

# 大学入試問題における下線部の適応的利用に基づく自動解答手法

中田 涼介<sup>†</sup> 木村 輔<sup>††</sup> 宮森 恒<sup>††</sup>

<sup>†</sup> 京都産業大学 コンピュータ理工学部 ネットワークメディア学科

<sup>††</sup> 京都産業大学 大学院 先端情報学研究科

〒 603-8555 京都府京都市北区上賀茂本山

E-mail: †{g1244820,i1458030,miya}@cse.kyoto-su.ac.jp

**あらまし** 近年、ユーザから与えられた自然言語での質問に対し、情報要求を満たして正確に返答する質問応答システムが研究開発されている。特に、NTCIR-12 QA Lab-2 Task では、大学入試における世界史問題の解答を目的とした質問応答システムの開発が進められている。設問に正答するためには、要点を正確に捉える必要がある。小問には度々「下線部①に関連して」といった記述が含まれるが、下線部のテキストが解答生成に必要な場合と不要な場合がある。従来研究では、下線部の適応的利用による自動解答手法の研究はほとんど行われていない。本稿では、下線部、小問、選択肢から特徴を抽出し、下線部を回答候補取得時のクエリとして利用するか否かを推定する識別器を構築し、実験において、下線部の適応的利用と正答率との関係について明らかにする。

**キーワード** 質問応答, 大学入試問題, 自動解答, 下線部, 機械学習, クエリ生成

## 1. はじめに

近年、ユーザから与えられた自然言語での質問に対し、情報要求を満たした正確な回答を返す質問応答システムが研究開発されている。特に、高度な質問応答システムの開発を目指す評価型ワークショップである NTCIR-12 QA Lab-2 Task [1] では、より現実世界に即した質問応答への第一歩として、大学入試における世界史問題の解答を目的とした質問応答システムの開発が進められている。

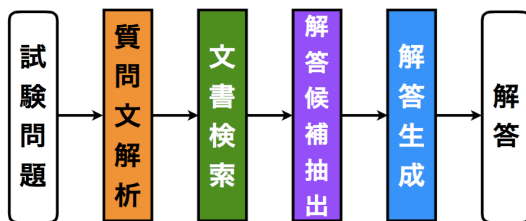


図1 大学入試における世界史問題の解答を目的とした質問応答システムの基本的な処理手順

大学入試における世界史問題の解答を目的とした質問応答システムの基本的な処理手順を図1に示す。XML形式の試験問題データが入力として与えられ、「質問文解析」、「文書検索」、「解答候補抽出」、「解答生成」の4つのモジュールによる処理が順に実行され、解答が出力される。

一般的な世界史の試験問題は、いくつかの大問で構成される。各大問は、その冒頭に提示される問題の文脈を表す文章（以下、文脈部と呼ぶ）と、いくつかの小問から構成される。解答者は、文脈部や各小問の内容、選択肢を適宜参照しながら解答する。小問には、度々「下線部①に関連して」といった記述が含まれるが、下線部の内容が問題を解くために必要不可欠な問題もあ

れば、直接的には不要とみなせる問題も存在する。従来研究では、このような下線部の要不要を適応的に利用することがどの程度正しい解答に結びつくのかについては明らかとはされていない。そこで、本稿では、下線部の要不要を判断する識別器を構築し、その結果を、解答候補を含む文書を検索するためのクエリ生成に利用した際の正答率との関係について明らかにする。

本稿の構成は以下の通りである。2章では、関連研究を紹介し、3章では、提案手法の詳細を説明する。4章では、提案手法と正答率との関係を確かめる実験を行い、5章で考察を述べる。最後に、6章で、全体と今後の課題をまとめる。

