

総合学習による質問応答システムの新しい構成法 ～CLQAに向けて

佐々木 裕

ATR 音声言語コミュニケーション研究所

〒619-0288 京都府相楽郡精華町光台2丁目2-2

E-mail: yutaka.sasaki@atr.jp

あ ら ま し 本報告では、CLQA(Cross-Language Question Answering)に向けた新たなアプローチとして、質問応答を「質問文によりバイアスされた用語抽出 (QBTE: Question-Biased Term Extraction) 問題」と捉え、質問文と文書の特徴から解答を直接抽出する枠組みを提案する。このアプローチは、非常に複雑な処理となるため、人手作成のパターンや評価関数による実現は不可能に近いが、実現すれば、質問タイプ(解答タイプ)の体系が不要になるという大きな利点がある。本報告では、最大エントピー法を用いることにより、質問・正解データから、質問文の答えを直接抽出する方法を総合的に学習することにより、このようなアプローチが可能であることを示す。CRL QA データを用いて、10 分割交差検定を行った結果、人手による評価で、従来手法に近い MRR 0.35、5 位以内正解率約 50% が得られることが明らかになった。

キーワード 質問応答 QBTE 最大エントロピー法

A Comprehensive Learning Approach to Trainable QA: Toward CLQA

Yutaka Sasaki

ATR Spoken Language Translation Research Laboratories

2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan

E-mail: yutaka.sasaki@atr.jp

Abstract This report regards *Question Answering (QA)* as *Question-Biased Term Extraction (QBTE)*. This new QBTE approach liberates QA systems from the heavy burden imposed by *question types* (or *answer types*). In conventional approaches, a QA system analyzes a given question and determines the question type, then select answers from among answer candidates that match the question type. This means that the output of a QA system is restricted by the design of the question types. To confirm the feasibility of our QBTE approach, we conducted experiments on the CRL QA dataset based on 10-fold cross validation, using the maximum entropy model as an ML technique. Experimental results showed that the trained system achieved approximately 0.35 in MRR and 50% in TOP5 accuracy.

Keywords: question answering, question-biased term extraction, maximum entropy model

1 はじめに

従来、大量の文書を用いて自然文によるユーザからの質問に答える質問応答 (QA: Question Answering) システムは、下記のような、質問解析、文書検索、解答候補抽出、解答候補選択の4つのコンポーネントからなっていた。

質問解析 質問文を解析し、質問文の質問タイプ（または、解答タイプ）を同定する。

文書検索 質問文に関連する文書を大量の文書データから取り出す。

解答候補抽出 文書から質問タイプに合った表現を解答候補として取り出す。

解答選択 質問タイプ、検索語等の特徴を利用して、解答候補から解答を優先度を付けて選択する。

質問タイプには、主として PERSON, DATE といった固有表現、および質問の対象になりやすい FISH, BIRD といったクラス名が用いられてきた。また、多くのシステムで、質問タイプの体系をそのまま解答タイプの体系として利用している。例えば、質問タイプが PERSON である質問文に対しては、文書中から人名を解答候補として抽出する。

しかしながら、このような従来の構成には、QA システムが解答できる質問文の対象が、質問タイプという中間表現によって限定されるという欠点がある。また、質問タイプの体系にあわせて解答候補の抽出ツールを準備することは、非常に人手と時間のかかる作業である。特に、CLQA (Cross-Language Question Answering) システムを構築する際には、それぞれの言語について解答候補の抽出ツールを準備する必要がある。

近年、SAIQA-II [10] 等、機械学習技術により QA システムの各コンポーネントを構築する研究も行われている [3, 4, 6, 7, 11, 12, 13, 14]。しかしながら、このようなアプローチをとったとしても、各システムの質問タイプの体系に合せた大量の学習データの作成には、多大な労力を要する。さらに、質問タイプの体系の追加変更は、大量の学習データ全体の修正変更を意味する。

例えば、中国語の質問文とその正解のペアが、10,000 ペアあったとすると、中国語の質問文に対して、各システム独自の質問タイプの体系にあわせて、人手により質問文の分類を行う必要がある。加えて、そ

