

早押しクイズの解候補の絞り込み

杉山 宏輝^{1,a)} 角 康之^{1,b)}

概要: 本研究は、早押しクイズに勝つための人工知能 (AI) システムの開発が目的である。近年のクイズ人気に伴い、早押しクイズ AI も需要が高まっていると考える。クイズを AI に解かせるという試みは以前より盛んにおこなわれているが、早押しという点においては未だ人間のクイズプレイヤーには及ばないことも多い。本研究では、序盤で答えがいくつかの確実な選択肢 (解候補) に絞られる問題において、その解候補から妥当性を考慮して正解を推定することを試みる。早押しクイズの作問の際に意識される解の妥当性に着目し、問題文の前半で得られた解候補を、妥当性を構成する各要素に分類・評価することで、問題文の後半を聞くことなく正解を導き出すことができると期待する。

1. はじめに

本研究は、早押しクイズに勝つための人工知能 (AI) システムの開発が目的である。本研究で開発するシステムでは、従来のクイズ AI のように十分な情報が与えられたのち、最適解を見つけるようなものとは異なり [1], お手付きのリスクを冒しながらも、より早い段階での解答を目指す。近年、クイズの人気は高まり、多くのテレビ番組やメディアで目にする機会が増えてきた。特に早押しクイズと呼ばれる形式は、競技化され、各地で大会なども開かれている。それに伴い、早押しクイズ AI の需要も高まっていると考えられる。

クイズを AI に解かせるという試みは Watson[2] をはじめ、盛んに研究されてきた。しかし、その多くは正解推定のために十分な文量と単語が必要であり、早押しという点においては人間のクイズプレイヤーに及ばないことも多い。クイズプレイヤーが見せる異常なまでの押しの速さは、早押しクイズは妥当性を考慮して正解を設定しているという経験に基づくものであることが多々ある [3]。例えば、問題『『歳時記』における季語の5つの分類とは、「春」「夏」「秋」「冬」と何でしょう?』において、クイズプレイヤーはほぼ確実に『『歳時記』における季語の5つとは』で押し、『新年』と答える。これは慣れ親しんだ「春」「夏」「秋」「冬」のほかに「新年」が歳時記では一季節として分類されているという意外性により、他の解候補よりも「新年」を正解とするのが自然だという推測からくるものである。このように、早押しクイズには妥当性を考慮した問題作りをする

という、暗黙的なルールや作問者の意図が介在する。

そこで、本研究ではこのような作問のメタ的な観点から答えを絞ることで、より早い段階での解答を実現する。最終的な目標は、自ら早押しクイズに参加し、人間を相手に勝つ AI の開発であるが、その途中段階として、択一問題における正解推定の手法を検討する。ここでいう択一問題とは、[W とは、w1、w2、...、wN-1 と何でしょう?、wN] のように、グループ W が問題文前半で示され、後半にその要素が羅列されていく問題を指す。問題の例を表 1 に示す。

2. 関連研究

早押しクイズにおける関連研究として、Yamada & Tamaki & Shindo & Takefuji (2018) [4] がある。Yamada らは問題の構成を予測する学習モデルと、答えの分類 (人・物・場所など) を予測するモデル、そして従来の情報検索モデルを合わせた AI を開発した。開発された AI は人口知能に関する国際会議である NIPS 2017 で行われた質問応答のコンペティションで優勝し、人間のクイズ王 6 名に対しても勝利を収めた。一方で、Quiz Bowl と呼ばれる、複数のパッセージから 1 つの問題文を構成する形式のクイズデータセットにおいて、文のフルセットが与えられた場合の精度は 97% と非常に高かったが、単一の文では 56% にとどまった。これは推定の精度が問題文中の特徴的な単語量に依存していることに由来するものであると考察される。本研究では、択一問題において序盤に得られる解候補の関係性から、後半の情報を聞くことなく正解を推定することを目指す。

また、問題文の構造を利用した研究に橋元ら (2021-2022)

