

ニュース記事を用いた時事問題自動作問システム

久田 将史 前田 亮

立命館大学情報理工学部 〒525-8577 滋賀県草津市野路東 1-1-1

E-mail: is0230kx@ed.ritsumei.ac.jp, amaeda@is.ritsumei.ac.jp

あらまし コンピュータが問題を出題するシステムの多くが、あらかじめ用意しておいた問題文と解答を用いて出題している。コンピュータが自動で作問するシステムがあれば、人間が問題を考える手間も必要なくなる。自動作問システムが有効である例として、時事問題が挙げられる。時事問題は頻繁に問題の更新が必要であり、問題の作成に大きな手間がかかる。そこで、本論文では4択式時事問題の自動作問手法を提案する。まず、新聞記事コーパスから、記事前文を抽出して、問題の解答となる単語を選択し、穴埋め式の問題文を作成する。次に、解答となる単語のベクトル類似度上位語から解答の類義語を抽出し、誤答を作成する。そして、解答の選択肢を4択で出題する。評価実験では、本提案手法を用いて作成した時事問題の正答率と適切さを調査し、システムを評価した。ユーザが「良問」・「普通」と答えた問題を適切な問題とすると、実験の結果、作成した問題のうち80%が適切な問題であった。

キーワード 自動作問, 分散表現, word2vec

1. はじめに

ユーザが投げかけた質問文を解析し、質問に答えるシステムは質問応答システムと呼ばれている。質問応答システムは、Siriなどのスマートフォンアプリにも使われており、人々の生活にも浸透し始めている。一方で、コンピュータが質問を作成し、ユーザに出題するシステムは全く浸透していない。コンピュータが問題を出題するシステムの多くは、あらかじめ人間が作成した問題文と解答を用いて出題している。コンピュータが自動で作問するシステムができれば、問題を考える手間も必要なくなる。頻繁に問題の更新が必要な時事問題に対して、自動作問システムは特に有効であると考えられる。

そこで、本論文では、入力した文章から自動で作問するシステムを構築し、ニュース記事から時事問題の作問を行う手法を提案する。本論文は、時事問題の作成支援を目的とし、自動作成した問題が良問であるとユーザが感じる作問システムの構築を目標とする。

2. 関連研究

2.1 珠算における自動作問システムの構築

野崎ら[1]は、ユーザのレベルに応じて珠算の問題を作問するシステムの設計・開発を行うことを目的として、日本珠算連盟の検定試験と競技大会の問題のうち、見取算と見取暗算を作問対象として、作問システムを生成した。

そろばんを弾くときの珠の動きを「運珠」、珠を動かす指の動きを「運指」と呼ぶ。野崎らは、足し算と引き算における、珠の動きと運指数を16種類に分類し、

日本珠算連盟の問題の調査を行い、作問規定を設定した。そして、6桁の数字を生成し珠算問題を作成した。作成した問題はMicrosoft Excelに自動で生成される。また、運指の回数を考慮することで、問題の難易度を設定することができる。

野崎らの研究の作問対象は、珠算問題であるため、数字を生成して作問するシステムである。本研究では、時事問題を作成するため、数字ではなく文章問題を作成する。そのため、形態素解析を行い、品詞を考慮した上で問題を作成する。

2.2 連鎖型パズルゲームにおけるパズル問題の自動創作

山崎[2]は、あらかじめ配置されているパネルを決められた手数以内にすべて消す問題を出題するパズルゲームの自動作問を行った。このパズルゲームには5種類の図形が存在し、同じ種類の図形が縦もしくは横に3枚以上揃うことで消滅する。パズルゲームは、ランダムに図形を配置しているだけでは、図形を3枚以上揃える手段がなくクリアできない問題が作成されてしまう。そのような問題が作成されないように、図形の配置パターンを用いて問題を作成している。

このように、数字や図形を扱うパズル問題を作問するシステムの研究は他にも行われている。本研究では、数字や図形を扱うのではなく、日本語の文章問題を作問する新たなシステムを提案する。また、時事問題を自動で作問する新たなシステムを提案する。