の質問タイプの体系に合せて、解答候補抽出のために、中国語の文書にタグを付与した訓練データも必要となる。

もし、質問タイプの変更を行う必要が生じた場合には、大量のデータ全体の見直しが必要となる。例えば、質問タイプ ORGANIZATION を、COMPANY と SCHOOL とそれ以外の組織の3つの分類に詳細化するという修正を行うためには、ORGANIZATION に関する質問文を人手により再分類し、解答候補抽出のための訓練データの ORGANIZATION タグをこの3種類のいずれかに人手で修正するという作業が必要となる。

このような問題を解決するため、本報告では、新たなアプローチとして、質問応答を「質問文によりバイアスされた用語抽出 (QBTE: Question-Biased Term Extraction) 問題」と捉え、質問タイプという中間表現を用いることなく、質問文と文書の特徴に基いて解答を直接抽出する方法について述べる。

質問タイプを用いることなく、質問と正解の特徴から、解答を文書から直接抽出することは、様々なパラメータが複雑に影響する、非常に困難な問題であり、人手作成のパターンや評価関数により実現することは、ほとんど不可能である。そのため、そもそもこのようなアプローチが可能であるかというレベルから検証する必要がある。解答として適切な文字列が切り出されるのか、もしうまく固有表現やクラス名等の表現単位が切り出されたとしても、多くの候補の中から解答として適切なものを選択できるのかという点はこれまで全く解明されていない。

そこで、本報告では、最大エントロピー法という機械学習の手法を用いることにより、質問文の特徴、文書の特徴、および両者の特徴の組み合わせを素性としたデータから、解答を直接抽出する方法を総合的に学習する方法を与え、実験によりどの程度の性能が得られたかを報告する。

2 準備

2.1 学習データ

タグ集合とともに、学習型 QA システムの実現に必要な要素として学習データがある。本研究で用いた学習データは以下の通りである。

文書集合 毎日新聞 9 5 年の新聞記事

質問・正解集合 質問・正解集合として、CRL QA データを用いた。このデータセットは、質問

表 1: CRL QA データセットの質問タイプ

質問文数	質問タイプ数	例
1～9	74	賞名, 罪名, 競技名
10～50	32	割合表現, 製品数, 年期间
51～100	6	国名, 企業名, グループ名
100～300	3	人名, 日付表現, 金額表現
計	115	

文 2000 問からなり、各質問文には、質問タイプと正解、正解の出現する毎日新聞 9 5 年の記事 ID が与えられている。各質問文には、階層的に細分化された 115 種類の質問タイプのうちの 1 つが与えられている。

本報告では、QA システムの構築および実行フェーズの両者において、学習データに付与されている質問タイプの情報は一切利用しない。

ちなみに、質問・正解データの各質問タイプについて、その質問タイプが質問文に与えられている数をカウントすると、表 1 に示すように質問数が 9 以下の質問タイプが 6 割以上を占める。このように、詳細な分類については、データ数が少なくなるため、SAIQA-II [10] と同様に学習させることは困難である。¹

文書集合は、学習データの作成のために用いられるとともに、評価実験において、評価用の質問の答えを抽出するための知識源としても利用される。

2.2 最大エントロピー法による学習

この節では、学習型 QA システムの説明のための準備として、機械学習アルゴリズム「最大エントロピー法」の概要、および最大エントロピー法のためにどのような学習データを準備するかについて述べる。

2.2.1 最大エントロピー法

全ての入力記号の集合を $\mathcal{X}$ 、全てのクラスラベルの集合を $\mathcal{Y}$ とする。入力 $x = \{x_1, \dots, x_m\}$ ($x_i \in \mathcal{X}$) と出力 $y \in \mathcal{Y}$ の対 (x, y) により事例 を表現する。

入力 x が与えられた時の出力 y に関する条件付き確率を $p(y|x)$ とすると、最大エントロピー原理 [1] は、 $p(y|x)$ に関するエントロピー $H(p)$ が最大にな

¹ 但し、質問分類の学習については、質問タイプの階層に沿って、質問分類を段階的に詳細化する分類器を学習する研究は [12] で行われている。

る確率モデル $p^* = \operatorname{argmax}_{p \in C} H(p)$ を、制約を満たす確率分布 C から求める最適化問題である。

