ける素性として上位 10 位内の素性を表 6 に示す。ここで、 $\{ \}$ に記述されているものは、意味カテゴリーのカテゴリ名を表し、 $()$ 内に記述されているものは品詞名を表す。また、何も括られていないものは、単語 (形態素) そのものである。 $(*)$ は、任意の単語列を表す単語属性である。

ここで注意しなくてはならないのは、実験で得られた質問タイプ同定の分類性能は素性をすべて考慮した場合のものであり、この上位 10 位の素性のみで高精度な分類が行えるということではない。しかし、質問タイプを特徴付けるのにより有効である素性の指標の 1 つと考えることができる。

まず、全体の傾向として、上位に残る素性は意味カテゴリーに関連するものが多く見られた。

個別に見ると質問タイプ AGE で意味カテゴリー { 年齢 }, MONEY で { 値・額 }, PERCENT で { 計算値 }, TIME で { 時刻 } 等、直観的にもあてはまる意味カテゴリーを含む素性が上位に多く現れた。また、1 種類の意味カテゴリーではないが、EVENT, LOCATION, NORGANIZATION, PERIOD, PERSON, PRODUCT, PTITLE, SUBSTANCE, TITLE で

表 6 各質問タイプを有効に特徴付けると考えられる素性例
Table 6 Examples of effective features for each question type.

AGE	DATE	EVENT	LOCATION	MONEY	NORGANIZATION
{ 年齢 }	いつ	{ 挙行 }-(格助詞)	どこ	いくら	{ 格助詞 }-(*)-{ 企業 }
年齢	{ 時間 }	{ 式・行事等 }	{ 地域(人間活動) }	{ 値・額 }	{ 企業 }
{ 年齢 }-(副助詞)	{ 日程 }	{ 格助詞 }-(*)-{ 挙行 }	どこ-で	{ 収入 }	数
{ 年齢 }-は	いつ-(*)-(格助詞)	イベント	会場	{ 値・額 }-(格助詞)	{ 企業 }-(*)-{ 数 }
{ 年齢 }-(*)-(助動詞)	{ 時間 }-(*)-(格助詞)	{ 格助詞 }-(*)-{ 行事 }	{ 地域(範囲) }-(格助詞)	いくら-(*)-(助動詞)	{ 公共機関 }-(*)-{ 数 }
{ 年齢 }-(*)-(記号)	{ 日程 }-(*)-(記号)	{ 日 }-(*)-で	{ 名詞 }-(*)-(格助詞)	{ 値・額 }-(*)-(記号)	{ 公共機関 }-(格助詞)-{ 数 }
{ 格助詞 }-(*)-{ 年齢 }	{ 日程 }-(*)-(助動詞)	{ 祭 }	{ 位置 }-で	{ 負債 }	数-(*)-(記号)
何-歳	{ 日程 }-(格助詞)	{ 会議 }	どこ-(格助詞)	額	幾-つ
{ 数 }-{ 年齢 }	は-いつ	{ 行事 }	{ 地域(人間活動) }-で	総額	幾-{ 単位 }
何-{ 年齢 }	は-{ 時間 }	{ 公共機関 }-(*)-{ 日 }	{ 地域(人間活動) }-(格助詞)	{ 格助詞 }-(*)-{ 収入 }	幾-{ 個数・回数等 }

NPERSON	ORGANIZATION	PERCENT	PERIOD	PERSON
人	どこ	{ 計算値 }	期間	誰
何-人	{ 企業 }-を	{ 格助詞 }-(*)-{ 計算値 }	{ 格助詞 }-(*)-期間	{ 不定称 }
{ 数 }-人	の-(*)-{ 企業 }	{ 計算値 }-(*)-(記号)	{ 時 }	誰-(格助詞)
人数	{ 機関 }	の-(*)-{ 計算値 }	期間-(格助詞)	{ 不定称 }-(格助詞)
の-{ 個数・回数等 }	どこ-(格助詞)	{ 計算値 }-(副助詞)	の-(*)-期間	{ 偉人 }
{ 個数・回数等 }	{ 位置 }	割合	どれぐらいの	{ 男 }-{ 競技場 }
{ 個数・回数等 }-(*)-(記号)	{ 記号 }-(名詞)-(格助詞)	{ 計算値 }-(*)-(助動詞)	{ 副助詞 }-(*)-{ 期間(自然・人間活動等) }	{ 記号 }-(*)-(格助詞)-{ 人名 }
数	{ 名詞 }-(*)-(副用語)	{ 計算値 }-は	期間-(*)-(記号)	{ 人名 }
{ 格助詞 }-{ 個数・回数等 }	{ 事務所 }	何-{ 計算値 }	期間-(*)-(助動詞)	{ 人名 }-(*)-(記号)
居る	どこ-が	{ 数 }-{ 計算値 }	{ 時点 }	は-誰

