

# 東大入試世界史第1問（大論述問題）を解く 質問応答システムの検討

阪本浩太郎<sup>†1†2†3</sup> 中山周<sup>†3†4</sup> 渋谷英潔<sup>†1</sup> 石下円香<sup>†2</sup> 森辰則<sup>†1</sup> 神門典子<sup>†2</sup>

<sup>†1</sup>横浜国立大学 <sup>†2</sup>国立情報学研究所 <sup>†3</sup>Carnegie Mellon University <sup>†4</sup>東北大学

E-mail: {sakamoto,shib,mori}@forest.eis.ynu.ac.jp,  
{ishioroshi,kando}@nii.ac.jp, nakayamas@ecei.tohoku.ac.jp

## 1 はじめに

近年、大学入試問題をコンピュータに解かせる試みとして、「ロボットは東大に入れるか」プロジェクト<sup>1</sup>（以下、東ロボプロジェクト）や東ロボプロジェクトと連携したNTCIR<sup>2</sup>のQA Lab-2<sup>3</sup>[3]がある。QA Lab-2では、世界史の大学入試問題（センター試験、東京大学を含む5大学の二次試験、センター試験模試、及び東大模試）を対象とした課題が設定されており、我々はこれに参加し主に東大入試と東大模試を対象にしたシステムを構築した。東大入試の問題は大問3問からなっており、第1問は8個程度の指定語句を必ず含ませる500字程度の大論述問題である。本稿では、東大入試の第1問に解答するシステムについて述べる。

第1問を情報要求記述として考えたときの、与えられた知識源に対する文選択を基本とする文章要約として解くシステムとして指定語句ごとに指定語句を含む文を1文ずつ解答に含める手法でQA Lab-1に参加した阪本ら[2]がある。我々は同様のシステムでQA Lab-2のPhase-1と-2に参加した。Phase-1と-2では東大模試としてそれぞれ駿台予備学校の2013/2014第2回東大入試実戦模試《世界史》（以下、Phase-1東大模試）と2015/2016第1回東大入試実戦模試《世界史》（以下、Phase-2東大模試）が出題された。提出したシステムの解答に対して駿台予備学校の講師による採点結果、講評、模範解答、採点基準、解説が得られた。それによると第1問のPhase-1と-2の結果はそれぞれ28点中4点<sup>4</sup>と26点中9点で東大志望受験生の結果に基づく偏差値はそれぞれ42.1（平均点7.3点、標準偏差4.2）と61.8（平均点4.3点、標準偏差4.0）だった。

採点基準には「フランス革命が勃発した」のような解答に含まれていると加点対象になる命題（加点ポイ

ント）が一覧されている。この加点ポイントを基にシステムの間接結果を第一著者が調査したところ、同じ指定語句を含む文の組の中から加点対象となる文を選択するための文の順位付けにおいて、1位の文が加点箇所を含んでいても字数制約のために解答には選択されない場合があった。この加点箇所を解答に含むためには不要な文字を減らす文内要約が必要である。

このような背景からPhase-1と-2の東大模試を対象に文の順位付けについて加点ポイントを用いて人手で評価し、まずPhase-1や-2で用いた手法の有効性を確認する。次に命題に着目した文内要約を提案し、そして文内要約を適用したシステムにおける結果と考察、今後の課題について述べる。

## 2 関連研究

文内要約の研究では「連体修飾要素の削除」に着目した研究が多くある。大竹ら[6]は「連体修飾要素の削除」に加え「直接引用表現の削除」、「補足情報の削除」、「例示の削除」、「換言処理」による文内要約を行うが、本稿で提案する文内要約はそれらのいずれでもない「命題の削除」に着目する点で大竹らとは異なる。

## 3 東大入試世界史第1問

東大入試世界史は3問の大問から成っており、大問毎に、大論述、小論述、語句と解答の形式が異なっている。本稿では、第1問の大論述問題を対象とする。<sup>5</sup>

第1問の質問文は、図のようにコンテキスト部と制約指定部からなっており、コンテキスト部を読ませ、制約指定部で指定される字数制約450字～600字、7個～9個の指定語句、「17世紀から18世紀末まで」のような時間制約などを満たすコンテキスト部に関連する解答が求められる。