データ $(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})$ が与えられたとき、本報告では、素性関数 f_i を以下のように定義する。 $\bigcup_k (x^{(k)} \times \{y^{(k)}\}) = \{\langle \tilde{x}_1, \tilde{y}_1 \rangle, \dots, \langle \tilde{x}_i, \tilde{y}_i \rangle, \dots, \langle \tilde{x}_m, \tilde{y}_m \rangle\}$ により $\tilde{x}_i, \tilde{y}_i (1 \leq i \leq m)$ を定め、

$$f_i(x, y) = \begin{cases} 1 & \text{if } \tilde{x}_i \in x \text{ and } y = \tilde{y}_i \\ 0 & \text{otherwise} \end{cases}$$

すなわち、 x に含まれる各入力記号とそのクラスラベルの組み合わせをそのまま最大エントロピー法の素性（関数）とする。

ラグランジュ乗数 $\lambda = \lambda_1, \dots, \lambda_m$ を用いて、 H の双対関数 Ψ を考える。

$$\Psi(\lambda) = - \sum_x \tilde{p}(x) \log Z_\lambda(x) + \sum \lambda_i \tilde{p}(f_i)$$

ここで、 $Z_\lambda(x) = \sum_y \exp(\sum_i \lambda_i f_i(x, y))$ 、 $\tilde{p}(x)$ および $\tilde{p}(f_i)$ は、それぞれ x と f_i の学習データでの分布である。

双対な最適化問題 $\lambda^* = \operatorname{argmax}_\lambda \Psi(\lambda)$ は制約のない最適化問題として効率的に解くことができ、その結果、目的の確率モデル $p^* = p_{\lambda^*}$ を以下の式で求めることができる。

$$p_\lambda(y|x) = \frac{1}{Z_\lambda(x)} \exp \left(\sum_i \lambda_i f_i(x, y) \right)$$

2.2.2 学習データ

従来の QA システムで行なわれてきた、質問文の分類や解答の選択に相当する機能を機械学習により実現するためには、質問文集合に含まれる質問文とその質問タイプや、文書中での正解の現われ方に関する特徴をベクトルまたは集合で表現する必要がある。一般的には、特徴（素性）の値をベクトル化した素性ベクトルとクラスラベルを学習データとするが、最大エントロピー法においては、素性関数が素性に相当し、素性関数の値が素性の値に相当する。

本報告では、入力に含まれる入力記号とクラスラベルの組み合わせがそのまま素性関数を決定する。以下、簡単に入力データの作成法を説明する。ある生徒が特徴として「身長 125cm、体重 35kg、兄弟なし、好きな色は黄色」という特徴を持っていたとする。数値は、いくつかの区間（bin）にわけて表現する。例えば、身長は 150cm 未満を S、150～

170cm を M, 170 以上を L で表し, 例えば, 体重は 50kg 未満を S, 50~70kg を M, 70kg 以上を L で表すとする. この生徒の入力データは, $x = \{ \text{身長:S, 体重:S, 兄弟:無, 色:黄色} \}$ となる. ここで, 入力記号の表現法は, 単なる一例であり, 入力記号集合において一意であれば良い.

3 総合学習による学習型 QA モデル

本節では, 質問・正解データから QBTE アプローチにより QA システムを構築および実行する枠組み (ATR QA-model 1) について述べる.

ユーザから質問文が与えられると, 以下のような 2 つの処理を行なう.

文書検索 質問文の単語により, 関連する文書を大量の文書データから上位 N 件取り出す.²

解答抽出 質問文と文書から入力データを作成し, 学習されたモデルによって素性集合を評価し, 質問タイプに合った表現を解答として抽出する.

本報告では, 文書検索は, 従来の検索手法である *idf* 値を利用できるため, 以下, 解答抽出に絞って述べる.

単語を w_i で表現し, 文書を単語列 $w_1, w_2, \dots, w_k, \dots$ と表記する. QBTE アプローチによる, 質問文に対する解答の抽出問題とは, 各単語 w_i を, 質問と文書の特徴に基いて, 解答単語列に含まれる単語か否かに分類する問題となる.

つまり, 入力 $x^{(i)}$ は, 文書集合の各単語すべてに対して作成され, クラスタブル集合 $\mathcal{Y}$ は, 下記のようなになる.

I: 正解文字列の途中または最後の単語である.

O: 正解文字列の単語列に含まれない.

