

# 主題と焦点の同定に基づく『世界史』小論述問題の自動解答

高田 拓真 土居 裕典 松崎 拓也 佐藤 理史

名古屋大学大学院 工学研究科 電子情報システム専攻

## 1 はじめに

本研究は、大学入試問題における世界史論述問題のうち、特に小論述と呼ばれるタイプの問題に対する自動解答を目標とする。世界史論述問題には大きく分けて数百字で答える大論述問題と、数十字で答える小論述問題がある。大論述問題では、解答に含めることを要求される指定語句が6~8程度与えられることが多く、これらは解答の自動生成における大きな手がかりになる。一方、小論述問題では指定語句が与えられない場合が多い。そのため、小論述問題に対する自動解答では、問題文を解析し解答すべき内容を定め、それを短い字数にまとめることが必要となる。よって、小論述問題に挑むことで、質問文解析技術、解答に含めるべき内容を知識源から検索・抽出する技術、制限字数に収めるための圧縮技術の向上に繋がると言える。

また、世界史小論述問題は、従来の質問応答で着目されてきたものとは異なるタイプの質問を多く含む。具体的には、従来は「初代江戸幕府征夷大將軍は誰?」といった factoid 型の質問や、「日本が開国したのはなぜですか?」といった why 型の質問などが主な研究課題であった。しかし、世界史小論述問題では「北イタリアに結成された都市同盟について 60 字以内で答えなさい」といった問われている対象の名称（ここでは「ロンバルディア同盟」）が明らかでない問題や、「ポリスの形成過程を 60 字以内で答えなさい」といった問われている対象のある特定の側面（ここでは「形成過程」）を答える問題がある。本文献では、問題文で問われている対象の名称を「主題」、さらにその主題に関し何が問われているかを「焦点」と呼ぶ。

世界史小論述問題の自動解答に関する先行研究 [1] では、問題文との表層類似度に基づき教科書から文を抽出することで自動解答を行なっている。しかし、前述したように問題の主題が表層的に明らかでない問題では、表層類似度が低くなるため正解となり得る文を抽出できないという課題があった。本稿では、この課題に対応するために、まず問題の主題と焦点を同定し、主題に関する文のうち問題の焦点にマッチしたものを

キリスト教徒がローマ皇帝に迫害された理由を 60字以内で説明しなさい。(2013年度東大)
5世紀におけるフン族の最盛期とその後について、 60字以内で説明しなさい。(2012年度東大)

図 1: 世界史小論述問題例

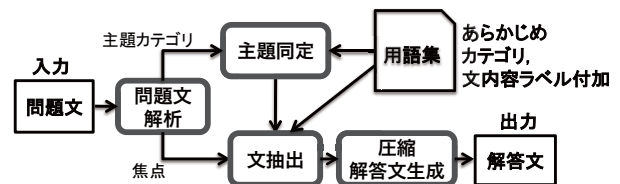


図 2: 解答システムブロック図

を抽出する手法を提案する。さらに、提案手法を用いた世界史小論述自動解答システムを作成し、入試過去問によって評価・分析を行なった結果を報告する。

システムの評価は、解答文の人手による採点に加え、文抽出結果に対して平均逆順位 (MRR) を用いた評価を行い先行研究と比較した。また、出力結果を分析し、現在のシステムの問題点について考察した。

## 2 世界史小論述自動解答システム

本システムでは、先行研究と同様に、知識源から文章を抽出し繋ぎ合わせて解答を行う。世界史小論述問題の例を図 1 に示す。前節で述べた先行研究の問題点を解決するため、まず、主題を同定する。現在のシステムでは、山川出版世界史用語集の見出し語を同定すべき主題の候補としている。また、問題文から焦点を同定し、同定した主題を見出し語とする用語集の項目から問題の焦点と一致する文を抽出し、それらを圧縮することで、解答文生成を行う。以下では、手法の詳細を記述する。

### 2.1 システム概要

図 2 にシステムブロック図を示す。システムは問題文解析、主題同定、文抽出、圧縮・解答文生成の 4 ブ

ロックから成る。また、世界史に関する知識源として世界史イベントオントロジー (EVT)[2] および山川出版の世界史用語集を用いた。

### 2.1.1 問題文解析

本システムでは、まず与えられた問題文を解析し、主題のカテゴリと問題の焦点を明らかにする。具体的には、構文解析器 knp を用いて「述べよ」、「説明せよ」などの特定のフレーズに係る名詞や動詞を抽出する。これらの名詞や動詞から、あらかじめ作成した辞書を用いて、主題のカテゴリ及び焦点を同定する。