3. 提案手法

本章では、本論文で提案する、ニュース記事を用いた4択式時事問題作成の手法について述べる。

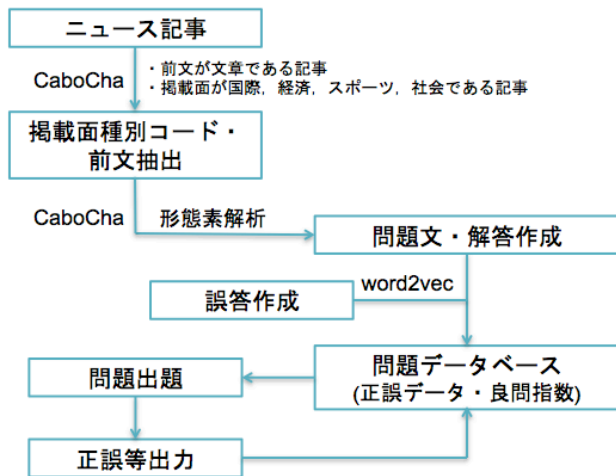


図1: システムの概要図

3.1 システム全体の概要

本論文で提案するシステムの概要を図1に示す。本提案システムでは、日本語の係り受け解析や形態素解析を行うソフトウェアであるCaboCha[3]を利用して、新聞記事コーパスから記事前文と掲載面種別コードを抽出し、形態素解析を行い、問題文と解答を作成する。次に、単語のベクトル化を行い、類似度が高い単語を抽出できるword2vec[4]を利用する、コサイン類似度上位の単語は類義語であるため、解答の単語のコサイン類似度上位語を誤答の選択肢として加えて、4択問題で出題する。最後に、掲載面、問題文、選択肢をデータベースに保存する。システムを利用するユーザは、システムを起動すると、国際・経済・スポーツ・社会いずれかのニュース掲載面を選択し、問題に答える。問題の正誤判定と解答が示された後に、ユーザは「良問」「普通」「悪問」から選び、自動作成された問題の問題文、解答部、選択肢が時事問題として相応しいか評価する。問題出題と評価を繰り返し、「悪問」と評価された数が「良問」と評価された数よりも一定回数多くなった問題は、出題されなくなる。

3.2 作問対象とする新聞記事

作問対象となる新聞記事の抽出方法を説明する。毎日新聞の全文記事データ集である、毎日新聞データ集[5]を例として説明する。毎日新聞データ集には、図2のようにタグが付与されている。まず、記事前文（第一段落）のタグである「\S2\」と、掲載面種別コードのタグである「\AD\」の行を抽出する。

毎日新聞データ集には、図3のように問題文として

不適切な、文章になっていない前文がある。そのため、前文に句点（。）が含まれている記事のみを抽出対象とした。また、本研究では掲載面が経済、国際、スポーツ、社会の4種類の記事を作問対象としたことから、毎日新聞データ集の場合、掲載面種別コード（\AD\）が経済（07）、国際（08）、スポーツ（35）、社会（41）である記事を作問対象とした。記事前文と掲載面種別コードの抽出例を図4に示す。

```
\ID\00000370
\C0\110101037
\AD\07
\AE\Y
\AF\110101M09
\T1\エストニア：きょうユーロ加盟 旧ソ連国初、「独自通貨、困難」
\S1\ '11. 1. 1 朝刊 9頁 写真有 (全242文字)
\S2\ 【ロンドン会川晴之】91年に旧ソ連から独立したバルト沿岸の小国・エストニアが1月1日にユーロに加盟。旧ソ連諸国初のユーロ加盟となる。欧州共通通貨・ユーロの導入国は、これで17カ国となる。エストニアのユーロ加盟は、今年7月の欧州連合（EU）理事会で承認された。
\T2\ 【ロンドン会川晴之】91年に旧ソ連から独立したバルト沿岸の小国・エストニアが1月1日にユーロに加盟。旧ソ連諸国初のユーロ加盟となる。欧州共通通貨・ユーロの導入国は、これで17カ国となる。エストニアのユーロ加盟は、今年7月の欧州連合（EU）理事会で承認された。
\T2\ エストニアは、独立前は、旧ソ連（現ロシア）のルーブルを使用。独立後は独自通貨クロウンを導入した。同国のリグ財務相は、ロイター通信などに「小国が、独自の通貨制度を維持するのは難しい」と、ユーロ加盟の理由を述べている。
\KA\FSU
\AA\FSU
\KA\エストニア
\AA\エストニア
```

図2: 新聞記事データの一例

```
\T1\新成人：124万人、総人口比初の1%割れ 1次ベビーブーム世代の半数
\S1\ '11. 1. 1 朝刊 2頁 写真無 (全368文字)
\S2\ ◇4年連続で過去最低

\T1\「働きの時代」に：第1部・ものづくりの現場から／3 海渡る日本人技術者
\S1\ '11. 1. 3 朝刊 1頁 写真有 (本文なし)
\S2\ <働きの時代>

\T1\★ニッポン再生：突破口を探る 小川弘一・東京大特任教授に聞く
\S1\ '11. 1. 3 朝刊 6頁 写真有 (本文なし)
\S2\ 【現在著作権交渉中の為、本文は表示できません】
```

図3: 作問が不可能な前文の例

```
\AD\07
\S2\ 【北京・共同】中国共産党機関紙、人民日報は31日、一部地域で配られた30日付紙面で温家宝首相の名前を「温家宝」と誤って印刷したため、担当者らの処分を検討していることを明らかにした。同紙は中国で最も権威のある新聞として字句の使用にも厳格なことからインターネット上で話題となり、関係者計17人が降格などの処分を受けたとの情報も流れている。
\AD\07
\S2\ 【ソウル大澤文護】韓国の李明博（イミョンバク）大統領は31日、文化体育観光相に鄭炳国（チョンビョングク）ハンナラ党国会議員（52）、知識経済相に崔重鼎（チェジュンギョン）大統領経済首席秘書官（54）、監査院長に鄭東基（チョンドンギ）前大統領民政首席秘書官（57）を指名するなどの内閣人事を発表した。国会聴聞会の同意を経て正式任命される。10年8月に、知識経済、文化体育観光相閣僚に、別の候補者が指名されたが、不動産投機などの指摘を受け指名辞退し、人事は白紙に戻っていた。
\AD\07
\S2\ 【ソウル西原真一】韓国の李明博（イミョンバク）大統領は1日付で「新年の辞」を発表、11年には「必ず朝鮮半島の平和を成し遂げ、経済も継続して成長させていくことができると確信する」と強調した。
```

