

平成 27 年（2015 年）11 月 14 日

日本ユニシスがセンター試験模試「世界史 B」で好成績 ～NII 人工知能プロジェクト「ロボットは東大に入れるか」～

大学共同利用機関法人 情報・システム研究機構 国立情報学研究所（NII、東京都千代田区、所長：喜連川 優）が取り組む人工知能（AI）プロジェクト「ロボットは東大に入れるか」（プロジェクトディレクター：NII 情報社会相関研究系教授 新井 紀子）の本年度の研究活動に参加した日本ユニシス株式会社（日本ユニシス、本社：東京都江東区、代表取締役社長：黒川 茂）は、株式会社ベネッセコーポレーションから本プロジェクトにご提供いただいた「進研模試 総合学カマーク模試」の世界史 B に挑戦し、平均点を 30 点上回る 76 点（偏差値 66.5）という好成績を収めました。

大学入試センター試験（センター試験）の世界史の問題をコンピュータで解くためには、自然言語で記述されている教科書などを知識源として、各設問に記述されている質問文に対する適切な答えを解答群の中から正しく選択することが求められます。自然言語処理分野の研究において、自然言語で記述された知識源に関する質問に答えるために有効とされる主な手法には、「含意関係認識技術^(*1)の適用」「事実型質問応答技術^(*2)の応用」「こうぶんぎ構文木の類似度評価」があります。日本ユニシスでは、これらの手法に基づき、センター試験の世界史の観点から解法を独自に定式化して解答を導き出す仕組みを実装することで、今回の成果を達成しました。

「ロボットは東大に入れるか」は、昭和 50 年代後半以降細分化された AI の分野を再統合することで AI 研究の新たな地平を切り拓こうと、平成 23 年度（2011 年度）に始まりました。大学入試問題は問題文を解析する自然言語処理をはじめ様々な技術が求められる統合的な課題で、点数と偏差値により成果を定量的に評価することが可能なタスクです。こうした特性を持つ大学入試問題に AI が挑戦することで、「AI が人間に取って代わる可能性のある分野は何か」といった問題を考える際の指標となりうる AI の進化の客観的なベンチマークを指し示すことが、本プロジェクトの目的です。

本成果は国立情報学研究所と日本ユニシス株式会社が共同で発表するもので、日本ユニシスも同内容のニュースリリースを発行しています。重複して配信される場合がありますことを、ご了承お願いいたします。

【取り組みの背景】

日本ユニシス総合技術研究所^(*)3)は平成26年度(2014年度)、含意関係認識に関する同社の研究のベンチマークとして、NIIの情報アクセス技術の評価と性能比較のための研究基盤プロジェクト「NTCIR(エンティサイル)」^(*)4)の含意関係認識部門(RITE-VAL)に参加し、文に含まれる単語の一致率に着目した機械学習と全文検索技術を駆使した含意関係認識システムを構築してトップの成績を収めました。しかし、同部門で構築したシステムの実力を試すためにNTCIRの大学入試問題の世界史を解く部門(QALab)でセンター試験の過去問に挑戦した結果、同システムをそのまま適用しても正解を導き出せないケースがあることがわかりました。

このため、同研究所ではセンター試験の世界史の問題をあらためて観察し、正解を導き出すために必要となる手法について検討を重ねて、本年度、NTCIRのQALabに再挑戦することを通じて、NTCIRと協賛関係にある「ロボットは東大に入れるか」プロジェクトのセンター試験の世界史受験に取り組みました。

【解法の概要】

センター試験の世界史では、解答群の各選択肢の文が意味する事柄が教科書などの知識源に含まれていれば正文、知識源から導き出すことができなければ誤文と判断する「正誤判定問題」が最も難易度が高く、出題の大部分を占めます。

日本ユニシス総合技術研究所では、知識源を有効に活用できるようにするため、教科書やWikipedia、イベントオントロジーEVT^(*)5)から問題に関係する箇所を局所的に抽出可能な全文検索システムを構築しました。また、問題に含まれる人名、出来事名、地名といった固有表現が問題を解く上での鍵となるため、固有表現や語彙の意味関係^(*)6)の辞書を知識源から半自動で情報抽出することで拡充しました。その上で、正誤判定問題を解くため、以下の手法により解答を生成しました。なお、正誤判定問題のほかに、穴埋め問題、出来事が起こった年代順を答える問題にも同じ手法を適用することができます。

(1) 事実型質問応答技術の適用

正誤判定問題の誤文では、年代、国名、人名などの固有表現に誤りがあるケースが多くあります。誤りを検出するためには、解答群の選択肢の文に含まれる固有表現一つ一つを順番に空欄として扱うことで、穴埋め問題として捉えて事実型質問応答技術を適用することができます。今回、知識源における質問と回答の位置関係の凝集度を独自に定式化し、回答候補を精度良くスコアリングする手法を実装しています。