主題のカテゴリについては、EVT 内で定義されている概念クラスに、「時代・時間」、「Role」、「その他」の3種を加えた23種のうち、問題文から前述のように抽出した名詞が当てはまるものを全て抽出する。どのカテゴリが当てはまるかは、あらかじめ作成した、各カテゴリごとの名詞辞書を用いる。辞書内の名詞は、EVT 内で定義されている用語に関する Wikipedia のページの冒頭の文から抽出し、人手で修正を行った。例えば「エラステネス」という用語は EVT 内で“人物”に分類されている。その Wikipedia ページから「学者」、「館長」などの名詞を抽出し、それらが“人物”を表す名詞とする。

各焦点を表す名詞・動詞が記載された辞書は人手で作成した。本システムで用いた焦点を表1に示す。本研究では、問題文から上述のように抽出した名詞・動詞に対し辞書で対応付けられた焦点は全て出力するようにした。また、辞書を用いた焦点の特定ができなかった場合は、「内容」及び「活動」を焦点とする。

以上の手法により同定された主題カテゴリと焦点は、それぞれ次の主題同定と文抽出に入力される。

### 2.1.2 主題同定

用語集を検索し、主題の同定を行う。具体的には、問題文と用語集の各項目との名詞一致数をスコアとし、最大スコアを与える項目名を主題として同定する。また、EVT から世界史用語に関する同義語辞書を作成し、スコア計算の際に同義語辞書に含まれる名詞は辞書中で定めた正規形に統一した。

スコアリングの際、問題文解析で出力された主題カテゴリに属する用語のみを対象とする。また、問題文から時間情報と地理情報を抽出し、検索対象の用語の時間情報 [1] と地理情報それぞれに対し合致するか判定を行い、合致しないものは除去する。

表 1: 焦点一覧

焦点	内容	理由	結果	確立	過程
焦点を表す名詞例	内容	理由, 誘因	結果	確立, 経緯	過程, 経緯
焦点	変化	情勢	特徴	活動	役割
焦点を表す名詞例	変化, 変遷	情勢, 動向	特徴, 特質	活動	役割

以上のようにスコアリングされた用語の内、スコアが最大のものを全てを主題として同定する。ただし、問題文解析で得た名詞が用語集の見出し語である場合は、上記の処理は行わずその名詞そのものを主題とする。

### 2.1.3 文抽出

同定された主題に対する用語集の項目から、問題文解析で同定した焦点と内容が一致する文を抽出する。具体的には、用語集の各説明文にはあらかじめルールベースで文内容ラベルを割り当てておき、文内容ラベルと焦点が一致した文全てを出力する。表2に文内容ラベル割り当てルールの例を示す。表中の cat は用語集の見出し語に与えられているカテゴリである。また、判別に用いている動詞については、EVT に記載されている動詞群を参考に人手で作成した。例えば、表中における「結果を表す動詞」には「奪う、略奪する、敗北する、負ける」などがある。

## 2.2 圧縮・解答文生成

最後に抽出された文集合をルールベースで圧縮し、字数制限に収まる解答文を生成する。解答文生成の大きな流れを以下に示す。

1. 抽出された文集合の各文に、問題文と共通する名詞の個数でスコアを与え、スコアの降順に並べる
2. スコア順の文リストから解答候補文のリストをつくる
3. 同義語辞書を用い、文中の単語をより短い同義語で置き換える
4. 解答候補文リストの先頭の文を、5つのルールで圧縮する。圧縮の途中で字数制限以下となった時点でそれを解答とする
5. 全てのルールで圧縮しても字数制限以下に収まらなければ、解答候補文リストの次の文を圧縮する

スコアの降順に並んだ文を  $s_1, s_2, \dots, s_n$  としたとき、 $s_1 + s_2, s_1, s_2, \dots, s_n$  を解答候補文のリストとした。上記の  $s_1 + s_2$  というのは、 $s_1$  の後に  $s_2$  を結合したもので、もし、 $s_1$  が字数制限よりも短い文だと  $s_1$  のみの解答となり、解答に含まれる情報が少なくなってしまうため、 $s_1$  よりも先に解答候補文とした。