¹ 公立はこだて未来大学

^{a)} h-sugiyama@sumilab.org

^{b)} sumi@acm.org

表 1: 択一問題の例

問題	答え
1 ベートーベンの三大ピアノソナタに数えられる 3 つとは、『悲愴』、『月光』と何でしょう？	『熱情』
2 『新約聖書』の 4 つの福音書とは、『マタイの福音書』『マルコの福音書』『ルカの福音書』と何でしょう？	『ヨハネの福音書』
3 夏目漱石の作品のなかで、タイトルと書き出しが同じものは『吾輩は猫である』と何でしょう？	『硝子戸の中』

[5,6] がある。彼らはパラレル問題の構造を調査し、パラレル問題を成立させている対比関係を明らかにした。パラレル問題とは「アイルランドの首都はダブリンですが、アイスランドの首都はどこでしょう？」のような早押しクイズ特有のひっかけ問題のことである。結果として、対比関係を 1) 対義 2) 連続 3) 慣例的に一対一対応 4) 共通の狭い集合 5) 音の類似・部分一致の 5 種類に整理し、それらを元に問題文の途中まで解答するシステムおよび問題文を生成するシステムを構築することができた。このことから、早押しクイズ特有の要素に着目した推定は有効であると考えられる。

3. 妥当性の分類

妥当性を構成する要素を分類するために、クイズ大会の過去問やアプリで出題される問題に含まれる択一問題を調査した結果、以下の 4 つに大別されることが分かった。

仲間外れ 解候補のうち、他に共通している上位概念がないもの

文字列の類似度 解候補のうち、文字の表記や音が 1 つだけ似ていないもの

難易度 解候補のうち、1 つだけ単語の難易度が高いもの

順序 日付など何かしらの番号が付与されている単語の順番

各要素に分類される問題文の例を表 2 に示す。

(a) は「ジャイアントパンダ」「オカピ」「コビトカバ」の 3 種がまとめて世界三大珍獣と呼ばれるため、ボンゴが仲間外れとなる。

(b) は「アルニラム」「アルニタク」という似た音の単語と比べ、「ミンタカ」は明らかに音が異なっていることがわかる。

(c) は「フランス料理」「中華料理」に比べ「トルコ料理」はあまり一般的ではなく、単語としての難易度が高いといえる。

(d) は葵祭は 5 月、祇園祭は 7 月、時代祭は 8 月に開催されるため、葵祭<祇園祭<時代祭という順序関係が存在する。

4. 評価方法

前項で得られた分類ごとに、それぞれ評価する手法を提案する。

仲間外れ

仲間外れであるものは、他の候補間で共通している上位語がないものである。よって上位-下位関係の抽出を行い、上位語/下位語ペアを作成することで、共通する上位語をカウントすることにより評価する。

文字列の類似度

文字列の類似度を測る代表的な手法として編集距離がある。これは 1 文字の挿入・削除・置換によって、一方の文字列をもう一方の文字列に変形するのに必要な手順の最小回数をカウントするものである。例えば“AAA”と“ABA”を比べた場合、2 文字目の‘A’を‘B’に置換するだけでよいため編集距離は 1 となる。しかし、“ABCXYZ”と“XYZDEF”を比べた場合、距離は 6 となり“XYZ”が部分一致しているにもかかわらず、位置が異なるためまったく別の文字列として扱われてしまう。そこで、本手法ではゲシュタルトパターンマッチングを使用する。2 つの文字列 S_1, S_2 類似度 D_{ro} は、式

$$D_{ro}(S_1, S_2) = \frac{2L}{|S_1| + |S_2|}$$

で表され、完全に一致する場合は 1、いずれの部分列も一致しない場合は 0 となる。ここで、 L は一致文字列の長さを表す。

難易度

単語の難易度を、その単語がどれだけ知られているか（知名度）として考え、指標とする。知名度を測る手法として、アンケートをとるといった方法が挙げられるが、あらゆる単語についてアンケートをとるのは現実的でない。そこで、本研究では単語のレア度を知名度とし、あらゆる文脈においてその単語が広く使われているかどうかを指標とする。具体的な方法として IDF (Inverse Document Frequency) を用いる。IDF は文書の総数をその単語を含む文書の数で割った商の対数を取ることで得られる、単語が普遍的かどうかを表す指標の 1 つである [7]。具体的には、以下の式で表される。