PRODUCT	P.TITLE	SUBSTANCE	TIME	TITLE
{ 菓子 }	{ 敬称 }	{ 物質(本体) }	{ 時刻 }	{ 創作物 }
{ 仕事場 }	氏	物質	{ 時刻 }-(*)-(助動詞)	{ 出版物 }
{ 仕事場 }-(格助詞)	{ 立場 }	{ 属性 }	{ 時刻 }-(*)-(記号)	作品
{ 格助詞 }-{ 菓子 }	役職	{ 格助詞 }-(*)-{ 物質(本体) }	{ 格助詞 }-(*)-{ 時刻 }	{ 格助詞 }-(名詞)
{ 売り }	{ 職業 }	{ 格助詞 }-(*)-物質	何-時	{ 題名 }
{ 乗り物(本体(移動(陸圏))) }-は	{ 名詞 }-氏	{ 物質(本体) }-(*)-(記号)	{ 時刻 }-を	{ 動詞 }-(*)-{ 出版物 }
{ 乗り物(本体(移動(陸圏))) }-(副助詞)	{ 敬称 }-は	は-(*)-{ 物質(本体) }	{ 時刻 }-(格助詞)	{ 創作物 }-(*)-(格助詞)
{ 名詞 }-(格助詞)	{ 敬称 }-(副助詞)	{ 副助詞 }-(*)-{ 物質(本体) }	の-(*)-{ 時刻 }	{ 創作物 }-(格助詞)
{ 電気部品 }	{ 名詞 }-{ 敬称 }	{ 動詞 }-(*)-{ 物質(本体) }	{ 時刻 }-(*)-(助動詞)	タイトル
{ 企業 }-(格助詞)	どうい	{ 副用語 }-(*)-{ 物質(本体) }	{ 時刻 }-(格助詞)-{ 助動詞 }	{ 出版物 }-(格助詞)

も、それぞれ、質問タイプに関連する意味カテゴリが含まれる素性が上位を占めていた。このことから、質問タイプと意味的に近い意味カテゴリは、密接なつながりがあることが証明された。また、質問タイプ同定は質問の意図を解析することととらえることもできることから推測できるように、質問タイプを特徴付ける素性として意味情報が非常に有効であることが分かった。

単語では、DATE に対する「いつ」、LOCATION・ORGANIZATION での「どこ」、MONEY の「いくら」、単語の接続では、NPERSON の「何-人」、TIME の「何-時」等、各質問タイプで多く出現すると考えられる素性が上位に出現していた。このように、SVM RFE によって質問タイプに関連してよく現れる単語を上位に選出することができた。ただし、前節の単語のみを用いた実験結果からも明らかのように、上記の単語が出現しない質問文も数多くあることから、これらの単語のみを考慮するようなルールを作成しても、高精度な解析を行うことはできない。

上位 10 位に出現する単語属性 N -gram は、 $N =$

1, 2, 3 程度の次数が小さい素性が多かった。これは、次数の大きい素性は次数の小さい素性を含んだ形であるため、必然的に次数の小さい素性の方がより重要と判定される可能性が高くなったと考えられる。このことから、上位に位置する素性で最も多い単語属性 N -gram のパターンは、 $N = 1$ の意味カテゴリ、或は、 $N = 2$ の意味カテゴリと助詞の連鎖であった。