## 2. 関連研究

質問応答システムに関する研究は、これまでも数多く行われている。

質問応答システムは、単純な事実や出来事について回答するファクトイド型質問応答システムと、事物間に存在する関係や一連の出来事、理由等について回答するノンファクトイド型質問応答システムに分類できる。近年は、大規模な知識源が利用できるようになったため、特定分野に限定されない質問に回答可能な、オープンドメイン質問応答システムの研究が活発に行われている。佐々木ら[2]は、新聞記事をもとにしたファクトイド型質問2,000問とその正解データを用いた学習型質問応答システム「SAIQA-II」を開発し、質問解析と解答生成に機械学習を用いるシステム構成法を提案している。2,000問の質問・正解データを用いた5分割交差確認の結果、5位以内正解率で約55%の正解率を達成している。遠藤ら[3]は、固有表現を階層的に詳細化し、質問タイプ解析を詳細化している。QAC-1で作成された質問文とその答えの情報、新聞記事中に頻出するもの等から、固有表現を階層的に拡張・詳細化した結果、QAC-1から大幅にスコアが向上している。Horiら[4]は、大語彙音声

認識エンジンを用いて質問に答える、音声駆動型質問応答システム「SPIQA」を開発している。曖昧性解消のモジュールを備え、質問の曖昧な点を判定し、ユーザに問い直すシステムである。田村ら [5] は、ノンファクトイド型質問における Why 型質問を対象とした質問応答システムを試作している。質問文の言い換えと、Web 検索エンジンを用いた検索結果を利用することで、10 位以内正解率で 54% を達成している。諸岡ら [6] は、Why 型質問に加えて、Definition 型質問と How 型質問の回答を抽出するシステムを開発している。QAC4 の質問応答サンプルデータの分析に基づき、検索キーワードを抽出し、検索結果の上位 10 位のスニペットを用いて、再度記事検索することで、検索精度を向上させている。いずれのシステムも、細部の違いはあるものの、基本的に、図 1 で示した構成がとられている。

一方、大学入試問題に対する質問応答システムに関する研究も近年活発に行われている。例えば、大学入試の世界史問題の場合は、世界史に分野が限定されるという点で必ずしもオープンメイン質問応答ではないが、人名や地名といった単純な名称の回答だけでなく、適切な内容説明や理由の回答、年代の正しい並びや地図の理解、写真の理解と組み合わせられた回答が求められるという点で、従来にはない、より現実的な回答生成機能が要求される。すなわち、ファクトイド型質問やノンファクトイド型質問に、柔軟かつ適切な形式で正確に回答することが課題の一つであると考えられる。

松井ら [7] は、大学入試の穴埋め問題を解答する手法を提案している。大学入試センター試験より高難度と考えられる 2 次試験において、質問数の割合が高い Factoid 型と Blank 型の質問に焦点を当て、質問文と知識源のそれぞれでキーワードベクトルを作成し、コサイン類似度を用いて類似度を計算することで、回答を生成する手法を提案している。本稿では、以下の 2 点が異なる。

- 抽出したキーワードを「文書検索」に利用している
- 下線部を、識別器の識別結果によって適応的に利用している

木村ら [8] は、各選択肢に出現する歴史的出来事の単語が表す期間を相互比較し、回答を生成する手法を提案している。この手法では、大学入試センター試験で約 20~35% の正答率を達成している。Okita ら [9] は、専用構文解析プログラム「KBarse」を使った質問応答システム「DCUMT」を開発している。自然言語で記述された問題文を、論理演算子で表された形 (ラムダ計算式) に変換し、回答を生成している。このシステムは、大学入試センター試験に加えて、東京大学、京都大学、北海道大学、早稲田大学、中央大学の 2 次試験問題 (記述式) も解答している。その結果、センター試験、2 次試験ともに、得点が 70 点を超え、NTCIR-11<sup>(注1)</sup> QA Lab Task に参加したシステムで最もよい成績を収めている。Kano ら [10] が開発した「KJP」は、知識源に教科書データのみを用いる手法を提案している。

本稿では、大学入試センター試験の世界史問題を対象とし、その下線部の適応的利用に着目した手法を提案する。

### 3. 提案手法

#### 3.1 提案手法の概要

大学入試における世界史問題の解答を目的とした質問応答システムが正答するためには、問題中の文脈部や各小問の内容、選択肢を適宜参照し、解答を生成する必要がある。小問には、度々「下線部①に関連して」といった記述が含まれるが、下線部の内容が問題を解くために必要不可欠な問題もあれば、直接的には不要とみなせる問題も存在する。本稿では、下線部の要不要を判断する識別器を構築し、解答生成に利用する手法を提案する。本稿では、下線部の要不要の識別結果を、解答候補を含む文書を検索するためのクエリ生成に利用した場合、正答率がどのように変化するかについて明らかにする。