B: 正解文字列の開始単語である.

クラスタブルとして, Ramshaw らの IOB [8] のバリエーションである IOB2 [9] を用いた.

各単語の入力 $x^{(i)}$ は, 次の定義に従って与えられる.

3.1 特徴抽出 (feature extraction)

学習データとして, 大きく分けて下記の 3 つグループの特徴を入力データに採用した.

- 質問文の特徴 (質問特徴集合)

² 本報告では $N = 1$.

- 文書の特徴 (文書特徴集合)
- 特徴の組み合わせ (組み合わせ特徴集合)

3.1.1 質問特徴集合

質問特徴集合 (question feature set) は, 質問文のみから得られる特徴である. 1 つの質問文に対して, 1 種類の質問特徴集合が定まるので, ある質問文に対する解答単語列を抽出する場合, 各単語に与えられる質問特徴集合は同一である.

作成される質問特徴集合の各特徴は以下の通りである. なお, 品詞体系は形態素解析ツール ChaSen が出力する IPA の最大 4 階層の品詞体系を用いている. 例えば, 「多岐川」の品詞は「名詞-固有名詞-人名-姓」であり, 助詞「が」の品詞は「助詞-格助詞-一般」である. 以下, 最左の品詞から順に, 品詞 1, 品詞 2, 品詞 3, 品詞 4 と呼ぶ.

(qw) 質問中の単語の n -gram ($1 \leq n \leq N$, n は整数) の列挙 (例: 「首相は誰」に対し, $N=2$ の場合, 「qw:首相,qw:は,qw:誰,qw:首相は,qw:は誰」を特徴とする)

(qq) 質問中の疑問詞 (「誰」「どこ」「何」「いつ」等)

(qm1) 質問中の単語の品詞 1 の異なりの列挙 (例: 「首相は誰」に対し, 「qm1:名詞,qm1:助詞」を特徴とする)

(qm2) 質問中の単語の品詞 2 の異なりの列挙

(qm3) 質問中の単語の品詞 3 の異なりの列挙

(qm4) 質問中の単語の品詞 4 の異なりの列挙

本報告では, qw については, 4-gram まで作成した.

3.1.2 文書特徴集合

文書特徴集合 (document feature set) は, 文書のみから得られる特徴である.

各単語 w_i について, 以下の各特徴を抽出する.

(dw-K, ..., dw+0, ..., dw+K) 単語 w_i の前後 K 単語の出現形

(dm1-K, ..., dm1+0, ..., dm1+K) 単語 w_i の前後 K 単語の品詞 1

(dm2-K, ..., dm2+0, ..., dm2+K) 単語 w_i の前後 K 単語の品詞 2

(dm3-K,...,dm3+0,...,dm3+K) 単語 w_i の前後 K 単語の品詞 3

(dm4-K,...,dm4+0,...,dm4+K) 単語 w_i の前後 K 単語の品詞 4

3.1.3 組み合わせ特徴集合

組み合わせ特徴集合 (combined feature set) は、文書のみから得られる特徴である。

各単語 w_i について、作成される特徴集合の各特徴は以下の通りである。

(cw-K,...,cw+0,...,cw+K) 質問文のいずれかの単語と単語 w_i の前後 K 単語の出現形一致の有無

(cm1-K,...,cm1+0,...,cm1+K) 質問文のいずれかの単語と単語 w_i の前後 K 単語の品詞 1 の一致の有無

(cm2-K,...,cm2+0,...,cm2+K) 質問文のいずれかの単語と単語 w_i の前後 K 単語の品詞 2 の一致の有無

(cm3-K,...,cm3+0,...,cm3+K) 質問文のいずれかの単語と単語 w_i の前後 K 単語の品詞 3 の一致の有無

(cm4-K,...,cm4+0,...,cm4+K) 質問文のいずれかの単語と単語 w_i の前後 K 単語の品詞 4 の一致の有無

(cq-K,...,cq+0,...,cq+K) 質問文の疑問詞と単語 w_i の前後 K 単語の組み合わせ (例: cq+1: 誰&さん)

3.2 構築と実行

構築フェーズにおいては、CRL QA データから生成されたデータ $(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})$ を用いて確率モデルを推定し、実行フェーズでは、そのモデルを用いて、入力 $x^{(i)}$ のクラスラベル $y^{(i)}$ を計算する。