(2) 照応解析を含めた、構文木の類似度評価

解答群の選択肢の文と教科書の文それぞれから生成した構文木の類似度を評価することで、文の構造の矛盾や反義語を検知し、正誤を判定します。ただし、教科書内の単文においては、主語、時間、場所などの省略や、代名詞による参照を含んでいるため、教科書の文をそのまま構文木に変換しても正しく類似度を評価できないこともあります。そのため、対象となる文の前に記述されている複数の文を参照し、省略の補完や、代名詞の参照先の解決を行う照応解析を組み込んでいます。また、構文木の類似度評価で判定できない場合には、含意関係認識により判定します。

(3) 単語の共起確率

世界史におけるこの「出来事」ならこの「人物」といった固有表現のペアで記憶するという受験テクニックを取り入れ、解答群の選択肢に含まれる固有表現のペアが、教科書や Wikipedia 内で近接して現れる確率を算出することで、ペアとして現れやすい固有表現、現れにくい固有表現を判別し、正誤を判定します。

(4) アンサンブル学習による解答

各手法の不安定さを抑えるアンサンブル学習^(*)の効果を得る目的で、上記(1)～(3)の3手法それぞれの解答から合議制で最終的な解答を導き出しました。

【今後の展開】

今回、既存の自然言語処理の手法をセンター試験の世界史に適応した形で再定式化し、これまでにない高得点を獲得することができました。ここで使用した世界史という特定ドメインへの適応の方策や手順などは、専門性が高い分野で精緻な情報が必要とされる場面において広く応用できると考えています。

一方で、過去問や模試での誤りを解析した結果、今回のアプローチ(知識源からの検索として解く手法)にも限界があることを感じており、さらに上を目指すには、より広い意味での常識や、きめ細やかな日本語運用能力などを備えた仕組みと組み合わせる必要があります。

日本ユニシス総合技術研究所では、会議室などコミュニケーションの場として利用される空間が、その場の雰囲気や会話の文脈を理解し、状況に応じた情報を提示することで会話の進展をサポートできる「空

間プラットフォーム」の実現を目指しています。現在は、人間がコミュニケーションを取るときの前提として共有している常識的な背景知識や感覚（コモンセンス）を備える AI としての「コモンセンス AI」を基にして、空間内で人と AI とが自然なインタフェースを介して相互作用できる環境の研究開発を進めています。

今後は、会話の文脈を理解し、コミュニケーションの目的や内容に応じて、より専門性の高い精緻な知識や過去の正確な事象についての知識を備える AI も必要となります。今回の取り組みで得られた研究成果とコモンセンス AI を融合させることで、空間プラットフォームの実現に一步近づけることができると考えています。

以上

※記載の会社名および商品名は、各社の商標または登録商標です。

〈メディアの皆様からのお問い合わせ先〉

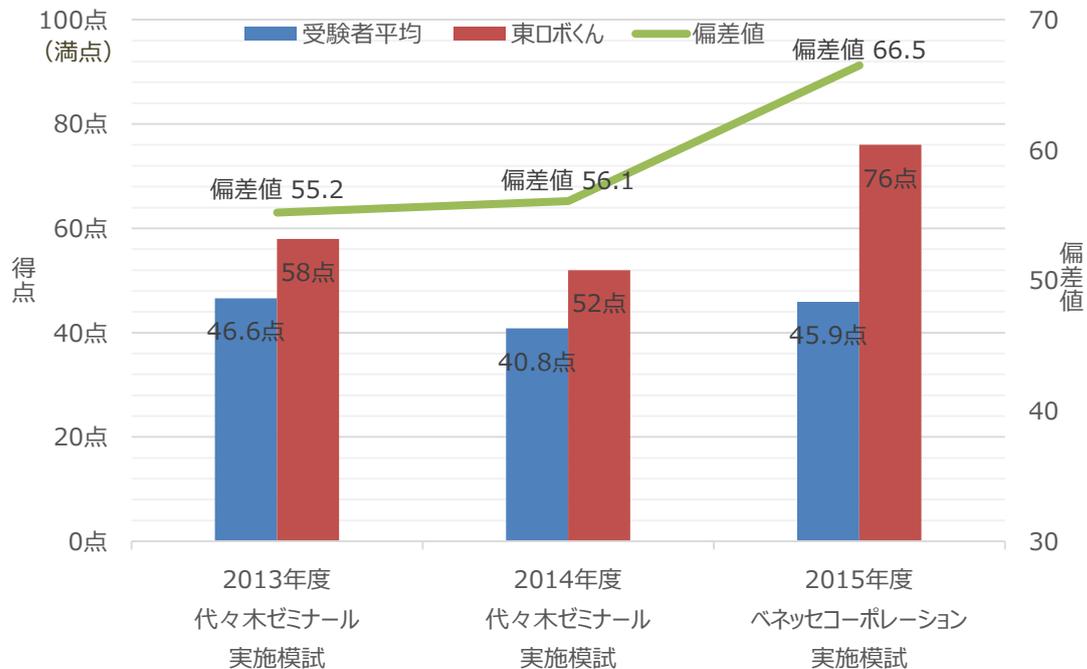


国立情報学研究所
広報チーム（担当：美土路）
TEL：03-4212-2164
E-Mail：media@nii.ac.jp



日本ユニシス株式会社
広報部 報道室（担当：小林、滝澤）
TEL：03-5546-7404
E-Mail：i-Box@unisys.co.jp

■過去2年間の受験結果との得点比較（世界史 B）



※日本ユニシスの成績は本年度（2015年度）の「東口ポくん」の成績のみ

■ 問題と解答の例（正誤判定問題）

分権的な統治が行われた③グプタ朝では、バラモンが周縁地域の農村に移住し、土地の寄進を受ける例が多くみられるようになった。彼らは未開拓地を開墾し、この時代の農業の発展に貢献した。村落の農民たちは、地方の部族長が発する命令によって、農村で存在感を増しつつあったバラモンへの貢納を課せられるようになった。

