

# 大学入試世界史論述問題における 非指定重要語句生成に関する検討

阪本浩太郎<sup>†1†2</sup> 陸宇傑<sup>†1</sup> 福原優太<sup>†1</sup> 渋木英潔<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

E-mail: {sakamoto,luyujie,shib,mori}@forest.eis.ynu.ac.jp,  
{ishioroshi,kando}@nii.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 [1]<sup>1</sup> (以下、東ロボプロジェクト) や NTCIR<sup>2</sup> の QA Lab-3 [2]<sup>3</sup> がある。QA Lab-3 では、世界史の大学入試問題 (センター試験, 二次試験) 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二次試験における論述問題を対象にしたシステムを構築している。

我々のシステム [3] は、問題中にある指定語句を検索語として文書検索をし、指定語句を含む文を抽出し、それらを組み合わせることで解答を作成する。ところが、知識源中にある加点対象となる文の中には、指定語句を用いても検索できず、問題文内の世界史用語を用いても検索できないものが存在する。例えば、図?? の問題の模範解答には「イスラームの暦学に基づいて元の郭守敬が「授時暦」を作成する」という記述が存在し、高校世界史の教科書 4 冊と用語集 1 冊を対象に同様の命題を含む箇所を調べたところ、「イスラーム天文学の知識にもとづいて中国で郭守敬が授時暦をつくった。」という記述が見つかった。この文内及び隣接する文には指定語句は含まれないため指定語句による検索でこの記述を得ることは困難であり、この記述を得るためには「郭守敬」や「授時暦」で検索することが一般的に考えられる。そこで、本研究では、「郭守敬」や「授時暦」のように問題文中にはないが加点対象となる記述に含まれる用語を非指定重要語句と呼び、問題からの非指定重要語句の生成を行う。非指定重要語句は、大量の問題と解答の組を用いることで効率的に得られるのではないかと仮定し、問題文に類似する教科書と用語集の記述から探す方法をベースラインとし、問題集の記述から探す方法を提案手法として結果を比較する。さらに、提案手法として一般的な検索を用い

歴史上、異なる文化間の接触や交流は、ときに軋轢を伴うこともあったが、文化や生活様式の多様化や変容に大きく貢献してきた。たとえば、7世紀以降にアラブ・イスラーム文化圏が拡大するなかでも、新たな支配領域や周辺の他地域から異なる文化が受け入れられ、発展していった。そして、そこで育まれたものは、さらに他地域へ影響を及ぼしていった。13世紀までにアラブ・イスラーム文化圏をめぐって生じたそれらの動きを、17行以内で論じなさい。その際に、次の8つの語句を必ず一度は用い、その語句に下線を付しなさい。インド、アッバース朝、イブン=シーナー、アリストテレス、医学、代数学、トレド、シチリア島

図 1: 2011年東大世界史第1問

た方法と教師あり学習を用いた方法を比較する。

## 2 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 (7,038 用語) と論述問題が含まれる問題集 6 冊 (全部で論述問題 1,205 問) (東京書籍株式会社の 3 冊「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 3 冊「世界史 B 詳説世界史 改訂版」, 「世界史 B 用語集 改訂版」, 「詳説世界史論述問題集」, Z 会の 2 冊「実力をつける世界史 100 題改訂第 3 版」, 「段階式世界史論述のトレーニング」, 河合出版の「判る解ける書ける世界史論述 (河合塾シリーズ)」, KADOKAWA/中経出版の「大学入試世界史 B 論述問題が面白いほど解ける本」, 株式会社パレードの「世界史論述練習帳 new」) を用いた。

### 2.1 Word Embedding

教科書 4 冊と用語集 1 冊と問題集 6 冊の中にある論述問題を fastText<sup>4</sup> を用いて学習し、ベクトル長 100 のベクトルを 52,939 語彙分得た。実行コマンドのオプショ

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

<sup>2</sup><http://research.nii.ac.jp/ntcir/ntcir-13/>

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

<sup>4</sup><https://github.com/facebookresearch/fastText>

ンとして、fastText のサンプルプログラムに含まれていた Skip-gram モデルで学習するための「skipgram」と、例えば「ウェストファリア条約」のように世界史の教科書中に頻度としてはほとんど現れないが重要な用語が存在するため「-minCount 1」を使用した。