#### 3.2 対象データ

本稿では、大学入試センター試験の世界史の問題を用いることとする。一般的な世界史の試験問題を図 2 に示す。

第 1 問 世界史上の法をめぐる問題について述べた次の文章 A~C を読み、下の問い (問 1~9) に答えよ。(配点 25)

中国の歴代王朝は、統治のため法の整備を心掛けた。秦の始皇帝は、法家の「ア」を登用して、全国統一と支配確立を進めた。唐代には律令が整備されたが、そのうち律は、今日の「イ」に相当する内容を主とする。明清時代にも、法典の編纂・改訂が進められた。20 世紀に入ると、清は近代法典の編纂に着手した。その試みは① 中華民国の時代に次第に実を結び、欧米や日本に倣った刑法や民法が制定された。近代法典編纂の努力がなされた背景には、列強が中国の国内で行使している領事裁判権を撤廃させようとする意図があった。そして、その撤廃が実現していったのは、② 第二次世界大戦中のことである。

問 文章中の空欄「ア」と「イ」に入れる語の組合せとして正しいものを、次の①~④のうちから一つ選べ。 1

① アー李 ス イー民法  
② アー李 ス イー刑法  
③ アー荀 子 イー民法  
④ アー荀 子 イー刑法

図 2 一般的な世界史の試験問題  
(2013 年度 大学入試センター試験 第 1 問 問 1)

世界史の試験問題は、いくつかの大問で構成され、各大問は、冒頭の文脈部 (図 2 の (a)) と、いくつかの小問 (図 2 の (b)) から構成される。文脈部には数本の下線部が引かれ、解答者は、下線部、小問、選択肢 (図 2 の (c)) の内容を適宜参照しながら解答する。小問には、度々「下線部①に関連して」といった記述が含まれるが、下線部の内容が問題を解くために必要不可欠な問題もあれば、直接的には不要とみなせる問題もある。以下、下線部が解答内容の判断に必要な場合、不要な場合についてそれぞれ具体例を挙げる。

##### 3.2.1 下線部が解答内容の判断に「必要」である問題例

下線部が解答内容の判断に必要な問題例を表 1 に示す。

表 1 の問題に正答するためには、各選択肢中の「この石版」という情報のみでは不十分であり、「この石版」が「ロゼッタ石」とであると特定することが必要不可欠である。このように、小問

(注 1) : NTCIR-11 : <http://research.nii.ac.jp/ntcir/ntcir-11/>

表 1 下線部が解答内容の判断に必要である問題例  
(1997 年度 大学入試センター試験 第 2 問 問 4)

問題区分	内容
小問	下線部④について述べた次の文 ①～④ のうちから、正しいものを一つ選べ。
下線部④	ロゼッタ石
選択肢①	この石版には、上から神聖文字、民衆文字、ギリシア文字が記されている。
選択肢②	この石版には、上からベルシア文字、アラム文字、フェニキア文字が記されている。
選択肢③	この石版の碑文は、セレウコス朝シリアの勅令を伝えている。
選択肢④	この石版の碑文は、アンティゴノス朝マケドニアの勅令を伝えている。

や選択肢中には事物を特定するための具体的な記述が不足し、下線部のみにそのような記述が現れる場合、下線部は解答内容の判断に「必要」であると判断する。

### 3.2.2 下線部が解答内容の判断に「不要」である問題例 下線部が解答内容の判断に不要である問題例を表 2 に示す。

表 2 下線部が解答内容の判断に不要である問題例  
(1999 年度 大学入試センター試験 第 3 問 問 5)