表 2: 文内容ラベル割り当てルール例

ラベル	ルール
理由	/理由/誘因/きっかけ/背景/そのため/ が現れる
特徴	/特徴/側面/特色/性格/ が現れる
活動	(cat ∈ { 人物, 組織, 国家・王朝 })&活動を表す動詞が現れる
確立	(cat ∈ { 組織, 国家・王朝, 政策・体制, 社会制度, 思想, 美術様式, 文化・文明, 宗教 })&確立を表す動詞が現れる
結果	/結果/が現れる or (cat=戦争・戦闘&結果を表す動詞が現れる)

圧縮のルールには、接続詞や連帯修飾節を削除するものなどがある。また、一般に世界史小論述において不必要なものから削除するように、ルールに優先順位をつけた。以上を指定字数に収まる解答が得られるまで、全解答候補文で繰り返す。解答候補文リストの最後の文でも解答が得られない場合は解答は無しとする。

### 3 評価実験

東京大学の過去問を入力とし、出力の人手による採点及び抽出された文のランキングによる自動評価により先行研究 [1] のシステムによる結果と比較、評価した。以下ではその詳細について述べる。

#### 3.1 評価データ・評価方法・環境

1989～2003 年度の東京大学入学試験世界史のうち 30～150 字で答える小論述形式のもの 30 問を評価データとした。解答文の人手による採点は、駿台文庫「24 年徹底分析テーマ別東大世界史論述問題集」に記載されている加点ポイントに従って行った。加点ポイントとは、加点の対象となる内容のことで、例えば「アッパース朝が衰えた」、「清朝は義和団弾圧に転じた」といった項目が問題ごとに列挙される形で与えられている。採点基準として、各加点ポイントと解答文が表層的に類似、もしくは同義であれば、点を加算するようにした。また、主題同定部及び文抽出部の有効性を測るため、文抽出で出力された文の内、上位 5 文を MRR により評価した。正解とする文の選択については、人手による採点と同様に加点ポイントから判断した。加点されうる文を正解とした。先行研究のシステムは知識源として教科書を用いていたが、提案手法と揃えるため、用語集の項目を 1 文区切りにし、主語の省略を補うため、各文に「〈見出し語〉は」を文頭に付け加えたものを知識源として使用した。

用語集およびシステム出力の形態素解析には世界史辞書 [3] を追加した Juman を用いた。また構文解析器には knp を用いた。

表 3: 採点結果-獲得した点数ごとの問題数

手法 \ 獲得点数・MRR	0	1	2	3	MRR
先行研究 [1]	24	5	1	0	0.187
提案手法	19	5	5	1	0.381

表 4: 配点ごとの平均点

配点	2	3	4	5	6	7
先行研究 [1]	0	0	0.50	0.23	0	0.67
提案手法	0	0.67	0.75	1.38	0	1.33
問題数	2	3	4	13	5	3

#### 3.2 結果

人手による採点と MRR の結果を表 3、配点ごとの平均点を表 4 に示す。表から先行研究に比べて点を取れた問題が増え、獲得点数が多い問題数が増えていることが分かる。MRR の値も提案手法が先行研究を上回った。主題及び焦点の同定によって文抽出の性能が上がり、MRR が上昇したことが、獲得点数つまりシステム性能の向上に繋がったと言える。表 4 から、点が獲得できなかった 2 点満点、6 点満点の問題以外は平均点が上昇しており性能が向上していると言える。一方、配点に対して獲得点数は非常に低く、現在のシステムの限界も見て取れる。

### 4 分析・考察

前節で示した評価結果では先行研究を上回る結果を得て、提案手法の有効性を示した。しかし、未だ全体の 3 分の 2 の問題に対しては点数を獲得することができていない。そこで、現在のシステムにおける問題点について考察した。以下にその詳細を記述する。

#### 4.1 複数主題の比較・関係などを問う問題

提案手法では、複数主題を同定することは想定しており、2.1.2 節で述べたようにスコアが最大のもの全てを出力することで複数の主題を同定する。そして、問題は各主題について述べられた文により解答できる



