2A4-6

質問応答システムでの解答に向けた大学入試問題の分析

An Analysis of the Questions of the University Entrance Examination to Answer Using the Question Answering System

石下 円香 *1　狩野 芳伸 *2　神門 典子 *1
Madoka Ishioroshi　Yoshinobu Kano　Noriko Kando

*1国立情報学研究所
National Institute of Informatics

*2科学技術振興機構 さきがけ
PRESTO,Japan Science and Technology Agency (JST)

We aim to develop a system which solves the questions of the university entrance examination with the question answering system (QA system). We conducted a preliminary experiment in order to investigate a possibility and problems of solving the questions of the National Center Test for University Admissions using the QA system. The format of the questions of the examination is different from the format of questions that the existing QA system to answer. So, we classified the questions of the examination into four types. Then, we manually converted the questions of the examination into questions that the QA system to answer and entered into the QA system. From the results of the preliminary experiment, we found problems of the QA system.

1. はじめに

現在、国立情報学研究所は、大学入試問題を解く計算機プログラムを開発することを目的とした、人工知能プロジェクトを進めている [新井 12]。このプロジェクトでは、2016 年までに大学入試センター試験 (以下、センター試験とする) において高得点をマークし、2021 年までに東大二次試験に合格することを目指している。

大学入試問題を解くような大規模な解答器は、複雑で多くの要素技術が必要となることが考えられ、一つの研究チームでの単独での開発は困難である。そこで、多くのチームが参加したり、それぞれが得意な部分を出し合って、相互に成果を活用できるようにすることを目指し、ツールやデータの共有・組み合わせ・実行を容易にする統合研究基盤の構築が行われている [狩野 12]。

試験問題を解くのに応用可能な仕組みとして、既存の質問応答システム (以下、QA システムとする) の仕組みが応用できると考えられており、ベースシステムとして MinerVA[Mori 03, 石下 09] と Javelin[Shima 08] の 2 つの既存の QA システムのコンポーネント化を行った。システム開発者は、UIMA のフレームワークにより、これらのコンポーネントを自由に組み合わせたり、特定のコンポーネントを新規のものに入れ替えたり、改良したり、新たなコンポーネントを追加したりして用いることができる。

コンポーネントは、具体的には、図 1 に示すように、質問解析、文書検索、回答抽出、回答選択の 4 つである。質問解析では、入力された質問文の解析を行い、キーワードの抽出や質問が何を聞いているかを表す質問タイプの判定を行う。文書検索では、抽出されたキーワードを用いて質問に関連する文書を検索する。回答抽出では検索された文書から解候補となる文字列を抽出する。回答選択では、抽出された解候補の周りのキーワードや、質問タイプと一致するかなどの情報から、回答らしさのスコアが付けられ、回答らしさの高い解候補が回答として出力される。2 つの QA システム中のコンポーネントは、互換が可能である。本研究では、このコンポーネント化された既存の QA システムを利用して、大学入試問題に解答するベースシステムの開発を目指している。

連絡先: 国立情報学研究所, ishioroshi@nii.ac.jp

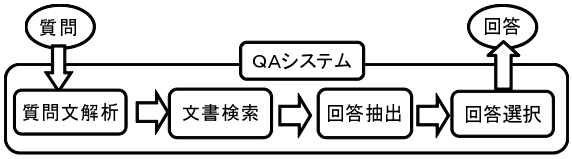


図 1: QA システムのコンポーネント

本稿では、大学入試問題として、センター試験に焦点を当て、その中でも特に、受験者の知識を問う問題が多く、QA システムにもっとも合致していると考えられる歴史の科目に焦点を当てた。本稿では、QA システムを用いてセンター試験の問題に解答することの実現性と課題を洗い流すことを目的として行った、センター試験の問の分析と予備実験について述べる。