問 下線部③に関連して、この時代の出来事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。

- ① プルシャブラを都とした。
- ② カーリダーサが『シャクンタラー』を著した。
- ③ ナーランダー僧院が建設された。
- ④ エフタルが侵入した。

2014年度6月総合学カマーク模試世界史B(トレーニングデータ)

選択肢	手法 (1) (事実型質問応答)	手法 (2) (構文木類似度評価)	手法 (3) (単語の共起確率)
① プルシャブラを都とした。	1	1	1
② カーリダーサが『シャクンタラー』を著した。	3	4	2
③ ナーランダー僧院が建設された。	3	3	3
④ エフタルが侵入した。	3	2	4

正解!
解答は

①

各手法で10点を振り分け、合計が小さい選択肢を解答とする

(1) ~ (3) の手法それぞれで、各選択肢の正文・誤文である可能性のスコアを算出する。選択肢「① プルシャブラを都とした」についての各手法の処理の概要は以下のとおり。

手法 (1) : 下線部から「グプタ朝」を取得。「プルシャブラ」を隠して質問応答を用いると、「パータリプトラ」を発見。「プルシャブラ」はなかなか見つからないので、誤文である可能性のスコアを高く設定する。

手法 (2) : ガ格として下線部「グプタ朝」を取得。教科書内の関連度が強い文章と構文木を比較すると、教科書の文章はワ格が「プルシャブラ」ではなく「パータリプトラ」という異なる固有表現になっているため、誤文である可能性のスコアを高く設定する。

手法 (3) : 下線部「グプタ朝」と「この時代」というキーワードから、イベントオントロジーEVTを検索し「320年から550年」という年代を取得。単語の共起確率は高いものの、年代と共起しないため、誤文である可能性のスコアを高く設定する。

(*1)「**含意関係認識技術**」: 2つの文の間に含意関係が成り立つかどうかを判別する自然言語処理の技術。2つの文の間に含意関係が成り立つとは、一方の文から他方の文が推論できることを指す。

(*2)「**事実型質問応答技術**」: 「日本で一番高い山は?」という質問に対して「富士山」と答えるなど、質問に対して事実に基づいて適切な単語で回答する自然言語処理の技術。

(*3)「**日本ユニシス総合技術研究所**」: 『『データ』を『価値のある情報』に変えるソフトウェア』『企業価値向上を支える情報システムの方法論とアーキテクチャ』『社会プラットフォームを支える技術』の提供をミッションとし、「アイデアをすぐカタチに」をモットーにスピード感のあるR&D(研究開発)の実現を目指して活動。

(*4)「**NTCIR**」: 正式名称は「情報検索システム評価用テストコレクション構築プロジェクト(NII Testbeds and Community for Information access Research)」。検索システムの性能評価という切り口から、情報検索やテキスト要約、テキストマイニング、意見や動向の分析など膨大なデータの中から求める情報にアクセスするためのテキスト処理技術の研究に取り組むNIIのプロジェクト。RITE-VALでは参加チームがそれぞれ独自のアプローチを含意関係認識の評価データに適用し、その手法について相互評価する。QALabは大学入試問題を解くことを目的とした部門で科目は世界史に限定し、センター試験および複数の大学の二次試験が対象。

(*5)「**イベントオントロジーEVT**」: 「出来事(イベント)の成立を記述する文」の真偽判断をサポートするオントロジー(体系的な知識の記述)。「イベントオントロジーEVT」の公開ウェブページ(関連リンク参照)では、XML形式のオントロジーファイルと、当該オントロジーに基づいて高校世界史教科書において言及される重要な人物、組織、国家、歴史的出来事などについて収集した情報をまとめたExcel形式のファイルが提供されている。今回はこれら二つのファイルを利用。

(*6)「**語彙の意味関係**」: 語彙どうしの階層関係(上位・下位)、類義関係などのこと。

(*7)「**アンサンブル学習**」: 複数の手法を融合させて一つの手法を構成すること。個々の手法の不安定さ(ばらつき)を低く抑え、未知の問題に対応する能力を向上できる。

■関連リンク

- 「ロボットは東大に入れるか」:
<http://21robot.org/>
- 日本ユニシスの研究開発方針(日本ユニシス総合技術研究所):
<http://www.unisys.co.jp/com/tech/index.html>
- 「NTCIR」:
<http://ntcir.nii.ac.jp/jp/>
- イベントオントロジーEVT:
<http://researchmap.jp/zoeai/event-ontology-EVT/>