ことを仮定している。しかし、評価データには、上記の仮定では正解できない問題、具体的には各主題について述べられた文内容の比較や関係性の把握が必要な問題が計7問含まれていた。以下に問題例と、提案手法の出力及び正解例を示す。

**問題** ムガル帝国とオスマン帝国には、それぞれジャーギール制と、ティマール制と呼ばれる類似の制度がみられた。両者の共通の特徴を2行以内で記せ。(2002年度東大)

**出力** ティマールは騎士に代償と徴税権。イクター制や、東ローマでのプロノイアと同様な制度その起源については不明な点が多い。

**正解例** イクター制を継承し、軍人・官僚に軍役・奉仕を課し、給与支払いの代わりに一定の土地からの徴税権を与えた。

この問題ではジャーギール制とティマール制の「共通」の特徴を求められている。つまり、主題であるジャーギール制、ティマール制それぞれの「特徴」について述べられた文を解答としても正解にはならず、「共通」する特徴を解答とする必要がある、よって、単に主題について述べられた文を抽出するだけではなく、さらにそれらの文を比較する必要がある。

## 4.2 より具体的な焦点を必要とする問題

今回用いた焦点、文内容ラベルでは解けない問題がある。以下に問題例を示す。

**問題** 10世紀後半にイスラーム世界に起こった変革の影響を受けて、従来栄えていたペルシア湾ルートに代わって、紅海ルートが栄えるようになった歴史的要因について、政治的状況にも留意して、3行以内で説明せよ。(1997年度)

**出力** アッバース朝は中央集権を整備して、後半から初頭にかけて全盛期を迎えたが、政治的に分裂・弱体化し、西征で滅亡した。特権を解消して、平等を実現し、アラブ帝国から大食への転換をかけた。

**正解例** アッバース朝が衰退し、ブワイフ朝のバグダード入城以後も、イクター制により地方分権化が固定された。一方、エジプトを征服したファーティマ朝は、首都カイロを建設し、紅海貿易を保護した。

**加点ポイント** (一部抜粋)

【ペルシア湾ルートの衰退】

1. アッバース朝が衰えた。
2. ブワイフ朝はイクター制を採用/地方分権化。  
【紅海ルートの発展】
3. ファーティマ朝は、カイロを首都とした。
4. ファーティマ朝は、紅海貿易を保護した。

この問題では、「政治的状況に留意して」など、現在の焦点や文内容ラベルの粒度では判別することができない情報も考慮する必要がある。より多くの問題を観察し、文内容ラベル及びその付与ルールを増やしていくことで対応することも考えられるが、問題を観察することにより詳細な文内容ラベルが必要になる可能性がある。例えば、この問題では「政治的状況」であるが、別の問題ではより細かい「国内政治」というものありえる。つまり、ありとあらゆる問題に対応できるよう適切な粒度で網羅的に文内容ラベルを設け、それを付与することは困難であると考えられる。

## 5 まとめ

大学入試『世界史』小論述問題に対する自動解答システムを開発し、人手による採点及びMRRを用いて先行研究と比較評価を行った。先行研究を上回る結果を得られたが、複数主題を同定し比較することが必要な問題などについては課題が残る。また、現在の文内容ラベルの粒度では対応できない問題がある。

## 謝辞

本研究で用いた世界史イベントオントロジー [2]、山川出版世界史用語集を提供していただいたNTCIR12 QALab-2タスクオーガナイザーの皆様と「ロボットは東大に入れるか」プロジェクトに深く感謝いたします。

## 参考文献

- [1] 高田拓真, 松崎拓也, 佐藤理史. 大学入試「世界史」論述問題解答システムの開発. 言語処理学会第22回年次大会発表論文集, pp. 925–928, 2016.
- [2] Ai Kawazoe, Yusuke Miyao, Takuya Matsuzaki, Hikaru Yokono, and Noriko Arai. World history ontology for reasoning truth/falsehood of sentences: Event classification to fill in the gaps between knowledge resources and natural language texts. In *New Frontiers in Artificial Intelligence*, pp. 42–50. Springer, 2014.
- [3] Mio Kobayashi, Hiroshi Miyashita, Ai Ishii, and Chikara Hoshino. Nul system at qa lab-2 task. In *NTCIR-12 workshop*, pp. 413–420, 2016.