センター試験の問は、既存の QA システムが想定する入力の形式とは異なっているため、問を解くのに適した質問文を生成する必要がある。また、センター試験の問には、空欄に入る語を選ぶ問題や、文の真偽を判定する問題など様々な形式があり、形式ごとに質問文の作り方を変える必要がある。そこで、本稿では、まず、センター試験の問の形式の分類を行った。次に、それぞれの形式に対して想定する質問文の作成方法がうまくいくかどうかを調べるために、センター試験の問から人手で質問文を作成し、QA システムに入力する予備実験を行った。QA システムの出力を分析することで、QA システムを用いてセンター試験の問題に解答することの実現可能性と課題、必要な知識リソースについての考察を行った。

2. 関連研究

歴史問題などの知識を問う問題に対するアプローチとして、含意関係認識を使ったアプローチと、世界史オントロジーを用いるアプローチなどがある [宮尾 12]。これらのアプローチでは、主に言明の真偽を問うタイプの問題に焦点を当てているが、本研究ではそれ以外のタイプの問題にも対応できるようにする。また、真偽を問うタイプの問題においても、これらの技術を本研究で用いる QA システムに導入することによって精度向上を狙うことが可能である。

また、金山ら [Kanayama 12] は、DeepQA[Ferrucci 12] を用いて言明の真偽判定をする手法を提案している。DeepQA は、アメリカのクイズ番組 Jeopardy!のクイズに答えるシステムである。本研究でも、言明の真偽判定に同様の手法をとることを検討している。ただし、金山らの研究では、センター試験の選択肢を人手で DeepQA が受け付ける文に変更しているが、本研究では自動で変換することを目指している。また、DeepQA では、Jeopardy!のクイズに対応するため、多数の質問タイプが用意されているが、本研究で用いる既存の QA システムでは人名、地名、組織名、数量といった代表的な質問タイプしか用意されていない点で異なる。

3. 大学入試問題の分析

分析対象として、センター試験の世界史 B の 2003 年から 2011 年までの奇数年の問題を用いた。センター試験の世界史 B の問題は約 36 問の問から成っており、各問には、指示文と選択肢が含まれており、一部の問にはそれ以外の文章や画像のデータも含まれている。問の例を図 2 に示す。本研究では、XML 化されたセンター試験データ [宮尾 12] を利用することを前提としている。

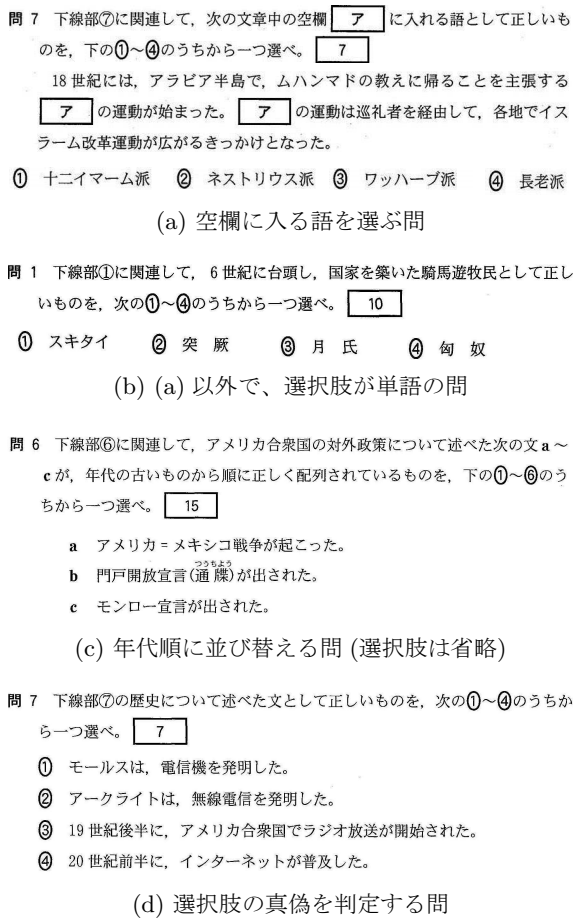
図 2 にあるように、センター試験の問の指示文は QA システムが回答できる質問文の形式とはなっていないため、指示文をそのまま QA システムに入力しても、適切な解答は得られない。問を解くために適した質問文に変換し、QA システムに入力し、その出力を集約することによって問に対する解答が可能となると考えられる。そこで、問を解くための、質問文への変換方法の観点から問の形式の分類を行った。分類の結果、センター試験の問は以下の 4 つの形式に分けられた。