## 4 知識源

知識源として高等学校教育で使用される世界史科目の教科書4冊と用語集1冊（東京書籍株式会社の3冊

<sup>5</sup>第2問・第3問への取り組みは、石下ら[1]を参照されたい。

<sup>1</sup><http://21robot.org/>

<sup>2</sup><http://research.nii.ac.jp/ntcir/ntcir-12/>

<sup>3</sup><http://research.nii.ac.jp/qalab/>

<sup>4</sup>ラベルノイズにより問題から時間制約が得られなかった。

近世のヨーロッパでは、それまでにないタイプの国家群が姿をあらわすとともに、海外からの情報が飛躍的に増大した。以下の史料のように、ヨーロッパ人はそれらの情報を基に自国の文化を相対化し、時に異国の文化を賛美した。

(ビュベク『トルコ使節の四書簡』の引用文は本稿では省略)  
(ケネー『中国の専制政治』の引用文は本稿では省略)

17世紀から18世紀にかけての西ヨーロッパ・西アジア・東アジアでは、整備された統治機構と強力な軍隊を持ち、それぞれ異なる思想・理念に基づく集権的な体制を持つ国家が確立し、各地域の政治・外交面で中心的役割を担った。フランスやこれと外交面で提携したオスマン帝国、清朝がその典型である。しかし、18世紀に入るとそれぞれの地域における社会の変動から集権的支配に緩みが生じるとともに、あるいは国家体制を支える理念に対抗する思想も生まれた。そして、18世紀末には国家体制を動揺させる事件も起こった。

以上の点を踏まえ、17世紀から18世紀末までの西ヨーロッパ・西アジア・東アジアにおける集権的な国家体制の変遷について、地域ごとの思想・理念および、体制維持に影響を及ぼした宗教の役割にも留意しつつ、略述しなさい。解答は、解答欄(イ)に20行以内で記述し、必ず次の9つの語句を一度は用いて、その語句に下線を付しなさい。

ウエストファリア条約、ウラマー、クリム=ハン国、啓蒙思想、チベット仏教、ナントの王令廃止、白蓮教との乱、満漢併用制、メッカ

図 1: 第 1 問の例 (Phase-2 東大模試)

「世界史 A (平成 20 年度発行)」, 「新選世界史 B (平成 19 年度発行)」, 「世界史 B (平成 19 年度発行)」と株式会社山川出版社の 2 冊「世界史 B 詳説世界史改訂版」, 「世界史 B 用語集 改訂版」を用いた。

## 5 モジュール

本章では、システム内で使用する文書検索、時間解析、文内要約モジュールについて説明する。

### 5.1 文書検索

キーワード 1 つを入力として受け取り、キーワードを含む文を出力とする文書検索モジュールを Indri 検索エンジン<sup>6</sup>を基に作成する。知識源中の文字を 1gram 単位で、教科書データはパラグラフ、用語集は用語説明を 1 文書としてインデクシングする。

### 5.2 時間解析

テキストを入力として受け取り、テキスト内にある文にそれぞれ時間範囲情報を西暦何年から何年という形式で与える時間解析モジュールである。文に与える時間範囲情報は文で説明される歴史的なイベント全ての時間範囲の和である。テキストを句点区切りの文に分割し前方の文から順に時間解析する。文  $S_n$  から時間範囲を得られない場合は、一つ前の文  $S_{n-1}$  から取得した時間範囲情報を文  $S_n$  に与える。文の時間解析では「時間表現に基づいた時間解析」と「用語集に基づいた時間解析」の出力の和で時間範囲を取得する。

#### 5.2.1 時間表現に基づいた時間解析

「1946 年から 1949 年まで」や「大化 2 年」や「中世」のように入力テキスト内に出現する時間表現から年単位での時間範囲をそれぞれ (1946, 1949), (646, 646), (476, 1453) のように抽出する。「90 年代」のような省略