図4: 抽出した記事例

問題：安倍晋三内閣総理大臣（首相）は2014年9月、第2次安倍内閣として初めて、
【 】に踏み切り、国務大臣の一部を交代させました。
選択肢：① 組閣 ② 内閣改造 ③ 総辞職 ④ 解散

問題：政府は、2020年までに外国人観光客を【 】人まで増やす目標を立てています。
選択肢：① 2万 ② 20万 ③ 200万 ④ 2000万

問題：福井県にある【 】原発3、4号機の運転差し止めを住民が求めた裁判で、福井地方裁判所は2014年5月、この原発を運転しないよう関西電力に命じる判決を出しました。
選択肢：① 女川 ② 大飯 ③ 浜岡 ④ 川内

図5: ニュース時事能力検定の問題例(3級)

3.3 問題文と解答の作成

次に、抽出した記事から問題文を作成する。どのような問題が時事問題として出題されているか、時事問題を扱う唯一の検定試験である、ニュース時事能力検定を参考にした[6].

ニュース時事能力検定では、すべての問題の解答が4 択の選択肢から選ぶ形式で、図 5 のような穴埋め式の問題が多く採用されている。また、穴埋め式問題の解答の選択肢は、調査した限りすべて名詞であった。

そこで、本研究の提案手法では、名詞を解答とする穴埋め式の4 択問題を作成する。ただし、名詞の中で、形容動詞語幹、接尾、非自立の3 種類は、ニュース時事能力検定で解答になっていなかったため、解答に相応しくないとして対象外とした。形容動詞語幹の名詞は、「決定的」「あからさま」など、接尾の名詞は、「州」「ヶ国」「化」など、非自立の名詞は「こと」「もの」「限り」などである。以下、解答対象となる、形容動詞語幹、接尾、非自立以外の名詞を「A 名詞」とする。

問題文から解答を作成するフローチャートを図 6 に示す。

まず、抽出した記事の前文の形態素解析を行う。問題文が長文にならないように、3 文以上ある記事は2 文目までを入力文とした。

次に、記事前文タグである「\S2\」以降から、ランダムに「A 名詞」を選択する。

最後に、「A 名詞」の連続を1 つの解答とする操作を行う。選択した形態素の1 つ前の形態素が、「A 名詞」もしくは「・」であるか判定する。1 つ前の形態素が「A 名詞」もしくは「・」でない場合、もしくは入力文の最初の形態素に戻った場合になるまで1 つ前の形態素を選択する。

ただし、「・」の前の形態素が「A 名詞」でない場合、「・」の次の形態素を選択する。これは、連続した「A 名詞」の先頭の形態素を選択するためである。

同様に選択した形態素の次の形態素が、「A 名詞」もしくは「・」であるか判定する。次の形態素が「A 名詞」もしくは「・」でない場合、もしくは入力文の最後の形態素に達するまで次の形態素を連結して解答とする。ただし、選択した形態素が固有名詞である場合、次の形態素が固有名詞以外のときは解答に連結しない。

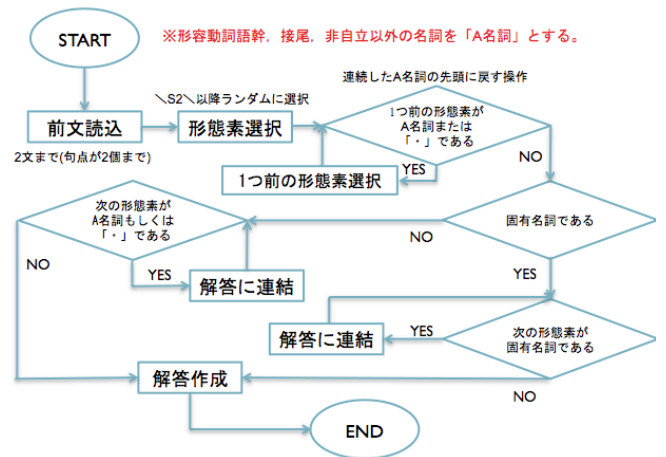


図 6: 解答作成のフローチャート

アメリカ	名詞, 固有名詞, 地域, 国, *, *, *, アメリカ, アメリカ, アメリカ
西部	名詞, 一般, *, *, *, *, 西部, セイブ, セイブ
の	助詞, 連体化, *, *, *, *, の, ノ, ノ
アリゾナ	名詞, 固有名詞, 地域, 一般, *, *, *, アリゾナ, アリゾナ, アリゾナ
州	名詞, 接尾, 地域, *, *, *, 州, シュウ, シュー
の	助詞, 連体化, *, *, *, *, の, ノ, ノ
グランド	名詞, 一般, *, *, *, *, グランド, グランド, グランド
・	記号, 一般, *, *, *, *, ・, ・, ・
キャニオン	名詞, 一般, *, *, *, *, キャニオン, キャニオン, キャニオン