問題区分	内容
小問	下線部⑤について述べた文として誤っているものを、次の ①～④ のうちから一つ選べ。
下線部⑤	修道院の活動
選択肢①	ベネディクトゥスは、モンテ=カッシーノに修道院を設立した。
選択肢②	修道院では、日常生活を信仰に捧げるために、労働は禁止された。
選択肢③	聖職売買などの教会の腐敗に対して、クリュニー修道院を中心とする改革運動が起こった。
選択肢④	フランチェスコ修道会は、托鉢修道会の一つであった。

表 2 の問題に正答するためには、各選択肢に含まれる情報のみで正誤判断することが可能であり、下線部の「修道院の活動」という情報は必ずしも必要ではない。このように、小問や選択肢中の記述で正答するための具体的な情報が十分提示されており、下線部を特に必要としない場合、下線部は解答内容の判断に「不要」であると判断する。

### 3.2.3 学習データおよびテストデータの作成

本稿では、NTCIR-12 QA Lab Task [1] で提供された以下の大学入試センター試験およびマーク模試の試験問題データを用いて学習データを作成した。

- 大学入試センター試験 (1997 年度～2009 年度、隔年 7 回分)

- ベネッセマーク模試 (2014, 2015 年実施、計 2 回分)
- 代ゼミセンターマーク模試 (2013 年実施、計 2 回分)

以上の試験問題データに対して、解答欄番号ごとに正解ラベルを手で付与した。各問題の下線部、小問、選択肢の内容を考慮し、3.2.1 節、3.2.2 節の基準に従い、下線部が解答内容の

判断に必要である場合は「あり」、不要である場合は「なし」とした。また、参照すべき下線部がない、あるいは、参照すべき下線部はあるが、画像を使って解くような問題は、正解ラベルを付与せず、データに含めないこととした。以上により、「あり」137 件、「なし」216 件の学習データを作成した。

さらに、NTCIR-12 QA Lab Task [1] で提供された、2011 年度 大学入試センター試験の問題データに対し、学習データと同様に正解ラベルを付与し、テストデータを作成した。

### 3.3 識別に用いる素性

各素性の計算式で用いる記号の定義を表 3 に示す。

記号	定義
$M_U$	下線部の形態素の集合
$M_S$	小問の形態素の集合
$N_U$	下線部の名詞の集合
$N_S$	小問の名詞の集合
$NNV_U$	下線部の名詞/固有名詞/動詞の集合
$NNV_S$	小問の名詞/固有名詞/動詞の集合
$NNV_C$	選択肢の名詞/固有名詞/動詞の集合
$NG_U$	下線部の N-gram の集合
$NG_S$	小問の N-gram の集合
$NO_C_U$	下線部の文字数
$NO_C_S$	小問の文字数
$W_U$	下線部の単語の集合
$W_S$	小問の単語の集合

#### 3.3.1 形態素重複率

下線部と小問で重複した形態素の数を、下線部と小問の少なくとも一方で出現する形態素の数で正規化する。

$$MorphemeOverlap = \frac{|M_U \cap M_S|}{|M_U \cup M_S|} \quad (1)$$

#### 3.3.2 名詞重複率

下線部と小問で重複した名詞の数を、下線部と小問の少なくとも一方で出現する形態素の数で正規化する。

$$NounOverlap = \frac{|N_U \cap N_S|}{|M_U \cup M_S|} \quad (2)$$

#### 3.3.3 cos 類似度

本稿では、類似度計算に用いる単語を名詞、固有名詞、動詞の 3 種類に限定し、学習データ中の下線部、小問、選択肢 (選択肢は小問ごとに該当する全選択肢を連結したものをを用いる) をそれぞれ文書とみなし、以下の 2 種類の方法で文書ベクトルを生成した。

- 各単語の出現頻度 (tf) を要素とする方法
- 各単語の tfidf を要素とする方法

さらに、各単語の tfidf を要素とする方法では、以下の 2 種類の方法で idf を算出した。本稿では、前者を「 $tfidf_1$ 」、後者を「 $tfidf_2$ 」と呼ぶ。