### 3 非指定重要語句生成

本章では、システム内で使用する非指定重要語句生成について説明する。

問題文中の内容語及び指定語句を入力とし、単語とその単語が解答に含まれるべきかを表す値の組のリストを出力する。

#### 3.1 ベースラインシステム

与えられた問題文に含まれる内容語に対して知識源内で共起する頻度が高い内容語を非指定重要語句として出力する。

##### 3.1.1 ベースライン 1 (類似文書検索 TF・IDF)

与えられた問題文と内容が近い記述を教科書や用語集の中から検索し、その周辺にある用語の頻度が高いものがより非指定重要語句に近いものだと仮定して順位付けて出力する。検索対象として単語 1-gram 単位でインデクシングをした教科書と用語集を用いる。検索文書の単位は、教科書は 1 パラグラフ、用語集は 1 見出し語とその説明文である。問題文と類似する記述の検索に TF・IDF を用いる。類似度上位 3 件の検索文書に含まれる内容語を頻度に基づいて順位づける。順位付けられた内容語の内、問題文中にはなく用語集の見出し語にある内容語のみを出力する。

##### 3.1.2 ベースライン 2 (類似文書検索 BM25)

ベースライン 1 の検索アルゴリズムを TF・IDF ではなく Okapi BM25 を用いる。

### 3.2 提案手法

#### 3.2.1 提案手法 1 (類似問題検索 TF・IDF)

与えられた問題文と内容が近い問題文を問題集から検索し、検索された問題の模範解答中に含まれる用語の頻度が高いものがより非指定重要語句に近いものだと仮定して順位付けて出力する。問題文と類似する問

インドアッパース朝 イブン=シーナー アリストテレス 医学  
代数学 トレド シチリア島 歴史上異なる文化間 接触 交流  
時 軋轢 伴う文化 生活様式 多様化 変容 大きい 貢献 来る  
例えば 7 世紀以降 アラブ イスラーム文化圏 拡大 中 新た 支  
配領域 周辺 他地域 異なる文化 受け入れる 発展 行く 所  
して 其処 育む 更に 他地域 影響 及ぼす 行く 13 世紀 アラブ  
イスラーム文化圏 巡る 生ずる 其れ等 動き 解答欄 510 字以  
内 論ずる 為さる 其の際 次 8 つ 語句 必ず 1 度 用いる 其の  
語句 下線 付する 為さる

図 2: 問題中の内容語の例

アラブ人 正統カリフ時代 ビザンツ帝国 エジプト 奪う ササン  
朝 滅ぼす イラン 支配 ウマイヤ朝時代 イベリア半島 インダ  
ス川流域 領域 広げる こう 領域 拡大 過程 ビザンツ帝国 保  
存 古代ギリシア ローマ文化 吸収 アッパース朝 首都 バグ  
ダード 立てる 知恵の館 アリストテレス 哲学 ヒポクラテス 医  
学 文献 アラビア語 翻訳 更に 研究 深める インド ゼロの概  
念 十進法 伝わる フォーリズミー 因る 代 代数学 確立 大き  
い 寄与 又 タラス河畔の戦い 因る 中国 製紙法 伝える 此れ  
等 十字軍 レコンキスタ 過程 ヨーロッパ 流入 トレド シチリア  
島 於く アラビア語 文献 ラテン語 翻訳 12 世紀 ルネサンス 大  
きな 影響 与える 医学 イブンシーナー 医学典範 中 世 ヨー  
ロッパ 大学 講義 使う イブンルシュド 因る アリストテレス 注  
釈 トーマス アクイナス 因る スコア 哲学 体系化 寄与 又 イスラ  
ム 暦学 基づく 元 郭守敬 授時 暦 作成 ヨーロッパ 無い 中国  
影響 及ぼす

図 3: 模範解答中の内容語の例

題の検索に TF・IDF を用いる。類似度上位 3 件の問題の解答中に含まれる内容語を頻度に基づいて順位付ける。順位付けられた内容語の内、問題文中にはなく用語集の見出し語にある内容語のみを出力する。

#### 3.2.2 提案手法 2 (類似問題検索 BM25)

提案手法 1 の検索アルゴリズムに TF・IDF ではなく Okapi BM25 を用いる。

#### 3.2.3 提案手法 2 (教師あり学習)

問題集にある問題と模範解答の組を用いて教師あり学習を行う。教師データとして、問題集 6 冊の論述問題文中の内容語と指定語句の組 2 を入力し、作製した Word Embedding の語彙リスト上に模範解答中にある内容語 3 があれば、その語彙に 1、なければ 0 を与えて作製したリストを正解情報として作製した。図 4 に示すように、Elman 型 Recurrent Neural Network (以下、Elman net) を用いて、Word Embedding で学習した語彙を順位付ける。順位付けた語彙の内、問題文中にはなく用語集の見出し語にある語彙のみを出力する。