表現については、テキストを前方解析して「90 年代」が出現する前の時間情報を 100 で割った商を、情報がなければ 1900 年を加え省略年数を補完する。

#### 5.2.2 用語集に基づいた時間解析

山川出版社の用語集を用いて見出し語が入力テキスト内に含まれていた場合にはその見出し語に用語集で紐付いた時間情報を取得する。

### 5.3 文内要約

文内要約モジュールは文を入力とし、文内に含まれる命題を文節のまとまりとして全て抽出する。要約文はそれぞれの命題の有無に着目して作成する。例えば、文  $S$  に含まれている命題集合  $\{P_1, P_2, P_3\}$  の冪集合から空集合と全体集合を除く 6 つの部分集合  $\{P_1\}, \{P_2\}, \{P_3\}, \{P_1, P_2\}, \{P_2, P_3\}, \{P_3, P_1\}$  を得る。ここから部分集合ごとに内在する命題のみを含む要約文 6 文を得たいが、命題の文節を単純に結合すると不自然な文となったため、不自然さの軽減として部分集合を持つ文節を文  $S$  の文節から除去する形で要約文を得る。

#### 5.3.1 文からの命題抽出

文から文に含まれる命題を全て自動抽出する。命題は 3 種類の文節「述語文節」「述語文節に係る文節」「述語文節に係る文節と並列関係にある文節」のまとまりとし、各文節抽出に述語項構造解析器 ChaPAS<sup>7</sup>の出力から述語、並列関係、係り受け関係の情報を用いた。

#### 5.3.2 命題除去による文内要約

5.3.1 で抽出した文節のまとまりとしての命題を文から除去することで文内要約をする。命題  $\{P_1, P_2, P_3\}$  を含む文  $S$  から  $\{P_1\}$  のみを含む要約文を得るために  $\{P_2, P_3\}$  を除去する場合、文  $S$  を文節単位に分解し文  $S$  の文節リストを得る。命題  $P_2$  が持つ文節と命題  $P_3$  が持つ文節を文  $S$  の文節リストから除去する。文  $S$  の文節リストに残る文節を結合し要約文を得る。

## 6 ベースラインシステム

阪本ら [2] を基に作成したベースラインシステムについて述べる。

処理の流れは図 2 に示すように、質問文として東大入試第 1 問の問題文を入力し、解候補として論述文を出力する。質問文解析で質問文から指定語句、字数制約、時間制約を抽出する。文書検索モジュールを用いて関連文書の検索で指定語句を含む文書を知識源から抽出する。文抽出では関連文書から指定語句を含む文を句点区切りで抽出する。この際に文に時間解析モジュールを用いて時間情報を与える。文の組の生成では、同じ指定語句を含む文をまとめて指定語句ごとに組を作成する。文の順位付けでは、文に解候補に含むべきか