構築フェーズ

1. 質問・正解として、質問 q 、正解 a 、記事 d が与えられたとする。
2. d 中の正解 a の前後にタグ $\langle a \rangle \langle /a \rangle$ を付ける。
3. d を形態素解析する。

4. $d = w_1, \dots, \langle a \rangle, w_j, \dots, w_k, \langle /a \rangle, \dots, w_m$ とすると、 $w_1, \dots, w_m$ について、特徴を抽出したものをそれぞれ $x^{(1)}, \dots, x^{(m)}$ とする。

5. クラスラベル $y^{(i)}$ は、 $\langle a \rangle \langle /a \rangle$ に挟まれている単語列の先頭に B、後続に I を与え、それ以外は 0 とする。

6. $(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})$ をデータとして、最大エントロピー法により、確率モデル p_{λ^*} を求める。

実行フェーズにおいては、検索された各記事から、解答候補を取り出す。

実行フェーズ

1. 質問 q 、記事 d が与えられたとする。
2. d を形態素解析する。
3. $d = w_1, \dots, w_m$ とすると、各 w_i について、特徴を抽出したものをそれぞれ $x^{(i)}$ とする。
4. 各ラベル $y^{(j)} \in \mathcal{Y}$ について、 $p_{\lambda^*}(y^{(j)} | x^{(i)})$ を計算することにより、各 $x^{(i)}$ に対するラベル $y^{(j)}$ の確率を求める。
5. 各 $x^{(i)}$ について、最も確率の大きい $y^{(j)}$ をそれぞれの単語 w_i のラベルとする。
6. ラベル付けされた単語列について、B から始まり、I または B が続く単語列を解答の候補とする。解答候補のスコアとして、先頭の単語の B ラベルの確率を用いる。
7. スコアの大きい上位 5 位までを解答とする。

本アプローチでは、正解部分のみを切り出す学習をするため、解答を 5 位まで出力する場合は、抽出の範囲を広げておく必要がある。そこで、実行フェーズにおいては、O ラベルの確率が 99% 以上の場合を O ラベルとし、それ以外は、B または I ラベルのいずれかを確率値に従って与えた。

また、実行フェーズにおいて、B から始まり I が続く単語列を解答候補とするだけでなく、B から始まり途中 B が続く単語列も解答候補としている。その理由は、一般の用語抽出とは異なり、質問の解答となる文字列を取り出す学習が行われた場合、ある質問の解答の候補が 2 つ連続して現れることは少なく、一連の単語列として扱う方が良いことが予備実験の結果判明したためである。

表 2: 10 分割交差検定による実験結果

	質問数	1 位	2 位	3 位	4 位	5 位	MRR	TOP5
完全一致	2000	378	175	82	45	40	0.26	0.36
部分一致	2000	579	274	143	76	75	0.40	0.57
人手評価	2000	512	232	123	68	53	0.35	0.49

4 評価

CRL QA データセット 2000 問の質問文・正解を 10 のセットに分割し、10 分割交差検定 (10-fold cross validation) を行なった。システムの最終的な出力結果として得られた解答を、標準的に用いられる次の 2 つの評価値により評価した。

Top5 スコア 5 位以内に正解が含まれた質問の割合。

MRR (Mean Reciprocal Rank) は各質問について、ランクの 1 位から 5 位まで順に正解かどうかをチェックしていき、最初に正解と判定されたランク n のポイント $1/n$ を与え、質問数で平均したもの。

正解の判定については、「完全文字列一致」と「文字列の包含」の 2 つの基準による自動評価および「人手評価」を行なった。

全体の評価結果を表 2 に示す。それぞれの評価法について、1 位～5 位での正解数、MRR 値、TOP5 値を表している。本報告で述べた構成法により、人手による評価で、全体で MRR=0.35、TOP5=50% の質問応答が実現できることが確認された。

5 考察

本報告のアプローチでは、質問タイプの体系を必要としないため、システムの構築は非常にシンプルであるが、MRR = 0.35、TOP5 = 50% が得られた。この精度は、評価用のデータセットは異なるが、人名、地名等の 8 種類の固有表現のみを対象にした SAIQA-II [10] の MRR = 0.4、TOP5 = 55% に近い精度である。