$$idf(q_i) = \log \frac{N}{n_i + 1}$$

ここで、 q_i をある単語、 N を総文書数、 n_i を q_i が含まれる文書数とする。値が大きいほどレア度が高く、小さいほど低くなる。

表 2: 妥当性の分類と例題

分類	問題	答え
a. 仲間外れ	世界四大珍獣に数えられる 4 種類の動物とは、ジャイアントパンダ、コビトカバ、オカピと何でしょう？	ボンゴ
b. 文字列の類似	オリオン座の中央にある 3 つ星を構成しているのは、アルニタク、アルニラムと何でしょう？	ミンタカ
c. 難易度	一般に「世界三大料理」と呼ばれる料理とは、フランス料理、中華料理と何料理でしょう？	トルコ料理
d. 順序	京都三大祭といえば、5 月の葵祭、7 月の祇園祭と 10 月にある何祭でしょう？	時代祭

順序

ある単語に紐づく数字の大小を比較することで評価する。

5. 解答システム構築

5.1 使用データ

知識源として 2022 年 11 月時点の Wikipedia ダンプデータ^{*1}を使用する。xml 形式のデータとなっているためパースし、各記事のタイトルと本文を抜き出して DB 化した。ただし、'Wikipedia:', 'File:', 'Portal:', 'Template:', 'Mediawiki:', 'User:', 'Help:', 'Book:', 'Draft:', 'Module:', 'Timedtext:' から始まる記事は除外した。本文は必要に応じてクリーニングした。

5.2 上位語/下位語ペアリストの作成

上位一下位関係を抽出する研究としては墨田ら (2008)[8]の研究がある。彼らは Wikipedia の記事構造を知識源として、単語の上位下位関係を抽出した。本研究でも同様に記事構造から上位下位関係を抽出する。Wikipedia の本文は

”日本三大一覽” (にほんさんだいいちらん) は、

[[日本]] を代表する三つのものの一覽。

== 総記 ==

See also—御三家

== 思想・哲学 ==

* [[寛政の三奇人]]

** 林子平、高山彦九郎、蒲生君平 Sfm—朝日新聞社知恵蔵編集部—2003—p=258

* [[寛政の三博士]] (寛政の三助)

** 尾藤二洲、柴野栗山《古賀精里、岡田寒泉》Sfm—広辞苑—2018—p=664

のようになっているため、

* 上位語

** 下位語

として抽出した。上記の引用の範囲では

寛政の三奇人/林子平

寛政の三奇人/高山彦九郎

寛政の三奇人/蒲生君平

寛政の三博士/尾藤二洲

寛政の三博士/柴野栗山

寛政の三博士/古賀精里

寛政の三博士/岡田寒泉

が得られる。ただし、択一問題の性質を顧みて、「世界三大〇〇一覽」のような名数をまとめた記事のみを対象とした。

5.3 単語/idf ペアリストの作成

全記事の本文を単語分割し、ある単語について idf を計算し、単語/idf ペアを作成した。形態素解析には MeCab^{*2}、辞書として mecab-ipadic-NEologd^{*3}を使用した。

5.4 単語/数値ペアリストの作成

Wikipedia では葵祭は以下のように説明されている。

葵祭 (あおいまつり、正式には賀茂祭) は、京都市の賀茂御祖神社 (下鴨神社) と賀茂別雷神社 (上賀茂神社) で、5 月 15 日 (陰暦四月の中の酉の日) に行なわれる例祭。石清水八幡宮の南祭に対し北祭ともいう。平安時代、「祭」といえば賀茂祭のことを指した。