なお、解くために画像や表の理解が必要となる問は、分析の対象から除いている。

1. 空欄に入る語を選択肢から選ぶ問
2. 1 以外で、選択肢が単語である問
3. 複数の出来事を年代順に並べ、正しい配列を選択肢から選ぶ問
4. 選択肢の内容の真偽を判定し、正しい (間違った) 内容の選択肢を選ぶ問

各分類の比率は、2 が 5%、1 が 12%、3 が 3%、4 が 74%(対象外が 6%) であった。4 が約 74%と最も多いが、高得点を目指すなら、それ以外の分類にも対応する必要があることが分かった。

質問文への変換方法としては、1 や 2 では、指示文や文章情報から、選択肢が解候補となるような質問文へ変換すればよいと考えられる。3 では、それぞれの事柄がいつ起こったかを問う質問文を作り、出力された年代順に並び替える方法が考えられる。4 では、金山らの手法 [Kanayama 12] と同様の手法を用いることができる。


問 7 下線部⑦に関連して，次の文章中の空欄 ア に入れる語として正しいものを，下の①～④のうちから一つ選べ。 7

18 世紀には，アラビア半島で，ムハンマドの教えに帰ることを主張する ア の運動が始まった。 ア の運動は巡礼者を経由して，各地でイスラーム改革運動が広がるきっかけとなった。

① 十二イマーム派 ② ネストリウス派 ③ ワッハーブ派 ④ 長老派

(a) 空欄に入る語を選ぶ問

問 1 下線部①に関連して，6 世紀に台頭し，国家を築いた騎馬遊牧民として正しいものを，次の①～④のうちから一つ選べ。 10

① スキタイ ② 突 厥 ③ 月 氏 ④ 匈 奴

(b) (a) 以外で、選択肢が単語の問

問 6 下線部⑥に関連して，アメリカ合衆国の対外政策について述べた次の文 a ～ c が，年代の古いものから順に正しく配列されているものを，下の①～⑥のうちから一つ選べ。 15

a アメリカ＝メキシコ戦争が起こった。

b 門戸開放宣言(通牒)が出された。

c モンロー宣言が出された。

(c) 年代順に並び替える問 (選択肢は省略)

問 7 下線部⑦の歴史について述べた文として正しいものを，次の①～④のうちから一つ選べ。 7

① モールスは，電信機を発明した。

② アークライトは，無線電信を発明した。

③ 19 世紀後半に，アメリカ合衆国でラジオ放送が開始された。

④ 20 世紀前半に，インターネットが普及した。

(d) 選択肢の真偽を判定する問

図 2: センター試験世界史 B の問の例

4. 予備実験

4.1 実験内容

4.1.1 実験で用いた QA システムの概要

使用可能な QA システムとして、MinerVA と Javelin[Shima 08] があり、MinerVA には、facitoid 型質問 *1 に回答する MinerVA-N[Mori 03] と、non-factoid 型質問 *2 に回答する MinerVA-D[石下 09] の 2 種類がある。それぞれのシステムのコンポーネントは入れ替えて使用することが可能であり、組み合わせを変えることで多数のシステムが使用できることになるが、本稿では単純に一つのシステムを用いて予備実験を行った。

予備実験では、MinerVA-N の文書検索部分のコンポーネントを入れ替えたものを用いた。MinerVA-N では、解答を抽出するための情報源として Web 文書が想定されており、文書検索エンジンとして Web 検索エンジンの API を用いている。予備実験では、試験問題に解答する観点から、歴史教科書と Wikipedia の文書を情報源として利用できるように、文書検索部分のコンポーネントを入れ替えて使用した。教科書は、東京書籍の歴史分野の教科書 (日本史 A、日本史 B、新選日本史 B、世界史 A、世界史 B、新選世界史 B) を利用した。一つの教科書を一文書とすると、一文書の文書量が膨大になってしまうため、節ごとに区切り、それぞれの節を一文書とした。wikipedia では、一つのエントリを一文書としている。文書検索エンジンは、indri*3 を用いた。