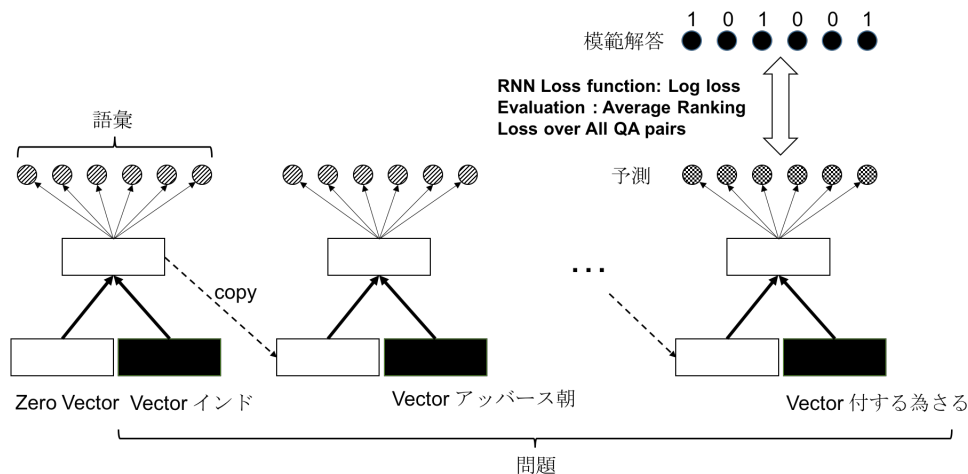


図 4: Elman net を用いた提案手法

## 4 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。

### 4.1 実験目的

システムが問題ごとに出力した語の順位付けの評価により、提案した類似問題検索や教師あり学習の、語の順位への影響を調べる。

### 4.2 実験設定

2000年から2011年の東京大学世界史科目第一問計12問を実験対象とする。2つのベースラインと3つの提案手法の出力を調査対象とする。システムが出力した語の順位の上位の語が模範解答に含まれるかどうかを調査する。模範解答は、世界史科目で塾講師や教師の経験がある専門家3名がそれぞれ作成したものと赤本に掲載されているものの計4つを用いる。

### 4.3 評価手法

システムの出力された語のランキングが、4つの模範解答のいずれかに含まれれば正解として評価する。語の順位の上位に解答に含むべき語が集まっているのかを見たいので、評価尺度として Mean Average Precision at N ( $N = 5, 10, 30$ ) を用いる。

## 5 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。

### 5.1 結果

表1が示すように、教科書と用語集を用いた手法より提案した問題集を用いた手法の方が非指定重要語句生成の精度が高かった。さらに、提案手法の内、類似問題検索に基づく手法より、Elman netを用いた手法の方が精度が高かった。

### 5.2 考察

教科書と用語集を用いた手法より提案した問題集を用いた手法の方が非指定重要語句生成の精度が高かったため、問題集を用いる方が効率的に非指定重要語句を生成できることが示された考えられる。

## 6 まとめ

教科書と用語集を用いた手法より提案した問題集を用いた手法の方が非指定重要語句生成の精度が高かつ

表 1: MAP@N による評価結果

手法	@5	@10	@30
Baseline (TFIDF)	0.291	0.284	0.224
Baseline (BM25)	0.427	0.416	0.284
Proposed (TFIDF)	0.522	0.433	0.343
Proposed (BM25)	0.493	0.425	0.333
Proposed (Elman net)	<b>0.594</b>	<b>0.572</b>	<b>0.461</b>

た。提案手法の内、教師あり学習として Elman net を用いた手法が最も精度が高かった。今後は他の学習モデルとの比較を行いたい。

## 謝辞

本研究の実施にあたって、模範解答を作成していただいた株式会社 IR-ALT の皆様と問題集の電子化作業を手伝ってくださった横浜国立大学の学生の皆様に深く感謝いたします。

## 参考文献

- [1] 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).
- [2] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, and Noriko Kando. Overview of the NTCIR-12 QA Lab-2 Task. In Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, pp. 392–408, (2016).
- [3] 阪本浩太郎, 中山周, 渋谷英潔, 石下円香, 森辰則, 神門典子. 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. 言語処理学会第22回年次大会発表論文集, pp. 553–556, (2016).