上記のように、何かしらの順序があるものの説明には対応する数値が記載されていることが期待できる。("5 月 15 日 (陰暦四月の中の酉の日) に行なわれる")。よって概要文から数値を抽出することとした。比較される主要な数値として日付、番号が挙げられる。そこで、以下に示される正規表現を用意し、Wikipedia 記事のタイトルとその概要文において、概要文内にある数値を抽出することでタイトル/数値のペアを作成した。マッチオブジェクトの探索は以下の順番通りに走査し、最初にマッチした数値を抽出した。¥d は半角数字を、+ は直前の文字の連続を表し、丸括弧で囲んだ部分を抽出する。

1. { 原子番号 (¥d+)
2. { 第 (¥d+)
3. {(¥d+) 年 }
4. {(¥d+) 月 }

^{*1} <https://dumps.wikimedia.org/jawiki/latest/jawiki-latest-pages-articles-multistream.xml.bz2>

^{*2} <https://taku910.github.io/mecab/>

^{*3} <https://github.com/neologd/mecab-ipadic-neologd>

5.5 確率ベクトル

システムに入力された解候補を妥当性の各要素ごとに評価した値で確率ベクトルにすることを考える。仲間外れは共通する上位語をカウントした数のベクトル。文字列の類似度はゲシュタルトパターンマッチングの和平均のベクトル。難易度はidfを並べたベクトル。順序は数値を比較した際、一番値が大きいものを1、それ以外を0としたベクトルにし、それぞれソフトマックス関数にかけて確率分布に正規化した。

5.6 解の選択

前節で述べた4つの確率ベクトルからその解候補がどの要素に分類される問題なのかを検査し、採用する確率ベクトルを選択する。適さない分類の確率ベクトルは値が平均的になると期待する。よって各ベクトル毎の分散を計算し、分散が最も大きい確率ベクトルを採用することとした。最終的な解として、採用した確率ベクトルの中で最大値をとる単語を出力する。

6. 評価実験

クイズを題材にした日本語 QA データセットである JAQKET[9] から抽出した択一問題全 293 問を使用し、問題文前半で得られる確実な解候補を入力した場合の正解率を算出した。確実な解候補とは、例えば問題「一般に『世界三大料理』と呼ばれる料理とは、フランス料理、中華料理と何料理でしょう?」であれば、「一般に『世界三大料理』の時点で得られる、「フランス料理」「中華料理」「トルコ料理」の3つのような解候補のことを指す。

また、解候補の単語が Wikipedia において 1 対 1 でリダイレクトされている場合は、リダイレクト先のタイトルをその単語として扱った。ベースラインとして、解候補数に応じた期待値の平均を算出している。