- 下線部、小問、選択肢の各文書の和集合を全文書集合として算出する方法
- 各文書の種別 (下線部、小問、選択肢のいずれか) と同一種別の文書全体のみを全文書集合として算出する方法

以上を踏まえ、以下の5つのコサイン類似度を素性として算出した。

- $Sim_1 = \text{CosineSimirality}(\overrightarrow{D_{tf,U}}, \overrightarrow{D_{tf,S}})$
- $Sim_2 = \text{CosineSimirality}(\overrightarrow{D_{tfidf_1,U}}, \overrightarrow{D_{tfidf_1,S}})$
- $Sim_3 = \text{CosineSimirality}(\overrightarrow{D_{tfidf_2,U}}, \overrightarrow{D_{tfidf_2,S}})$
- $Sim_4 = \text{CosineSimirality}(\overrightarrow{D_{tf,C}}, \overrightarrow{D_{tf,S}})$
- $Sim_5 = \text{CosineSimirality}(\overrightarrow{D_{tfidf_1,C}}, \overrightarrow{D_{tfidf_1,S}})$

ここで、

$$\text{CosineSimirality}(\vec{X}, \vec{Y}) = \frac{\vec{X} \cdot \vec{Y}}{|\vec{X}| |\vec{Y}|}$$

である。

また、 $\overrightarrow{D_{tf,U}}$  は、下線部の tf 重み付け文書ベクトルを表し、添え字の U, S, C はそれぞれ、下線部、小問、選択肢を表す。

### 3.3.4 N-gram の重複率 (N = 1, 2, 3)

下線部と小問で重複した N-gram の数を、小問の N-gram の数で正規化する。

$$\text{NgramOverlap} = \frac{|NG_U \cap NG_S|}{|NG_S|} \quad (3)$$

### 3.3.5 文字数の割合

下線部の文字数を、下線部の文字数と小問の文字数の和で正規化する。

$$\text{ProportionOfNumberOfCharacters} = \frac{NOC_U}{NOC_U + NOC_S} \quad (4)$$

### 3.3.6 Jaccard 係数

下線部と小問で共通して出現する単語の数を、下線部と小問の少なくとも一方で出現する単語の数で正規化する。

$$\text{JaccardCoefficient} = \frac{|W_U \cap W_S|}{|W_U \cup W_S|} \quad (5)$$

### 3.3.7 Simpson 係数

下線部と小問で共通して出現する単語の数を、下線部と小問のうち、単語数が少ない方の要素数で正規化する。

$$\text{SimpsonCoefficient} = \frac{|W_U \cap W_S|}{\min(|W_U|, |W_S|)} \quad (6)$$

## 3.4 識別器の構築

3.3節で示した素性を用いて識別器を構築した。下線部、小問、選択肢のそれぞれに対する形態素解析には Mecab<sup>(注2)</sup>を使用した。また、識別手法には、Weka<sup>(注3)</sup>に実装されている libSVM, SMO, RandomForest, J48 を用いた。

## 4. 実験

### 4.1 実験1：識別器の識別性能の評価

#### 4.1.1 実験目的・評価方法

本実験では、3.4節で示した識別器の識別性能がどの程度であるかを、学習データおよびテストデータを用いて明らかにする。学習データに対しては 10-fold 交差確認を行い、適合

率 (Precision), 再現率 (Recall), F 値 (F-measure), 分類精度 (Accuracy) により評価する。なお、ここでの「F 値」は、各正解ラベルの F 値の加重平均であり、重みとして、各正解ラベルの件数を用いた。

#### 4.1.2 実験結果

実験結果を表4に示す。なお、表中の「Acc」は分類精度、「Pre」は適合率、「Re」は再現率、「F1」は F 値をそれぞれ表している。

表4より、SMO および J48 による分類では、学習データ、テストデータともに、比較的高い精度を示している。また、RandomForest によるテストデータの分類では、学習データのものと比べて、精度が大きく向上していることがわかる。

### 4.2 実験2：下線部の適応的利用による質問応答システムの正答率の比較

#### 4.2.1 実験目的・評価方法

本実験では、大学入試における世界史問題の解答を目的とした質問応答システムが、提案手法による識別結果を、解答候補を含む文書を検索するためのクエリ生成に適応的に利用した際の正答率への影響について明らかにする。