図 7: 「アメリカ西部のアリゾナ州のグランド・キャニオン」の形態素解析結果

「アメリカ西部のアリゾナ州のグランド・キャニオン」という文章を形態素解析すると、図 7 のようになる。この入力文から図 6 の手順で解答を作成すると、最初を選択した「A 名詞」が「アメリカ」または「西部」の場合、解答は「アメリカ」になり、最初を選択した「A 名詞」が「アリゾナ」の場合、解答は「アリゾナ」になり、最初を選択した「A 名詞」が「グランド」または「キャニオン」の場合、解答は「グランド・キャニオン」になる。

問題文は、抽出した前文の解答部以外を出力する。そして、選択した解答の最初の形態素が人名を表す固有名詞ならば「誰」、地域を表す固有名詞ならば「何処」、一般名詞ならば「何」、それら以外の名詞はどのような名詞に対しても文脈が成立するように「【 ? 】」を解答部に置き換えて出力する。

3.4 選択肢の作成

3.4.1 word2vec

word2vec とは、単語の意味や文脈を理解するために単語をベクトル化して表現するツールである。word2vec にテキストデータを学習させることで単語の周辺の文脈から単語の意味を表現するベクトルを獲得できる。word2vec では、単語の定義によってベクトル化を行う分散表現を用いることで演算化を可能にしておき、ベクトルどうしの演算により意味の関係が表現できる。また、意味が似た単語同士は近い値を持つという性質がある。

3.4.2 日本語 Wikipedia エンティティベクトル

日本語 Wikipedia エンティティベクトル[7]とは、日本語版 Wikipedia の本文全文から学習した、単語と Wikipedia で記事となっているエンティティの分散表現ベクトルである。日本語 Wikipedia エンティティベクトルでは、word2vec を利用して学習した単語のベクトルと固有表現によって表されるエンティティのベクトルが図 8 のように 200 次元で格納されている。

word2vec では、人名や地名などの固有表現に対してのベクトル化を行う際、実世界エンティティに関連付ける処理であるエンティティリンキングが必要になる。そのため、Wikipedia の文章を word2vec に学習させただけでは、固有表現の特徴ベクトルを抽出することができない。そこで、日本語 Wikipedia エンティティベクトルを利用することで、固有表現に対してもベクトル化を行うことができる。

```
[東京大学] 0.913346 1.702978 -2.601626 -0.833778 1.134562 4.624570 0.937753 0.713276
-2.305333 -2.022263 -1.692672 1.657979 0.478895 -0.058124 -0.974068 -0.293815 0.456912
-1.448976 2.202936 -0.060279 0.523856 -0.758256 -0.910222 1.856699 0.253525 -0.757435
0.385498 -0.003326 -1.179806 0.829287 0.159821 -0.085337 -0.190831 -3.192611 -0.639016
0.744532 4.374451 -0.195011 0.837417 1.862437 -1.054375 -3.132215 0.550157 -1.161446
3.571697 -0.295730 0.820828 3.015656 -0.072501 -0.513194 3.655614 -0.742538 -0.247882
0.181987 0.807847 2.993727 -1.522233 1.054701 -1.572063 -0.632783 2.671748 1.048537
-2.892971 -0.632587 1.049019 1.140323 2.405315 3.620163 -2.255266 -0.900499 -1.122139
1.869303 0.537865 0.957488 -0.912345 1.126023 0.810905 -1.229677 -0.497341 1.049351
0.442222 0.267065 -0.138904 1.301769 -1.395499 0.842343 3.222328 0.664005 0.335558
-1.687110 0.982713 -0.575561 0.246242 -2.299703 -0.368103 0.076360 0.033748 -0.696398
-0.610437 0.122767 -0.639406 -1.376824 0.731923 -1.532060 -0.441889 -1.403357 -0.142522
-0.795035 0.699843 3.433924 2.702341 -1.527936 -1.354057 0.769708 -1.655478 0.974903
-0.068022 2.995333 1.386374 -0.926879 0.698566 -1.578082 -1.869689 -0.427349 0.475067
1.673807 -0.781144 -1.179037 -1.374926 -0.622608 1.160208 -1.599249 1.030984 3.969492
2.901484 -0.079041 -0.076098 -0.316273 -1.083471 -0.397007 0.175767 -0.663534 -1.183736
0.185450 -1.851932 -0.446278 1.678190 0.523367 1.069696 -0.789973 0.887630 2.596541
0.570117 -1.014195 0.764468 1.152348 -1.146720 0.178048 0.804208 -0.426525 0.907671
-0.157536 0.425058 -0.176795 -0.956073 -1.214763 1.658072 -0.706574 -0.048705 2.412612
-2.177432 -1.540455 -0.487441 -1.018857 1.932356 -0.854474 0.992267 0.327895 1.082534
-1.497752 0.062991 -2.433075 0.640830 0.611120 -0.564405 -2.410870 1.705194 -0.668869
1.009955 1.556106 0.633322 0.093374 -2.890807 -0.809733 -1.660049 1.487894 -0.003539
-0.733374 0.386249 -0.321303
```

図 8:「東京大学」の日本語エンティティベクトル

Enter word or sentence (EXIT to break): 東京大学

Word: 東京大学 Position in vocabulary: 21794

Word	Cosine distance
東北大学	0.902459
大阪大学	0.870136
北海道大学	0.853183
【東北大学】	0.849363
【名古屋大学】	0.849202
【長崎大学】	0.840205
【大阪市立大学】	0.839372
【京都大学】	0.838444
名古屋大学	0.836777
【大阪大学】	0.836531