<sup>6</sup><http://www.lemurproject.org/indri.php>

<sup>7</sup><https://sites.google.com/site/yotarow/chapas>

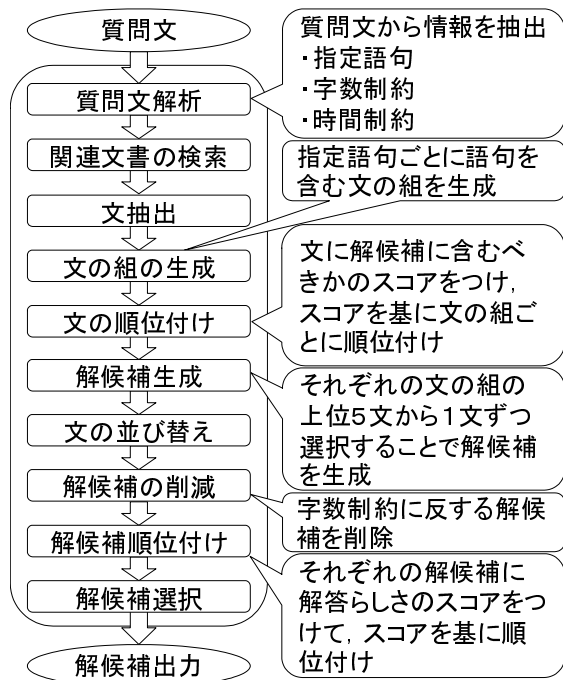


図 2: ベースラインシステムの処理の流れ

のスコアをつけ、そのスコアを基に文の組ごとに順位付けを行う。スコアが等しい場合は文の文字数が多い順に順位付ける。解候補生成ではそれぞれの文の組の上位5文から1文ずつ選択し組合せることで解候補を生成する。文の並び替えは解候補中の文を文の時間情報を基に古い順に並べる。解候補の削減では、字数制約に反する解候補は削除する。もし字数制約を満たす解候補が存在しない場合は最小字数の解候補のみ残す。解候補順位付けでは、それぞれの解候補に解答らしさのスコアをつけて、スコアを基に順位付けを行う。解候補の解答らしさのスコアは解候補に含まれる文のスコアの合計で計測する。解候補選択では、解答らしさのスコアが最大の解候補を選択する。

### 6.1 文の解候補に含むべきかのスコア

問題文の特にコンテキスト部は解答に含んで欲しい内容が抽象的に書かれていると考えられる。そのため、文の解候補に含むべきかのスコアとして、文の内容を問題文の内容が包含しているかの包含度を用いた。包含度は、UniDic 辞書を用いた形態素解析器 MeCab を基に内容語を取得し重複を削除し、問題文に含まれる内容語が文に含まれている割合で計測する。

## 7 提案手法

ベースラインシステムの改良として、時間解析モジュールを使用した提案手法1（時間解析）と時間解析と文内要約の両方のモジュールを使用した提案手法2（時間解析+文内要約）の2つの手法を提案する。

### 7.1 提案手法1（時間解析）

提案手法1は文の組の生成と文の順位付けの間に文の削減を加える。文の削減では時間解析モジュールを用いて問題文から抽出した時間制約を満たさない文の削減を行う。

### 7.2 提案手法2（時間解析+文内要約）

提案手法2は、提案手法1と同様に時間解析モジュールによる文の削減に加えて、さらに文の削減の後に文の拡張を加える。文の拡張では、文内要約モジュールを用いて文を命題単位に分割し、それぞれの命題を含むか否かによる組合せ数だけ要約文を生成し、指定語句を含む要約文を文の組に加える。

## 8 評価実験

本章では、評価実験の実験目的、実験設定、評価手法について説明する。

### 8.1 実験目的

システムが指定語句ごとに出力した文の順位付けの評価により、提案した時間解析や文内要約を用いた文の順位の評価への影響を調べる。

### 8.2 実験設定

問題と採点基準の両方を取得済みの、Phase-1 東大模試と Phase-2 東大模試の2回分の第1問の計2問を実験対象とする。ベースライン、提案手法1、提案手法2の3システムの出力を調査対象とする。評価者は情報工学を専攻する修士課程の学生4名であり、評価者には問題、採点基準、模範解答、解説、システムの出力のみを参照させる。

### 8.3 評価手法

システムが指定語句ごとに出力した文の順位付けの上位5文に対して、文が複数の駿台予備学校の加点ポイントの内1つでも含意しているかを評価者が評価する。順位付けの評価として何位に加点される文が最初に現れるかを見るため MRR を用いる。評価者間の信頼性は評価結果の一致度を Fleiss の  $\kappa$  係数で測る。

## 9 結果と考察

本章では、評価実験の結果と考察について説明する。

### 9.1 結果

表1が示すように MRR は、Phase-1 東大模試ではベースラインと提案手法1の両方が、Phase-2 東大模試では提案手法1が最も高かった。提案手法2の MRR は両問とも最も低かった。Fleiss の  $\kappa$  係数は中程度の一致であった。

## 9.2 考察

ベースラインと提案手法1のMRRはいずれも高い値であったが、より提案手法1のMRRが高いため時間解析の有効性が示せた。内容語の包含度による文の順位付けの改良のために内容語の一致物について調査した。評価者全員が加点箇所があると評価し提案手法1で上位1文だった「しかし土地の不足による農民の貧困化や開墾による環境破壊が社会不安をうみだし、18世紀末には四川を中心とする新開地で白蓮教徒の乱がおこった。」が問題文と一致した内容語は「然し, 化, 因る, 教, 徒, 18, 世紀, 末, 中心, 起こる」であった。基本的に問題文には具体的な語句が含まれないが、この結果からはかなり内容が薄い語の一致で順位付けていることが観察される。順位付けの精度向上には文の語の上位語による拡張が必要かもしれない。