実験結果は、テストデータ (2011 年度大学入試センター試験) に対する解答の正解数を、表5に示す6種類の解答タイプ毎に評価した。解答タイプは、選択肢の内容や、解答に必要な図表の有無に応じて分類される。

表5 解答タイプ一覧

解答タイプ	概要	問題数
COMBO	選択肢が穴埋め記号の組み合わせ	4
SENTENCE	選択肢が説明文	20
TERM	選択肢が人物や出来事の名称	2
TF	選択肢が正誤の組み合わせ	6
UNIQUE	選択肢が、出来事の年代順の並び	3
MAP	解答に、地図や写真等の画像が必要となるもの	1

また、下線部の要不要の識別には、テストデータに対して最も高い分類精度を示した、RandomForest による識別手法を用いることとする。

本実験で用いた質問応答システムは、CMU が開発したシステム<sup>(注4)</sup>を基に、日本語の試験問題に対応できるよう改良を加えたものである。システムは、問題読み取り部、質問文解析部、解答候補抽出部、解答候補選択部の4つから構成されている。問題読み取り部では、問題データの読み込みと形態素解析を行っている。質問文解析部では、解答タイプの推定と、文書検索に用いるクエリを生成する。解答候補抽出部では、生成されたクエリを用いて、以下の文書集合に対して全文検索する。

- Wikipedia を段落単位に文書とした文書集合
- Wikipedia や Web ページから抽出した年表データ

(注2) : Mecab : <http://taku910.github.io/mecab/>

(注3) : Weka 3: Data Mining Software in Java : <http://www.cs.waikato.ac.nz/ml/weka/>

(注4) : NTCIR-11 QA-Lab CMU Baseline :

<https://github.com/oaqa/ntcir-qalab-cmu-baseline>

表4 識別結果

データ種別	識別手法	Acc	Yes-Pre	No-Pre	Yes-Re	No-Re	F1
学習データ	libSVM	0.618	0.516 (33/64)	0.64 (185/289)	0.241 (33/137)	0.856 (185/216)	0.576
	SMO	0.7	0.585 (107/183)	0.824 (140/170)	0.781 (107/137)	0.648 (140/216)	0.703
	RandomForest	0.688	0.6 (81/135)	0.743 (162/218)	0.591 (81/137)	0.75 (162/216)	0.688
	J48	0.705	0.601 (98/163)	0.795 (151/190)	0.715 (98/137)	0.699 (151/216)	0.709
テストデータ	libSVM	0.647	0.75 (3/4)	0.633 (19/30)	0.214 (3/14)	0.95 (19/20)	0.584
	SMO	0.735	0.632 (12/19)	0.867 (13/15)	0.857 (12/14)	0.65 (13/20)	0.736
	RandomForest	0.735	0.647 (11/17)	0.824 (14/17)	0.786 (11/14)	0.7 (14/20)	0.737
	J48	0.706	0.6 (12/20)	0.857 (12/14)	0.857 (12/14)	0.6 (12/20)	0.706

解答候補選択部では、解答タイプに応じて回答の決め方を変更し、回答を決定する。解答タイプ別の回答の決め方を表6に示す。なお、解答タイプがMAPである問題には解答しないこととする。

表6 解答タイプ別の回答の決め方

解答タイプ	決め方
TERM	回答候補中の最上位1件のスコアを持つもの
SENTENCE	
COMBO	
TF	各クエリの最上位1件のスコアがある閾値を超えれば「正」、そうでなければ「誤」と判断
UNIQUE	年表データに対し、各クエリで得られた最上位1件のイベント開始時間表現を用いて並べ替え

検索方式は、OR検索とANDOR検索(AND検索かつOR検索し、文書が発見できなければOR検索のみを行う検索方式)でそれぞれスコアを出し、高い数値を最終スコアとする方式を用いる。