図 9:「東京大学」との類似度が高い単語

3.4.3 選択肢の作成方法

誤答となる選択肢の作成方法について述べる。問題が穴埋め形式であるため、誤答の選択肢を穴埋め部分に当てはめた場合でも、文脈が成立しなければいけない。どの選択肢を当てはめても文脈が成立するように、word2vec を利用して、正答の類義語を誤答の選択肢に加えたい。

そこで、日本語 Wikipedia エンティティベクトルを利用して学習した word2vec に単語を入力する。「東京大学」という単語を入力した場合、図 9 のような出力が得られる。このとき、解答が固有名詞の場合は解答の単語のベクトル類似度上位 3 単語を選択肢に加え、解答が固有名詞以外の場合は、ベクトル類似度上位 40 単語からランダムに 3 単語を選択肢に加える。図 9 の場合、「東北大学」「大阪大学」「北海道大学」の 3 単語を誤答として選択肢を作成する。

日本語 Wikipedia エンティティベクトルでは、記事名には[]を付けることで通常の単語との区別を行い、リンクによって言及されているエンティティを正しく扱うことを可能としている。そのため、「東北大学」と「【東北大学】」など同一の単語が 2 通り出力される場合があるが、そのような単語の重複は除外し、[]を付けない状態で誤答として作成する。

3.5 良問指数

3.3 節や 3.4 節の手法で作成した問題には、誤答の選択肢を当てはめた際に文脈が成立しないことや、時事問題として相応しくない記事や、重要でない部分を解答部にしているなど、ユーザが適切でないと感じる問題も含まれている。そのため、良問指数を設定することでそのような悪問をフィルタリングする。

良問指数とは、自動作成した問題が、どれだけ時事問題として相応しいかを表す指数である。ユーザは問題に解答後、「良問」「普通」「悪問」の中から 1 つ選び、自動作成された問題の問題文、解答部、選択肢が時事問題として相応しいか評価する。「良問」「普通」「悪問」を選択した場合、それぞれ良問指数は+1, ±0, -1 となり、複数のユーザが問題を解くことで良問指数が変化

する．予備実験の結果を踏まえて，良問指数が-3 以下の問題を出題しないようにする．これによって，作問システムが出題を繰り返すに連れて，多くのユーザが「悪問」と評価した問題が淘汰され，多くのユーザが「良問」「普通」と評価した問題のみが出題されるようになる．

4. 評価実験

システムが作成した問題の評価と，固有名詞と固有名詞以外が解答の場合の誤答選択肢の作成方法を検討するため，評価実験を行った．

4.1 評価実験 1：「A 名詞」を解答とした作問システム

4.1.1 実験の概要

2011 年の毎日新聞の全文記事データ集である，「毎日新聞 2011 データ集」から，国際 7 記事，経済，スポーツ，社会，各 6 記事の計 25 記事をランダムに選択し，本提案手法を用いて時事問題を作成した．このとき，すべての問題の誤答の選択肢は，解答の単語のベクトル類似度上位 3 単語を選択した．

そして，20 名の被験者(20 代，男性，大学生)は作成した 25 問の時事問題を解き，「良問」「普通」「悪問」の判定を行った．このとき，被験者全員が同じ問題を解くように，良問指数が-3 以下の問題も出題した．

4.1.2 実験結果

各問題の正答率および「良問」「普通」「悪問」の解答割合を表 1 に示す．

被験者が「良問」と答えた問題は，全体の 26%，「良問」・「普通」と答えた問題は全体の 68%であった．良問指数が低い問題と高い問題を図 10 に示す．良問指数が低い問題は，選択肢が「今後」と「これから」など，同義語であるため正解が複数ある問題が多くあった．一方，良問指数が高い問題は，選択肢が「長野」「栃木」などの固有名詞である問題が多くあった．

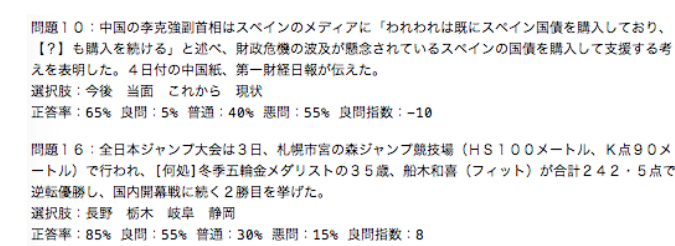


図 10：評価実験 1 での悪問と良問の例

表 1: 評価実験 1 の実験結果

問題番号	正答率	良問	普通	悪問
1	25%	5%	70%	25%
2	30%	25%	30%	45%
3	35%	0%	75%	25%
4	35%	55%	25%	20%
5	20%	55%	20%	25%
6	65%	30%	55%	15%
7	55%	5%	55%	40%
8	60%	15%	60%	25%
9	45%	25%	40%	35%
10	65%	5%	40%	55%
11	55%	10%	45%	45%
12	85%	45%	20%	35%
13	85%	20%	50%	30%
14	95%	10%	65%	25%
15	40%	50%	15%	35%
16	85%	55%	30%	15%
17	80%	15%	65%	20%
18	35%	20%	30%	50%
19	95%	60%	30%	10%
20	45%	10%	50%	40%
21	80%	30%	40%	30%
22	10%	25%	50%	25%
23	50%	40%	45%	15%
24	80%	15%	35%	50%
25	90%	25%	15%	60%
平均	58%	26%	42%	32%