提案手法2の文内要約による文の拡張で順位の評価が上がらなかった理由として、同じ文から要約された文が上位に多く固まってしまい、元の文に加点箇所がないため、ベースラインや提案手法1よりRRが下がってしまう問題が見られた。同じ文から要約された文が上位5文全てを占めてしまう場合は18件中11件見られ、この内4件は上位5文全てに加点箇所がないと全評価者が評価した。この問題の対策として同じ文由来で同スコアの文を削減する必要がある。例えば、文字数最大の文と最小の文以外の文を削除することが考えられる。他の理由として、「1920年には女性参政権が実現し、また、」のように文末が不自然な文や「…農民の貧困化や開墾によるうみだし…」の「うみだし」の目的格が除去されて非文にになってしまうといった文内要約の失敗が上位5件に含まれていた。この対策として、文内要約した文を整える処理を加えることと、不自然な文の順位を下げるように文に文の不自然さのスコアをつけることが考えられる。文の不自然さをどのような尺度や計算方法でスコアづけるのかは今後の課題である。

実験で最も評価が高かった提案手法1のPhase-1と-2の東大模試の評価者4名による最終得点の平均点はそれぞれ2.75点と10.5点であった。字数制約を無視し上位5文から最も加点される文を選択した場合の評価者

表 1: 上位5文に対する評価結果 (MRR)

手法\問題	Phase-1	Phase-2
ベースラインシステム	<b>0.551</b>	0.630
提案手法1	<b>0.551</b>	<b>0.692</b>
提案手法2	0.340	0.583
評価者間の一致率	0.538	0.531

4名による最終得点の平均点はそれぞれ6.5点と13.5点であった。この点から、文から加点対象にはならない箇所を削除する文内要約と順位付けの改良によってそれぞれ2.75点から6.5点や10.5点から13.5点に得点をあげられる可能性があることがわかる。しかし満点はそれぞれ28点と26点であるため、現在のシステムの文内要約と順位付けを改良したとしても満点には届かなそうである。満点を目指すには文を横断し加点箇所を組合せる文生成が必要かもしれない。

## 10 まとめ

Phase-1と-2の東大模試を対象に文の順位付けについて加点ポイントを用いて人手で評価した。Phase-1や-2で用いた手法の有効性を確認し、今後の改良点を述べた。命題に着目した文内要約を提案し、文内要約を適用したシステムの結果から要約文の不自然さを修正する処理や不自然さのスコアの必要性を述べた。満点を目指すには文の順位付けと文内要約の改良だけではなく文を横断し加点箇所を組合せる文生成が必要性的について述べた。

## 参考文献

- [1] 石下円香, 阪本浩太郎, 中川周, 渋谷英潔, 森辰則, 神門典子. 東大入試世界史第2問 (小論述問題) 及び第3問 (語句問題) を解く質問応答システムの検討. (2016年言語処理学会第22回年次大会にて発表予定).
- [2] 阪本浩太郎, 渋谷英潔, 石下円香, 森辰則, 神門典子. 大学入試の論述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, pp. 179-182, (2015).
- [3] 石下円香, 阪本浩太郎, 渋谷英潔, 狩野芳伸, 三田村照子, Wang Di, 森辰則, 神門典子. 大学入試問題を解くための質問応答システムにおける現状と課題. 言語処理学会第21回年次大会発表論文集, pp. 183-186, (2015).
- [4] 西川仁, 平尾努, 牧野俊朗, 松尾義博, 松本裕治, 冗長性制約付きナップサック問題に基づく複数文書要約モデル, 自然言語処理, Vol.20, No.4, pp.585-612, (2013).
- [5] 新井紀子, 松崎拓也. ロボットは東大に入れるか?-国立情報学研究所「人工頭脳」プロジェクト. 人工知能学会論文誌, (2012).
- [6] 大竹清敬, 岡本大吾, 児玉充, 増山繁, 重要文抽出, 自由作成要約に対応した新聞記事要約システムYELLOW, 情報処理学会論文誌「データベース」Vol.43, No.SIG 2(TOD 13), pp.37-47 (2002).