実験で用いた QA システムの構成図を図 3 に示す。

*1 人名数量などのなどの短い語句が回答となる質問
*2 定義や方法などの長い文章表現が回答となる質問
*3 http://www.lemurproject.org/indri.php

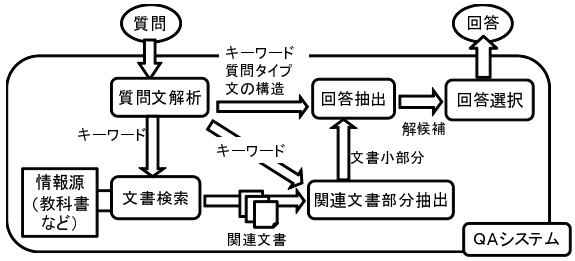


図 3: 実験で用いた QA システムの構成図

まず、質問解析で入力された質問文の解析を行い、キーワードの抽出及び質問が何を聞いているかを表す質問タイプの判定を行う。文書検索では、抽出されたキーワードを用いて質問に関連する文書を検索する。次に、検索された文書全体から解候補を抽出するのは時間がかかるため、検索された文書をまず数文の塊 (パッセージ) に分割し、パッセージの形態素解析や構文解析の結果から解候補となる文字列を抽出する。抽出された解候補には、質問文との構造の一致度や解候補の周辺に現れるキーワードの数、質問タイプと一致するかなどの情報から、回答らしさのスコアが付けられる。回答らしさのスコアや、同じ解候補が違う文脈から何回出てきたかの情報を用いて解候補の最終的なスコアを求め、スコアの大きい解候補を回答として出力する。

4.1.2 実験方法及び実験から得られた知見

本研究では、センター試験の問から自動で QA システムの入力となる質問文へ変換ことを目指しているが、図 2 の例で示したように、センター試験の問には、指示文の他に選択肢やそれ以外の文章などが含まれており、質問文を作る際に使える情報は様々である。予備実験では、人手で質問文を作成することで、どのような情報を用いて質問文への変換を行うと良いかの検討も合わせてすることとした。

試験問題として、2011 年のセンター試験世界史 B を用いた。大問 1 と大問 2 に含まれる 18 問の問を対象に、人手で質問文への変換を行った。各問を、3. 節のどの分類かを人手で判定し、質問文への変換を行った。

真偽を判定する問では、選択肢の文中に含まれる単語のうち、人名、地名、組織名、数量などの固有表現及び一般名詞 (名詞句) を疑問詞に置き換えることで質問文に変換した。質問文は、疑問詞に置き換える固有表現や名詞 (名詞句) の数だけ作成した。例えば、「パグウォッシュ会議で、科学者たちが核兵器の禁止を訴えた。」という言明の場合、「パグウォッシュ会議」「科学者たち」「核兵器」のそれぞれを疑問詞に置き換えた 3 つの質問文へと変換した。また、選択肢の文に指示文の一部を補うことによって、命題の真偽が判定できるようになる場合がある。選択肢の文のみで命題の真偽判定ができそうかどうかも人手で判断し、指示文で補う必要があると判断した時には必要な記述を補った上で、質問文へ変換した。

作成した質問文を QA システムに入力し、その出力を見ることで、問に解答できそうかどうかの検討をした。その結果、QA システムの精度が悪く、質問文への変換方法を検討する前に、QA システムを改良する必要があることが分かった。QA システムの出力を見ると、まったく的外れな解候補が多く含まれることが分かった。センター試験で問われている内容は、既存の QA システムが想定していなかった性質のものが多く含まれていることが要因として考えられる。次節において、既存の QA システムを試験問題の解答に利用する際に問題と考えられる点について述べる。

4.2 QA システムを用いたセンター試験の解答に向けた課題

本節では、試験問題を解くにあたって既存の QA システムを利用する際に問題となったと思われる点について述べる。予備実験を進めるうちに、以下のような点が問題点として浮かび上がった。