4.2 評価実験 2：固有名詞を解答とした作問システム

4.2.1 実験の概要

評価実験 1 の結果を踏まえて，解答を固有名詞に限定することで，より良問指数の高いシステムが作成できると考えられる．そのため，評価実験 1 と同記事から解答を固有名詞に限定して再度作問を行った．そして，評価実験 1 での被験者 20 名に対して同様の実験を行った．

4.2.2 実験結果

各問題の正答率および「良問」「普通」「悪問」の解答割合を表 2 に示す．

被験者が「良問」と答えた問題は全体の 37%，「良問」・「普通」と答えた問題は全体の 82%であった．評価実験 1 と評価実験 2 の結果の比較を図 11 に示す．図 11 のように，解答を固有名詞に限定することで出題した問題の良問の割合が高くなった．

評価実験 2 で良問指数が負の値になった 3 問を図 12 に示す．図 12 の問題 25 では，「都内」という単語に続く選択肢が「東京」の 1 つしかないため，良問指数が負の値になったと考えられる．

表 2：評価実験 2 の実験結果

問題番号	正答率	良問	普通	悪問
1	60%	25%	50%	25%
2	50%	50%	35%	15%
3	55%	50%	35%	15%
4	85%	45%	40%	15%
5	60%	45%	45%	10%
6	80%	60%	35%	5%
7	35%	50%	35%	15%
8	50%	25%	35%	40%
9	80%	35%	60%	5%
10	90%	25%	50%	25%
11	90%	45%	45%	10%
12	50%	30%	45%	25%
13	80%	40%	50%	10%
14	80%	45%	40%	15%
15	65%	30%	55%	15%
16	75%	40%	55%	5%
17	70%	45%	40%	15%
18	85%	25%	60%	15%
19	60%	25%	55%	20%
20	65%	25%	65%	10%
21	40%	10%	40%	50%
22	90%	50%	35%	15%
23	85%	35%	50%	15%
24	95%	45%	45%	10%
25	100%	20%	30%	50%
平均	71%	37%	45%	18%

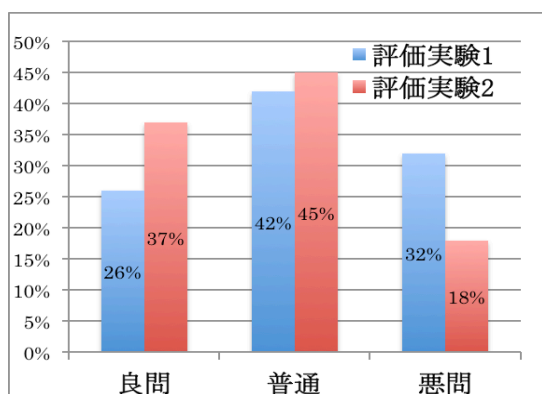


図 11：評価実験 1 と評価実験 2 の比較

問題 8：日産自動車のカルロス・ゴーン社長＝写真＝は 31 日までに【？】などのインタビューに応じ、成長の柱と期待する電気自動車（EV）について、13 年未だに提携する仏ルノーも含めて計 8 車種にラインアップを拡大する計画を示した。その上で、現在の主力 EV「リーフ」より性能を向上させた次世代 EV の開発を検討していることを明らかにした。

選択肢：毎日新聞 東京新聞 読売新聞 北海道新聞
正答率：50% 良問：25% 普通：35% 悪問：40% 良問指数：-3

問題 21：大みそかの「第 61 回 NHK 紅白歌合戦」の平均世帯視聴率は、関東地区で前半（午後 7 時半～同 9 時 25 分）35.7%、後半（同 9 時半～同 11 時 45 分）は 41.7% だった。【？】が発表した。

選択肢：ビデオリサーチ ビデオリサーチ社 関東地区 17.0%
正答率：40% 良問：10% 普通：40% 悪問：50% 良問指数：-8

問題 25：【何処】都内で正月三が日に 70～96 歳の男性 7 人が餅をのどに詰まらせて死亡したことが 3 日、東京消防庁のまとめで分かった。同庁によると、7 人は自宅で雑煮などの餅をのどに詰まらせたという。

選択肢：東京 大阪 名古屋 横浜
正答率：100% 良問：20% 普通：30% 悪問：50% 良問指数：-6

図 12：評価実験 2 での悪問

4.3 評価実験 3：固有名詞以外の名詞の選択肢の比較

4.3.1 実験の概要

評価実験 2 の結果、解答を固有名詞に限定することで、良問の出題率がより高くなった。しかし、ニュース時事能力検定の問題は、固有名詞以外の名詞も解答にある。そこで、ニュース時事能力検定の問題同様に、固有名詞以外の名詞も解答として出題したい。

評価実験 1 の結果から、固有名詞以外の名詞は、解答とのベクトル類似度上位 3 単語を選択肢に加えた場合、解答の同義語が選択肢に加わり、正答が複数存在する問題があった。ベクトル類似度上位 3 単語ではなく、上位語からランダムに 3 単語選択することで、解答の同義語が選択肢に含まれる場合を減らすことができると考えられる。予備実験の結果から、上位 40 単語までに適切な選択肢が含まれる割合が高かったため、上位 40 単語からランダムに 3 単語に選択する。