#### 4.2.2 実験結果

実験結果を図3に示す。

図3より、本実験で用いた質問応答システムに、識別器の識別結果を、文書検索のクエリ生成に導入したところ、試験問題全体の正解数および解答タイプ別の正解数に変化が見られないことがわかる。

### 5. 考 察

実験1の結果より、テストデータに対しては、識別手法にSMOおよびRandomForestを用いた場合で、最も高い精度を示した。そこで、この2つのケースにおいて、1つ抜き法により、各素性が与える影響を調査した。調査結果を表7に示す。なお、表中の「train(test)-all」は、学習(テスト)データに対

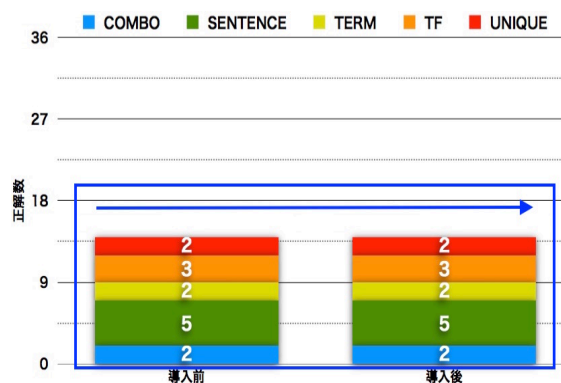


図3 テストデータに対する正解数の比較

する、すべての素性を用いた場合の分類精度、「train(test)」は、学習(テスト)データに対する、1つ抜き法を用いた場合の分類精度、「delta」は、「train(test)-all」と「train(test)」の差をそれぞれ表している。

表7より、識別手法にSMOを用いた場合、学習データに対しては、 $Sim_1$ と $Sim_5$ が分類精度向上への貢献が認められる。しかし、テストデータに対しては、どの素性を削除しても分類精度に変化が見られず、特定の素性による分類精度向上への貢献が見られなかった。

一方、識別手法にRandomForestを用いた場合は、学習データ、テストデータともに、1-gramの重複率が一貫して分類精度向上に貢献していることが認められた。

さらに、学習データにおいて、選択肢に「この」「その」といった代名詞を含む問題が、各正解ラベルでどの程度存在するのか調査した。なお、対象となる代名詞は、後に何らかの名詞句が続くと考えられる「この」「その」「あの」の3単語に限定した。本稿では、これら3単語は、ストップワードに含まれており、形態素解析の結果から除外されている。調査の結果、正解ラベル「あり」の問題では21問、「なし」の問題では8問が該当した。どちらも各正解ラベルの問題数に対し大きな割合を

表 7 各素性が与える影響

削除した素性	SMO						RandomForest					
	train-all	train	delta	test-all	test	delta	train-all	train	delta	test-all	test	delta
形態素重複率	0.7	0.703	0.003	0.735	0.735	0	0.688	0.714	0.026	0.735	0.647	-0.088
名詞重複率	0.7	0.705	0.005	0.735	0.735	0	0.688	0.703	0.015	0.735	0.676	-0.059
$Sim_1$	0.7	0.697	-0.003	0.735	0.735	0	0.688	0.714	0.026	0.735	0.676	-0.059
$Sim_2$	0.7	0.7	0	0.735	0.735	0	0.688	0.697	0.009	0.735	0.765	0.03
$Sim_3$	0.7	0.7	0	0.735	0.735	0	0.688	0.691	0.003	0.735	0.735	0
$Sim_4$	0.7	0.705	0.005	0.735	0.735	0	0.688	0.691	0.003	0.735	0.735	0
$Sim_5$	0.7	0.697	-0.003	0.735	0.735	0	0.688	0.717	0.029	0.735	0.676	-0.059
文字数の割合	0.7	0.703	0.003	0.735	0.735	0	0.688	0.725	0.037	0.735	0.735	0
1-gram 重複率	0.7	0.705	0.005	0.735	0.735	0	0.688	0.683	<b>-0.005</b>	0.735	0.676	<b>-0.059</b>
2-gram 重複率	0.7	0.711	0.011	0.735	0.735	0	0.688	0.7	0.012	0.735	0.794	0.059
3-gram 重複率	0.7	0.708	0.008	0.735	0.735	0	0.688	0.711	0.023	0.735	0.706	-0.029
Jaccard 係数	0.7	0.703	0.003	0.735	0.735	0	0.688	0.708	0.02	0.735	0.735	0
Simpson 係数	0.7	0.703	0.003	0.735	0.735	0	0.688	0.7	0.012	0.735	0.735	0