ドメイン特有の表記 「世紀」や「紀元前」といった歴史問題特有の年代表記に対応できていない。そのため、年代を問う質問文に対して適切な解候補が抽出されないことがあった。

ドメイン特有の質問カテゴリ 「○○の宗派は何派ですか？」という質問文に対し、宗派ではない解候補が抽出されるなど、質問文で聞いている内容と違うカテゴリの解候補が抽出されてしまっていた。適切な質問カテゴリが設定されていないことが要因として挙げられる。

本稿では、特に 2 つ目の問題点に着目する。質問タイプとは、質問文が聞いているもののカテゴリを表す情報で、例えば、「○○したのは誰ですか。」という質問文では、質問タイプは「人名」となる。既存の QA システムでは、質問タイプとして、「人名」、「地名」、「組織名」、「数量 (年代、高さなど)」があり、それ以外のものは「質問タイプなし」と判定される。「質問タイプなし」では、どのようなカテゴリの語でも解候補として扱われるため、例えば質問タイプが設定されていない「宗派」を問う質問文では、「宗派」以外のカテゴリの語が適切な解候補として出力されてしまうなど、解候補抽出時にエラーが生じる可能性が高い。歴史の試験問題で頻出しそうな「王朝」や「戦争」などといった質問タイプを新たに設けることで、問題の改善が見込まれる。質問タイプの拡張に対する検討を次節で詳しく述べる。

4.3 QA システムの質問タイプの拡張に対する検討

QA システムの質問タイプを増やすためには、「質問文がどの質問タイプに分類されるか」と、「解候補として抽出された語がどの質問タイプに対応するか」の両方が判定できる必要がある。質問文の質問タイプ分類に関しては、ルールベースや機械学習などの手法があり、比較的容易に対応が可能であると考えられる。解候補として抽出された文字列のカテゴリを求めることについては、「人名」、「地名」、「組織名」、「数量表現」などは固有表現抽出器で抽出が可能であるが、前述した「宗派」や「王朝」などは、固有表現抽出器では対応していない。質問タイプごとに個別に辞書などを用意して対応する必要があると考えられるが、個別に辞書を用意するのはコストがかかる。そのため、質問タイプを増やすにあたっては、どのような質問タイプを増やすべきかをよく検討する必要がある。

我々は、センター試験に出てくる用語のカテゴリに着目した。あるカテゴリに含まれる用語の頻度が高ければ、そのカテゴリを質問タイプすると有効に働くのではないかと考えた。そこで、以下の方法で、センター試験に出てくる用語の分析を行った。

1. まず、準備として、wikipedia データ解析ツール Wik-IE[*4] を用い、wikipedia の各エントリのタイトルとカテゴリのひも付けを行い、「語-カテゴリ DB」を作成した。

*4 http://wik-ie.sourceforge.jp/

歴史　人物　年生　戦争　年没　民族　ヨーロッパ史
都市　国際関係　言語　世紀　共和国　国家　姓
学者　男性名　宗教　法　経済史　地形　政治　文化

図 4: センター試験世界史 B に頻出する語のカテゴリ

人物　年生　存命人物　選手　企業　教員　年没　歴史
姓　国会議員　野球選手　俳優　食文化　文化　建築物
地理　東証一部上場企業　都市　参議院議員　地名　法

図 5: 毎日新聞に頻出する語のカテゴリ

2. 次に、専門用語自動抽出システム「termex」*5 を用いてセンター試験から用語を抽出した。

3. 抽出された用語が「語-カテゴリ DB」にあった場合、そのカテゴリを取得した *6。

4. すべての用語に対してカテゴリを取得し、カテゴリの頻度を求めた。

センター試験として、世界史 B の 1997 年〜2011 年の奇数年の問題を用いた。分析の結果、異なり数 3995 の用語が抽出され、そのうち 2292 の用語でカテゴリを取得できた。センター試験世界史 B に頻出した語のカテゴリの例を図 4 に示す。また、比較のために、毎日新聞 (2002 年前半) を使った結果を合わせて示す (図 5)。