評価実験 1 で使用した 25 問のうち、固有名詞以外の名詞が解答である 18 問の解答の選択肢を、ベクトル類似度上位 40 単語からランダムに 3 単語選び、再度作成する。評価実験 1 及び評価実験 2 と同様の被験者 20 名は選択肢を再度作成した 18 問を解き、「良問」「普通」「悪問」の判定を行った。

4.3.2 実験結果

各問題の正答率および「良問」「普通」「悪問」の解答割合を表 3 に示す。

被験者が「良問」と答えた問題は全体の 30%、「良問」・「普通」と答えた問題は全体の 71%であった。評価実験 1 との比較を図 13 に示す。評価実験 1 での固有名詞以外の名詞が解答である 18 問は、良問 20%、普通 48%、悪問 32%の割合であった。評価実験 1 では良問よりも悪問の割合が高かったのに対して、選択肢をベクトル上位 40 単語からランダムに選択することで良問の割合が悪問の割合を上回った。

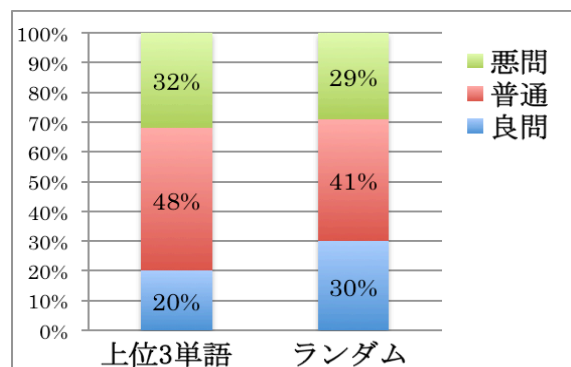


図 13：評価実験 1 と評価実験 3 の比較

表 3：評価実験 3 の実験結果

問題番号	正答率	良問	普通	悪問
1	70%	30%	45%	25%
2	45%	40%	30%	30%
3	70%	20%	55%	25%
4	55%	25%	65%	10%
5	80%	50%	40%	10%
6	85%	5%	45%	50%
7	90%	40%	40%	20%
8	55%	35%	40%	25%
9	85%	20%	25%	55%
10	85%	30%	35%	35%
11	95%	30%	35%	35%
12	90%	20%	45%	35%
13	95%	25%	35%	40%
14	20%	40%	40%	20%
15	95%	50%	40%	10%
16	100%	30%	60%	10%
17	15%	35%	30%	35%
18	100%	15%	40%	45%
平均	74%	30%	41%	29%

4.4 評価実験 4：システムの最終評価

4.4.1 実験の概要

評価実験 1, 2, 3 の結果を踏まえて、固有名詞が解答である場合、ベクトル類似度上位 3 単語を誤答の選択肢とし、固有名詞以外が解答である場合、ベクトル類似度上位 40 単語からランダムに 3 単語を誤答の選択肢とする。固有名詞と固有名詞以外の名詞を同じ割合で解答とし、システムの最終的な評価を行う。固有名詞を解答とした 25 問の評価は評価実験 2 で既に終えているので、固有名詞以外の名詞を解答とした問題を、新たな国際 7 記事、経済、スポーツ、社会、各 6 記事の計 25 記事からランダムに選択して作問した。評価実験 1, 2, 3 と同様の被験者 20 名は「良問」「普通」「悪問」の判定を行った。

4.4.2 実験結果

評価実験 4 で新たに作成した各問題の正答率および「良問」「普通」「悪問」の解答割合を表 4 に示す。

被験者が「良問」と答えた問題は全体の 35%、「良問」・「普通」と答えた問題は全体の 78%であった。これらの問題と評価実験 2 での各問題の評価を合わせると、システムの最終的な評価結果は、図 14 のように良問 36%、普通 44%、悪問 20%となった。

表 4：評価実験 4 で新たに作成した問題の実験結果

問題番号	正答率	良問	普通	悪問
1	65%	35%	40%	25%
2	100%	40%	45%	15%
3	40%	30%	40%	30%
4	55%	40%	40%	20%
5	40%	25%	45%	30%
6	15%	50%	35%	15%
7	70%	30%	55%	15%
8	75%	45%	40%	15%
9	55%	30%	45%	25%
10	85%	15%	50%	35%
11	40%	65%	25%	10%
12	55%	55%	25%	20%
13	90%	60%	35%	5%
14	70%	20%	55%	25%
15	45%	25%	50%	25%
16	30%	45%	40%	15%
17	65%	35%	35%	30%
18	35%	25%	50%	25%
19	85%	20%	65%	15%
20	60%	25%	35%	40%
21	85%	30%	55%	15%
22	35%	45%	35%	20%
23	55%	55%	20%	25%
24	60%	30%	50%	20%
25	70%	10%	65%	25%
平均	59%	35%	43%	22%