部分一致により正解と判定された解答は、正解例が想定していなかったが、正解と判定されたものがほとんどであった。例えば、「～は何倍ですか」「2」（想定正解は「2 倍」）のように解答が短い場合や、「～は誰ですか」「伊藤仙二さん」（想定正解は「伊藤仙二」）のように正解で想定していない敬

称が付加されている。このような過不足は、本報告では、固有表現といった固定的な単位を想定せず、質問の正解として与えられた単位を自由に切り出す学習をしているため必然的に起きるものである。また、取り出される文字列の切れ目は、ほとんどの場合、通常の固有表現や名詞句の切れ目となっていて、2000 問のデータからの学習でも用語の切り出しには 98% 程度という予想したよりも高い精度で成功していた。

一方、切り出された用語が解答候補として、順序でランキングされているかという点では、さらなる改良の必要性を感じた。現在は、文書検索にて、上位 1 位の文書のみから解答を探しているが、上位 5 位まで文書から解答を探してみたところ、正解率が半分近くまで低下した。これは、正解を含む文から解答の切り出しを学習しているため、正解を含まない文に対する学習が十分に行われていないためであると考えられる。

本報告における学習による構築、実行フェーズでは、CRL QA データに付与されている質問タイプの情報は一切利用していないが、参考のために、各質問タイプに対する自動評価の正解率を表 3 に示す。T5 は Top5、ダッシュ() は部分一致を表す。質問タイプを用いないで、学習を行った結果、海洋名や絵画名、期間など、1 件しか質問文を含まないマイナーな質問タイプに属する質問文にも正解できていることは興味深い。

関連研究として、文書検索以外の QA コンポーネントを統合する研究には、Echihabi ら [2] および Lita ら [5] の研究がある。Echihabi らは、Noisy-Channel Model により、QA システムを構築しているが解答の候補となる文字列の切り出し範囲は学習により決定されるのではなく、デコーダが解答タイプや統語タイプに従って切り出している。Lita らは、質問をクラスタリングすることにより、質問タイプを用いることなく、質問文と解答から QA システムを構築しているが、正解を含む短いパッセージ (snippet) を回答するシステムであり、本報告のよ

うに正解そのものを用語抽出として実現したものではない。

6 おわりに

本報告は、質問文およびその正解例のみから、質問に対する解答法を学習するという、新しい QA システムの構築法について述べた。我々の知る限り、質問タイプを用いることなく、質問文にバイアスされた用語抽出 (QBTE) 問題として、文書から直接解答を抽出する方法を総合的に学習する QA システムの研究はこれまで示されていない。最大エントロピー法を用いたシステムを実装し、システム全体を 2000 問の質問・正解データにより学習させた時の性能を交差検定により測定した。その結果、完全一致による正解評価でも MRR が 0.26、5 位以内正解率が平均約 36%、人手評価では、MRR が 0.35、5 位以内正解率が平均約 50% であることが明らかになった。特徴抽出の改良による、さらなる精度の向上が今後の課題である。

謝辞

CRL QA データを使わせていただいた NYU の関根聡氏および NICT の井佐原均氏と最大エントロピー法の学習ツールを使わせていただいた NICT の内山将夫氏に感謝いたします。