図 4 と図 5 を比較すると、世界史 B では新聞記事に比べ、「戦争」、「世紀」、「宗教」といったカテゴリが頻出しやすいということがわかる。また、どちらも人物に関するカテゴリが多く見られるが、新聞記事で見られる「存命人物」、「選手」、「国会議員」、「俳優」といったカテゴリが世界史 B では見られず、実際に現れる人物名の性質が違うことが伺える。

このように、ドメインによって、頻出するカテゴリに違いがあることが分かった。ドメイン特有の語のカテゴリが新たな質問タイプとして有効である可能性がある。本研究では、世界史以外の科目についても QA システムを利用することを目指しているため、ドメイン特有の語のカテゴリを自動で収集する手法を検討中である。

また、新たな質問タイプを設定した時には、解候補として抽出された文字列がどの質問タイプに適合するかを判断するため、各質問タイプに属する語を集めた辞書などが必要である。各質問タイプに対応した辞書の構築については、世界史については世界史オントロジー [宮尾 12] を用いることが考えられるが、他の科目についてもオントロジーが使えるとは限らない上、辞書をすべて人手で構築するのはコストがかかる。そのため、辞書の構築については、半自動の手法を検討中である。

*5 http://gensen.dl.itc.u-tokyo.ac.jp/win.html

*6 一つの語に対して複数のカテゴリがある場合、そのすべてを採用した。また、「○○の歴史」は「歴史」に統合するなど、一部のカテゴリについては簡単な一般化を行っている。

5. 終わりに

本稿では、QA システムを用いてセンター試験に解答することの実現可能性と課題の検討を目的とした、予備実験について述べた。

センター試験の問を人手で質問文へ変換し、QA システムに入力する予備実験の結果、センター試験への解答に利用するためには、既存の QA システムの精度に問題があることが分かった。QA システムの質問タイプが足りないことが、精度が悪い原因の一つ思われることから、質問タイプを増やすことに対する検討を行った。どのような質問タイプを増やすのか、新しい質問タイプに対応するための辞書などをどのように構築するのかが、今後の課題である。

今後は、既存の QA システムに見つかった問題点を解決すると共に、センター試験の問から質問文へ自動で変換する手法の検討を行う予定である。さらに、センター試験の入力から解答までを自動で行うシステムを作成し、その評価を行う予定である。

謝辞

本研究の実施にあたっては、東京書籍の教科書データ及びまた 2002 年度版の毎日新聞を使用した。東京書籍様と毎日新聞社様に感謝いたします。

参考文献

[Ferrucci 12] Ferrucci, D. A.: Introduction to "This is Watson", *IBM Journal of Reserch and Cevelopment*, Vol. 56, No. 3.4, pp. 1:1–1:15 (2012)

[Kanayama 12] Kanayama, H., Miyao, Y., and Prager, J.: Answering Yes/No Questions via Question Inversion, in *Proceedings of COLING 2012*, pp. 1377–1391 (2012)

[Mori 03] Mori, T., Ohta, T., Fujihata, K., and Kumon, R.: An A* Search in Sentential Matching for Question Answering, *IEICE Transactions on Information and Systems*, Vol. E86-D, No. 9, pp. 1658–1668 (2003), Special Issue on Text Processing for Information Access

[Shima 08] Shima, H., Lao, N., Nyberg, E., and Mitamura, T.: Complex cross-lingual question answering as a sequential classification and multi-document summarization task, in *Proceedings of the 7th NTCIR Workshop Meeting*, pp. 33–40 (2008)

[宮尾 12] 宮尾 祐介, 川添 愛 :「大学入試問題を解く」ことから見える言語, 知識, 世界理解に関する研究課題, 人工知能学会論文誌 (2012)

[狩野 12] 狩野 芳伸 : 統合研究基盤:質問応答システムの互換コンポーネント化による再利用性向上と開発自動化支援, 人工知能学会論文誌 (2012)

[新井 12] 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)

[石下 09] 石下 円香, 佐藤 充, 森 辰則 : Web 文書を対象とした質問の型に寄らない質問応答手法, 人工知能学会論文誌, Vol. 24, No. 4, pp. 339–350 (2009)