占めるものではないが、素性に組み込むことで識別精度向上に繋がる可能性がある。

実験 2 の結果より、識別器による判断が「なし」、かつ、解答タイプが「TERM」「SENTENCE」「TF」の問題に対して、文書検索時のクエリ数および最高スコアの変化を調査した。その結果、提案手法導入前後で、削減したクエリ数に関わらず、スコアが増加した場合は、増加値が一定でなく、検索上位の結果が同一であることが多い傾向が見られた。また、スコアが減少した場合は、減少値が小さく、検索上位の結果の変化を確認した。よって、識別器の識別結果を、解答候補を含む文書を検索するためのクエリ生成に適応的に利用した場合、文書検索の結果に影響を与えるが、その程度は問題によって異なることが考えられる。

## 6. まとめ

本稿では、大学入試の世界史問題における、下線部の適応的利用に基づく自動解答手法を提案した。下線部の要不要を判断する識別器を構築し、識別結果を、解答候補を含む文書を検索するためのクエリ生成に適応的に利用した際の正答率との関係性を、実験によって明らかにした。実験の結果、提案手法の導入による正答率の変化は確認できなかった。

今後は、学習データに対する分類精度の向上を目指すとともに、提案手法を解答生成時に利用する手法を検討することが課題である。

## 文 献

- [1] NTCIR-12 QA Lab-2 Task : <http://research.nii.ac.jp/qalab/index.html?locale=ja>
- [2] 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-I I", 情報処理学会論文誌 Vol.45 No.2, pp.635-646, 2004.
- [3] 遠藤 哲哉, 福本 淳一 : "詳細化された質問タイプによる質問応答システム", 情報処理学会研究報告自然言語処理 (NL) 2004-NL-159, pp.25-30, 2004.
- [4] Chiori Hori, Takaaki Hori, Hideki Isozaki, Eisaku Maeda,

Shigeru Katagiri : "Deriving disambiguous queries in a spoken interactive ODQA system", Proceedings of ICASSP-2003, Vol.1, pp.624-627, 2003.

- [5] 田村元秀, 村上仁一, 徳久雅人, 池原悟 : "Web 検索エンジンを用いた Why 型質問応答システム", 言語処理学会 第 14 回年次大会 発表論文集 (2008 年 3 月), pp.1021-1024, 2008.
- [6] 諸岡 心, 福本 淳一 : "非 Factoid 型質問に対応した質問応答システム", 日本言語処理学会 第 13 回年次大会 発表論文集, pp.958-961, 2007.
- [7] 松井兵庫, 阪本浩太郎, 松永詠介, 神貴久, 渋谷英潔, 石下円香, 森辰則, 神門典子 : "大学入試の穴埋め問題を解く質問応答システムの検討", 言語処理学会 第 21 回年次大会 発表論文集 (2015 年 3 月), pp.175-178, 2015.
- [8] Yasutomo Kimura, Fumitoshi Ashihara, Arnaud Jordan, Keiichi Takamaru, Yuzu Uchida, Hokuto Ototake, Hideyuki Shibuki, Michel Ptaszynski, Rafal Rzepka, Fumito Masui, Kenji Araki : "Using Time Periods Comparison for Eliminating Chronological Discrepancies between Question and Answer Candidates at QALab NTCIR11 Task", In NTCIR-11 QALab, 2014.
- [9] Tsuyoshi Okita, Qun Liu : "The Question Answering System of DCUMT in NTCIR-11 QA Lab", In NTCIR-11, 2014.
- [10] Yoshinobu Kano : "Solving History Exam by Keyword Distribution : KJP System at NTCIR-11 QALab Task", In NTCIR-11 QALab, 2014.