逆に、 N が大きい単語属性 N -gram は最上位にはほとんど現れない結果となった。しかし、これは、 N が大きい単語属性 N -gram が重要でないということではなく、上位数十から数百以内には、質問タイプを特徴付ける表現と考えられる素性が抽出されている。

表 7 に、ルールベース手法⁷⁾と比較して類似する素性の例を示す。なお、左側の数字は素性の順位付けアルゴリズムによる順位である。これは、素性集合から人手による抽出を行った結果なので、すべてのルールと比較しているわけではない。しかし、上位には直観的に質問タイプを特徴付けるのに有効であると思われる表現が現れている。このように、SVM における有効素性を調査すると、直観的にも質問タイプ同定問

表 7 質問タイプを特徴付ける典型的な定型表現となる単語属性 N -gramTable 7 Examples of word attribute N -gram typically used in each question type.

type	rank	feature
AGE	90	(格助詞)-(*)-{年齢}-(格助詞)
	94	{年齢}-を-(*)-.
	456	は-(*)-幾-つ
DATE	12	{日}-は-?-文末)
	71	{時機}-を-(*)-(記号)
	113	{日}-(*)-知る-たい-.
	289	は-いつ-(*)-(助動詞)
EVNT	587	{催し}-[副助詞]-何-[助動詞]-[終助詞]
LOCT	45	ある-(*)-の-{名称}
	73	どこで-(*)-(格助詞)
	392	(名詞)-(*)-は-どこ
ORGN	160	{企業}-[格助詞]-名前-[格助詞]
MONY	98	{値・額}-は-?-文末)
NORG	113	{企業}-は-{数}-{個数・回数等}
NPER	765	{数}-人-[助動詞]-[終助詞]
PERI	100	{数}-年-(*)-(助動詞)
	250	は-(*)-{年月日}-くらい
PTIT	437	{仕事}-は-何
SUBS	109	{物質(本体)}-(*)-何-です
TIME	58	は-(*)-(格助詞)-いつ
	588	いつ-[動詞]-[助動詞]-[準体助詞]
TITL	127	(格助詞)-(*)-の-{題名}

題でルールとして用いることが可能な表現が上位に見つけられる結果となった。

これらの検討結果を用いて、最終的には分類に有効であると考えられる素性のみ抽出し、素性空間を再構成した後に再学習をすることで、性能を向上させることが可能であると考えられる。ただし、このような作業を行うには、実際にどの順位までの素性を用いて再学習すると確実に性能が向上するという指標が不明であるため思考錯誤的に性能が良くなる素性集合を見つける必要があり、人手によるコストと計算時間が非常にかかる点が問題となる。

7. 今後の課題

7.1 係り受け解析の導入

本論文では、素性抽出時に単語の係り受け関係を考慮せずに、単語 N -gram により表層的な解析にとどめた。しかし、質問文中の係り受け関係を考えることで、より詳細な文の構造を素性に反映することができる。係り受け解析を質問タイプ同定問題に導入するには、文正解率という点で高精度であること、質問文のような口語的表現にも高精度に解析できること、単語の意味情報が付加できることが必要である。これらの条件から、現状で係り受け解析を用いると、解析誤りから素性抽出時に含まれるノイズが増加して、最終的

な性能が劣化することが考えられる。

よって、本論文では係り受け解析は行わなかったが、今後、高精度で意味情報も付加できる係り受け解析器が作成されたら、質問タイプ同定に適用することは有効であると考えられる。

また、係り受け解析を導入する場合でも、本論文で用いた単語属性 N -gram を、係り受け関係上の単語属性 N -gram へ変更するだけでよく、それ以外は本論文で用いた手法をそのまま使用できる。

7.2 質問タイプ同定問題の類似問題への適用

質問タイプ同定問題は、質問文を扱うタスクには共通した問題であると考えられる。本論文では、質問応答技術における質問タイプ同定を取り扱ったが、異なるタスクでも質問タイプを適切に変更することによって、本論文で提案した手法をそのまま使用することができる。たとえば、対話システムの話者の意図解析等に応用可能であると考えられる。

このように、今後は本論文で提案した質問タイプ同定手法を質問応答技術以外の分野への適用を考えていくことを検討している。

8. 結 論

本論文では、質問タイプ同定に適した素性抽出手法として単語属性 N -gram を提案し、この素性を用いた統計的機械学習手法 SVM による質問タイプ同定手法を提案した。