参考文献

- [1] Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra: A Maximum Entropy Approach to Natural Language Processing, *Computational Linguistics*, Vol. 22, No. 1, pp. 39–71 (1996).
- [2] Abdessamad Echihabi and Daniel Marcu: A Noisy-Channel Approach to Question Answering, *Proc. of ACL-2003*, pp. 16–23 (2003).
- [3] Abraham Ittycheriah, Martin Franz, Wei-Jing Zhu, and Adwait Ratnaparkhi: Question Answering Using Maximum-Entropy Components, *Proc. of NAACL-2001* (2001).
- [4] Abraham Ittycheriah, Martin Franz, Wei-Jing Zhu, and Adwait Ratnaparkhi: IBM’s Statistical Question Answering System – TREC-10, *Proc. of TREC-10* (2001).
- [5] Lucian Vlad Lita and Jaime Carbonell: Instance-Based Question Answering: A Data-Driven Approach: *Proc. of EMNLP-2004*, pp. 396–403 (2004).
- [6] Hwee T. Ng and Jennifer L. P. Kwan and Yiyuan Xia: Question Answering Using a Large Text Database: A Machine Learning Approach: *Proc. of EMNLP-2001*, pp. 67–73 (2001).
- [7] Marisu A. Pasca and Sanda M. Harabagiu: High Performance Question/Answering, *Proc. of SIGIR-2001*, pp. 366–374 (2001).
- [8] Lance A. Ramshaw and Mitchell P. Marcus: Text Chunking using Transformation-Based Learning, *Proc. of WVLC-95*, pp. 82–94 (1995).
- [9] Erik F. Tjong Kim Sang: Noun Phrase Recognition by System Combination, *Proc. of NAACL-2000*, pp. 55–55 (2000).
- [10] 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).
- [11] Jun Suzuki, Yutaka Sasaki, and Eisaku Maeda: SVM Answer Selection for Open-Domain Question Answering, *Proc. of Coling-2002*, pp. 974–980 (2002).
- [12] Jun Suzuki, Hirotoshi Taira, Yutaka Sasaki, and Eisaku Maeda: Directed Acyclic Graph Kernel, *Proc. of ACL 2003 Workshop on Multilingual Summarization and Question Answering - Machine Learning and Beyond*, pp. 61–68, Sapporo (2003).
- [13] 鈴木 潤, 佐々木 裕, 前田 英作: 単語属性 N-gram と統計的機械学習による質問タイプ同定, 情報処理学会論文誌, Vol. 44, No. 11, pp. 2839–2853 (2003).
- [14] Ingrid Zukerman and Eric Horvitz: Using Machine Learning Techniques to Interpret WH-Questions, *Proc. of ACL-2001*, Toulouse, France, pp. 547–554 (2001).

表 3: 質問タイプ毎の評価結果 (参考)

質問タイプ	#Q	MRR	T5	MRR'	T5'
G O E	36	0.29	0.36	0.39	0.56
G P E	4	0.62	0.75	1.00	1.25
イベント数	7	0.48	0.71	0.48	0.71
イベント名	19	0.20	0.26	0.36	0.42
グループ名	74	0.22	0.26	0.38	0.53
スポーツチーム名	15	0.19	0.27	0.29	0.40
テレビ番組名	1	0.00	0.00	0.00	0.00
ポイント	2	0.00	0.00	0.00	0.00
医薬品名	2	0.00	0.00	0.00	0.00
宇宙船名	4	0.38	0.75	0.38	0.75
運動行為名	18	0.18	0.39	0.36	0.61
映画名	6	0.33	0.33	0.38	0.50
音楽名	8	0.07	0.25	0.07	0.25
河川湖沼名	3	0.11	0.33	0.11	0.33
会議名	17	0.21	0.24	0.50	0.71
海洋名	1	1.00	1.00	1.00	1.00
絵画名	1	0.50	1.00	0.50	1.00
学校名	21	0.19	0.24	0.31	0.43
学問名	5	0.10	0.20	0.14	0.40
割合表現	47	0.34	0.43	0.48	0.62
企業名	77	0.35	0.48	0.45	0.65
期間	1	1.00	1.00	1.00	1.00
規則名	35	0.21	0.31	0.32	0.51
記念碑名	2	0.50	0.50	1.00	1.00
競技名	9	0.17	0.33	0.31	0.56
協会名	26	0.20	0.27	0.38	0.62
金額表現	110	0.43	0.52	0.58	0.72
空港名	4	0.30	0.50	0.51	1.00
軍隊名	4	0.08	0.25	0.08	0.25
芸術名	4	0.50	0.50	0.56	0.75
月期間	6	0.08	0.17	0.08	0.17
言語名	3	0.33	0.33	0.33	0.33
個数	10	0.47	0.60	0.68	0.90
娯楽施設	2	0.00	0.00	0.00	0.00
公園名	1	0.00	0.00	0.00	0.00
公演名	3	0.50	0.67	0.83	1.00
公共機関名	19	0.18	0.26	0.35	0.53
港名	3	0.17	0.33	0.17	0.33
国数	8	0.50	0.50	0.54	0.62
国籍名	4	0.12	0.25	0.12	0.25
国名	84	0.41	0.58	0.50	0.73
罪名	9	0.17	0.22	0.39	0.44
市区町村名	72	0.30	0.46	0.38	0.58
施設数	4	0.25	0.25	0.25	0.25
施設名	11	0.20	0.27	0.34	0.64
時間	3	0.00	0.00	0.33	0.67
時間表現	2	0.00	0.00	0.50	0.50
時刻期間	8	0.21	0.38	0.38	0.62
時刻表現	13	0.08	0.08	0.10	0.15
時代表現	3	0.11	0.33	0.44	0.67
自然現象名	5	0.30	0.40	0.34	0.60
自然災害名	4	0.38	0.50	0.42	0.75
自然物名	5	0.27	0.60	0.27	0.60
車名	1	0.00	0.00	0.00	0.00
宗教名	5	0.40	0.40	0.60	0.60
週期間	4	0.00	0.00	0.62	0.75
重量	12	0.29	0.33	0.38	0.58