正解数 137

正解率 0.468

期待値 0.370

結果、期待値は 37.0%、本システムは 46.8%の正解率となり、期待値と比べて約 10%増となった。

最後に、解の確率が変動する様子を見せる。

図 1 はそれぞれ

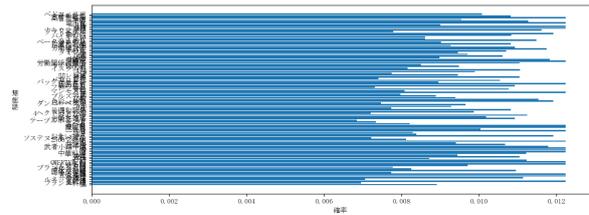
1a. 一般

1b. 一般に

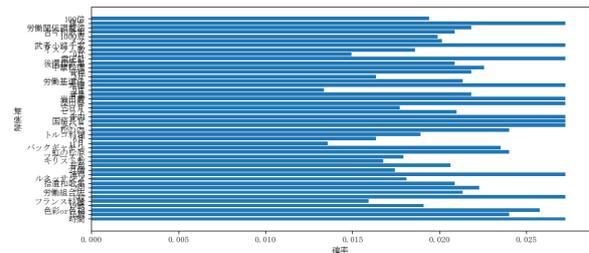
1c. 一般に世界三大

1d. 一般に世界三大料理

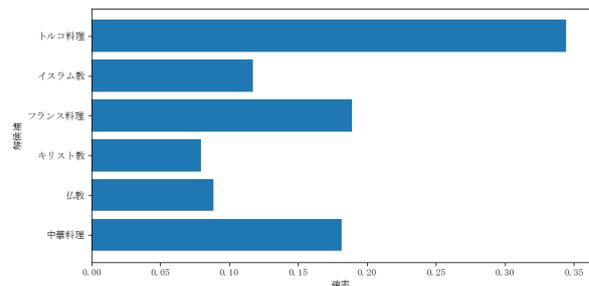
の4つのパターンにおいて得られる解候補と確率をグラフ化したものである。解候補は JAQKET の問題をデータ拡張し、先頭から全文一致する問題の解をその時点での解候



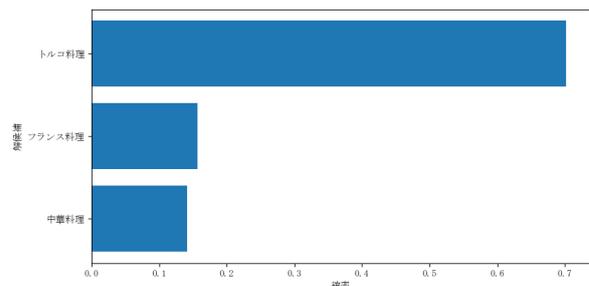
(a) 解候補数 99



(b) 解候補数 46



(c) 解候補数 6



(d) 解候補数 3

図 1: 解の確率変動の様子

補として扱った。

7. まとめ

早押しクイズの択一問題において、解としての妥当性を考慮した推定をするシステムを構築した。期待値によるベースラインと比べた正解率は約 10%増であり、課題の残る結果となった。1つは、評価基準の異なる値を確率ベクトルとしてそのまま比較しているため、公平でないといった

ことが挙げられる。実際、idf による確率ベクトルが優先して採用される傾向があり、誤った解が選択されることが多かった。また、単語/数字ペアにおいて、余計な数値が抽出されているせいで誤った順序評価がなされている例もあった。さらに、現状は抽出されたすべての択一問題の解候補で正解率を出しているが、妥当性の要素で分類できない、確率がフラットな問題も含まれている。今後の課題として、各評価の確率にバイアスをかけて適切な解が選択されるようにする、データ構築の改善、問題数の増加などをしていきたい。

参考文献

- [1] Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, Hal Daume III. “A Neural Network for Factoid Question Answering over Paragraphs”. EMNLP, ACL, 2014, 633-644.
- [2] David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty. “Building Watson: An Overview of the DeepQA Project”. AI Magazine, AAAI, 2010, 31(3), 59-79
- [3] 伊沢拓司. クイズ思考の解体. 朝日新聞出版, 2021, 479p.
- [4] Ikuya Yamada, Ryuji Tamaki, Hiroyuki Shindo, Yashiyasu Takefuji. “Studio Ousia’s Quiz Bowl Question Answering System”. The NIPS ’17 Competition: Building Intelligent Systems. Springer, 2018, 181-194
- [5] 橋元佐知, 佐藤理史, 宮田玲, 小川浩平. 競技クイズ・パラレル問題の基本構造と文型. 言語処理学会第 27 回年次大会発表論文集, pp. 1420–1424, 2021.
- [6] 橋元佐知, 佐藤理史, 宮田玲, 小川浩平. 早押しクイズのパラレル問題の自動生成. 言語処理学会第 28 回年次大会発表論文集, , 2022.
- [7] Salton, G., & McGill, M. J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.
- [8] 墨田飛鳥, 吉永直樹, 島澤健太郎, 萬成賢太郎. Wikipedia からの大規模な上位下位関係の獲得. 言語処理学会第 14 回年次大会予稿集, 2008.
- [9] 鈴木正敏, 鈴木潤, 松田耕史, 西田京介, 井之上直也. JAQKET: クイズを題材にした日本語 QA データセットの構築. 言語処理学会第 26 回年次大会, pp. 237-240, 2020.