本論文の実験において、質問タイプ同定には、単語の持つ複数の属性の組合せ的 N -gram である単語属性 N -gram が有効であることと、この素性を用いて質問タイプ同定を行う際には、サンプル数に対して素性空間が高次元でも高精度な分類を行うことが可能な SVM が有効であることを示した。

また、本論文で用いた素性に関して、素性選択手法の 1 つである SVM REF を用いて各質問タイプを特徴付ける素性を抽出し検証した。同時に、本論文で用いた単語属性 N -gram によって抽出される素性が質問タイプ同定に有効であることも明らかにした。

参 考 文 献

- 1) Vicedo, J.L. and Ferrandez, A.: A Semantic approach to Question Answering system, *Proc. 9th Text Retrieval Conference (TREC-9)*, NIST (2000).
- 2) Harabagiu, S., Pasca, M. and Maiorano, S.: FALCON: Boosting Knowledge for Answer Engines, *Proc. 9th Text Retrieval Conference (TREC-9)*, NIST (2000).

- 3) Hovy, E., Hermjakob, U. and Lin, C.-Y.: The Use of External Knowledge of Factoid QA, *Proc. 10th Text Retrieval Conference (TREC 2001)*, Gaithersburg, MD, U.S.A., NIST (2001).
- 4) 賀沢秀人, 加藤恒昭: 意味制約を用いた日本語質問応答システム, 情報処理学会自然言語処理研究会 NL-140, 情報処理学会, pp.173-180 (2000).
- 5) 村田真樹, 内山将夫, 井佐原均: 類似度に基づく推論を用いた質問応答システム, 情報処理学会自然言語処理研究会 NL-135, 情報処理学会, pp.181-188 (2000).
- 6) 佐々木裕, 磯崎秀樹, 平 博順, 廣田啓一, 賀沢秀人, 平尾 努, 中島浩之, 加藤恒昭: 質問応答システムの比較と評価, 電子情報通信学会技術報告 NLC2000-24, pp.17-24, 電子情報通信学会 (2000).
- 7) 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, 情報処理学会 (2001).
- 8) Zukerman, I. and Horvitz, E.: Toward Understanding WH-Questions: A Statistical Analysis, *Proc. Association for Computational Linguistics (ACL-2001)*, ACL (2001).
- 9) Ittycheriah, A., Franz, M., Zhu, W. and Ratnaparkhi, A.: Question Answering Using Maximum-Entropy Components, *Proc. North American Chapter of the Association for Computational Linguistics (NAACL 2001)*, ACL, pp.33-39 (2001).
- 10) Cortes, C. and Vapnik, V.N.: Support Vector Networks, *Machine Learning*, Vol.20, pp.273-297 (1995).
- 11) Vapnik, V.N.: *The Nature of Statistical Learning Theory*, Springer (1995).
- 12) Ittycheriah, A., Franz, M., Zhu, W. and Ratnaparkhi, A.: IBM's statistical question answering system, *Proc. 9th Text Retrieval Conference (TREC-9)*, NIST (2000).
- 13) Fellbaum, C.E.: *WordNet: An Electronic Lexical Database*, MIT Press (1998).
- 14) Harabagiu, S., Pasca, M. and Maiorano, S.: Experiments with Open-Domain Textual Question Answering, *Proc. COLING-2000 (COLING-2000)* (2000).
- 15) Ittycheriah, A., Franz, M. and Roukos, S.: IBM's Statistical Question Answering System — TREC-10, *Proc. 10th Text Retrieval Conference (TREC 2001)*, NIST (2001).
- 16) Pei, J., Han, J., Mortazavi-Asl, B. and Pinto, H.: PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth, *Proc. 17th International Conference on Data Engineering (ICDE 2001)*, pp.215-224 (2001).
- 17) Lodhi, H., Saunders, C., Taylor, J., Cristianini, N. and Watkins, C.: Text Classification using String Kernel, *Journal of Machine Learning Research*, Vol.2, pp.419-444 (2002).
- 18) Ikehara, S., Shirai, S., Ogura, K., Yokoo, A., Nakaiwa, H. and Kawaoka, T.: ALT-J/E: A Japanese to English Machine Translation System for Communication with Translation, *Proc. 13th IFIP World Computer Congress*, pp.80-85 (1994).
- 19) 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: 日本語彙大系, 岩波書店 (1997).
- 20) 津田宏治: サポートベクトルマシンとは何か, 電子情報通信学会誌, Vol.83, pp.460-466 (2000).
- 21) 高須淳宏: 発見科学とデータマイニング, Support Vector Machine による分類, pp.118-127, 共立出版 (2001).
- 22) 前田英作: 痛快! サポートベクトルマシン—古くて新しいパターン認識手法, 情報処理学会誌, Vol.42, No.7, pp.676-683 (2001).
- 23) Joachims, T.: Text Categorization with Support Vector Machines: Learning with Many Relevant Features, *Proc. European Conference on Machine Learning (ECML '98)*, pp.137-142 (1998).
- 24) Taira, H. and Haruno, M.: Feature Selection in SVM Text Categorization, *Proc. 16th Conference of the American Association for Artificial Intelligence (AAAI '99)*, pp.480-486 (1999).
- 25) 平 博順, 春野雅彦: Support Vector Machine によるテキスト分類における属性選択, 情報処理学会論文誌, Vol.41, No.4, pp.1113-1123 (2000).
- 26) Quinlan, J.R.: *C4.5: Programs for Machine Learning*, Morgan Kaufmann (1993).
- 27) Scholkopf, B., Burges, C. and Vapnik, V.: Extracting support data for a given task, *Proc. 1st International Conference on Knowledge Discovery & Data Mining*, AAAI Press. (1995).
- 28) Blanz, V., Scholkopf, B., Bulthoff, H.H., Burges, C., Vapnik, V. and Vetter, T.: Comparison of View-Based Object Recognition Algorithms Using Realistic 3D Models, *ICANN*, pp.251-256 (1996).
- 29) Weston, J. and Watkins, C.: Multi-class support vector machines, Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, Egham (1998).
- 30) Weston, J. and Watkins, C.: Support Vector Machines for Multi-Class Pattern Recognition, *Proc. 7th European Symposium on Artificial*