質問タイプ	#Q	MRR	T5	MRR'	T5'
出版物名	6	0.00	0.00	0.21	0.33
順位表現	7	0.25	0.43	0.46	0.71
書籍名	6	0.50	0.50	0.50	0.50
賞名	9	0.29	0.56	0.29	0.56
場所数	2	0.00	0.00	0.00	0.00
植物名	10	0.24	0.50	0.24	0.50
色	5	0.00	0.00	0.00	0.00
新聞名	7	0.43	0.57	0.43	0.57
神社寺名	8	0.27	0.50	0.30	0.62
震度	1	0.00	0.00	0.33	1.00
人数	72	0.26	0.43	0.29	0.50
人名	282	0.18	0.25	0.44	0.60
数値表現	19	0.03	0.11	0.18	0.32
寸法表現	1	0.00	0.00	0.00	0.00
政治的組織名	3	0.33	0.33	0.67	0.67
政党名	37	0.33	0.46	0.54	0.68
政府組織名	37	0.29	0.46	0.32	0.49
製品数	41	0.33	0.41	0.41	0.54
製品名	58	0.26	0.36	0.38	0.52
戦争名	2	0.25	0.50	0.42	1.00
船名	7	0.26	0.43	0.35	0.86
組織数	20	0.12	0.15	0.24	0.30
組織名	23	0.17	0.22	0.36	0.48
速度	1	0.00	0.00	1.00	1.00
体積	5	0.30	0.40	0.50	0.60
大会名	8	0.19	0.25	0.44	0.62
地位名	39	0.08	0.15	0.15	0.26
地域名	22	0.17	0.27	0.32	0.50
地形名	3	0.33	0.33	0.33	0.33
地名	2	0.12	0.50	0.62	1.00
長さ	22	0.11	0.14	0.27	0.41
通貨名	1	0.00	0.00	0.00	0.00
電車站名	3	0.17	0.33	0.17	0.33
電車路線名	1	0.00	0.00	0.33	1.00
電話番号	1	0.00	0.00	0.00	0.00
都道府県州名	36	0.33	0.42	0.59	0.75
動物数	3	0.11	0.33	0.44	0.67
動物名	10	0.24	0.50	0.24	0.50
道路名	1	0.00	0.00	0.00	0.00
日数期間	9	0.04	0.11	0.30	0.44
日付表現	130	0.26	0.40	0.39	0.60
年期間	34	0.29	0.38	0.58	0.79
年齢	22	0.41	0.59	0.46	0.73
倍数表現	9	0.58	0.78	0.66	1.00
犯罪名	4	0.62	0.75	0.71	1.00
飛行機名	2	0.10	0.50	0.10	0.50
美術博物館名	3	0.00	0.00	0.00	0.00
病気名	18	0.15	0.22	0.31	0.50
頻度表現	13	0.18	0.31	0.19	0.38
武器名	1	0.00	0.00	0.00	0.00
物質名	18	0.31	0.39	0.40	0.50
方式制度名	29	0.18	0.24	0.37	0.52
民族名	3	0.00	0.00	0.57	1.00
名前	5	0.25	0.40	0.25	0.40
面積	4	0.25	0.25	0.30	0.50
理論名	1	0.00	0.00	0.00	0.00
陸上地形名	5	0.51	0.80	0.51	0.80
列車名	2	0.17	0.50	0.17	0.50
計	2000	0.26	0.36	0.40	0.57