Neural Networks (1999).

- 31) KreBel, Y.H.-G.: *Advances in Kernel Methods, Pairwise Classification and Support Vector Machines*, pp.255–268, MIT Press (1998).
- 32) Guyon, I., Weston, J., Barnhill, S. and Vapnik, V.N.: Gene Selection for Cancer Classification using Support Vector Networks, *Mahine Learning*, Vol.46, pp.389–422 (2002).

(平成 14 年 11 月 19 日受付)

(平成 15 年 9 月 5 日採録)



鈴木 潤 (正会員)

1999 年慶應義塾大学理工学部数理科学科卒業。2001 年同大学院理工学研究科計算機科学専攻修士課程修了。同年日本電信電話株式会社入社。現在, NTT コミュニケーション

科学基礎研究所に所属。主として自然言語処理, 機械学習に関する研究に従事。ACL, 言語処理学会各会員。



佐々木 裕 (正会員)

1986 年筑波大学第三学群情報学類卒業。1988 年同大学院理工学研究科修士課程修了。同年日本電信電話株式会社入社。現在, NTT コミュニケーション科学基礎研究所に所属。

博士(工学)。1995 年~1996 年サイモン・フレーザー大学(カナダ)客員研究員。主として自然言語処理, 機械学習に関する研究に従事。ACL, 人工知能学会, 言語処理学会等各会員。



前田 英作 (正会員)

1984 年東京大学理学部卒業。1986 年同大学院理学系研究科修士課程修了。同年日本電信電話株式会社入社。現在, NTT コミュニケーション科学基礎研究所。知能情報研究部知識

処理研究グループリーダー。工学博士。1995 年~1996 年ケンブリッジ大学(英国)客員研究員。主としてパターン認識, 統計的機械学習, 生物情報処理の研究に従事。IEEE, 電子情報通信学会, 日本バイオインフォマティクス学会各会員。