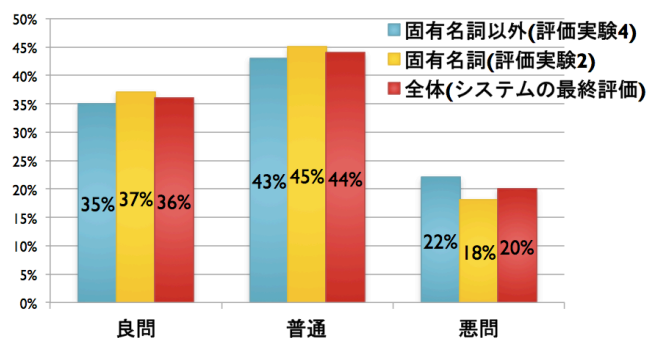


図 14：システムの最終的な評価結果

問題 6：米外交公電の暴露を続ける内部告発サイト「ウィキリークス」は 1 日、日本の調査捕鯨を妨害する反捕鯨団体「シー・シェパード」（ＳＳ、本部・米ワシントン州）をめぐり、日本の【？】幹部が米側に ＳＳへの対策強化を要求したとする在日米大使館発の公電を公表した。
選択肢：農林水産省 経済産業省 厚生労働省 林野庁
正答率：80% 良問：60% 普通：35% 悪問：5% 良問指数：11

問題 11：第 31 回「毎日経済人賞」の受賞が決まったコマツの坂根正弘会長は、中国など新興国市場への進出をいち早く強化し、日本を代表するグローバル企業に成長させた。東しの柳原定征会長は、航空機などに使われる「[何]」の事業拡大に粘り強く取り組み、世界をリードする事業に育てるなど、基礎素材メーカーの多角化に尽力した。
選択肢：炭素繊維 ポリ塩化ビニル(20) チタン(6) ポリテトラフルオロエチレン(33)
正答率：40% 良問：65% 普通：25% 悪問：10% 良問指数：11

問題 13：三菱電機は I H ジャー炊飯器「蒸気レス I H」の新製品を 2 月 1 日に発売する。誤操作を防ぐため、メニュー設定や予約時間を【何】で知らせる機能を備えた。
選択肢：音声 映像(5) 肉声(36) 字幕(6)
正答率：90% 良問：60% 普通：35% 悪問：5% 良問指数：11

図 15：良問指数が最も高い問題

問題8：日産自動車のカルロス・ゴーン社長=写真=は31日までに【?】などのインタビューに応じ、成長の柱と期待する電気自動車（EV）について、13年末までに提携する仏ルノー分も含めて計8車種にラインアップを拡大する計画を示した。その上で、現在の主力EV「リーフ」より性能を向上させた次世代EVの開発を検討していることを明らかにした。
選択肢：毎日新聞 東京新聞 読売新聞 北海道新聞
正答率：50% 良問：25% 普通：35% 悪問：40% 良問指数：-3

図 16：良問指数が負の値となった問題の例

5. 考察

5.1 良問指数の高い問題

良問指数の高い問題は、どのような問題だろうか。評価実験の中で最も高い良問指数は11で、図15に示した、評価実験2の問題6、評価実験4の問題11、問題13の3問である。良問指数が高い問題は、同義語や文脈に当てはまらない単語が選択肢に無い問題が多くあった。しかしながら、評価実験4の問題13の選択肢である「肉声」は「音声」の部分集合であるため、「肉声」が正答である場合、「音声」も正解になる不備があるが、最も良問指数の高い問題となった。また、図16の問題の選択肢「毎日新聞」「東京新聞」「読売新聞」「北海道新聞」のように、誤答の選択肢に不備が無い問題で良問指数が負の値となった場合もあった。

どの記事を選択するか、記事のどの部分を解答にするのか、どのような選択肢を作成するのか、それぞれを総合的に見て、ユーザは良問かどうか判断していると考えられる。ユーザによって良問の基準は異なるが、一般的にユーザが良問と感じる問題の傾向を理解することができれば、より良問指数の高い問題が出題できるだろう。しかし、問題を解くユーザが良問と感じる問題は、必ずしも時事問題作成者にとって良問とは限らないので、その点についても検討する必要がある。

5.2 提案手法の課題

本提案手法における改善点は、解答と記事の選択方法である。

本提案手法では、文章の形態素解析を行い、「A 名詞」を解答としてランダムに選択している。解答は形態素のみに注目して選択しているが、形態素解析に加えて、文脈や構文を考慮して、良問となりやすい文脈や構文から解答を選択する手法も考えられる。本提案手法で形態素解析に用いているソフトウェア CaboCha は、構文解析も行うことができる。ユーザが良問と感じる問題の構文の傾向を理解することができれば、構文解析を利用することでより良問を出題するシステムが作成できると考えられる。

また、本提案手法では、記事前文に句点が無い場合のみを作問対象から除外しているが、ユーザが悪問と感じる問題の傾向を理解することができれば、記事前文からフィルタリングする新たな基準を作成できると考えられる。

6. むすび

本論文では、新聞記事データから4択式の時事問題を自動作問する手法について述べた。評価実験では、作問システムの問題の評価を行い、80%の問題が、ユーザが「良問」・「普通」と感じる問題であった。80%の割合で、作問システムとして相応しい問題が出題できた。一方で、ユーザが「悪問」と感じる問題が20%あり、記事や解答部の新たな選択基準を設けて、これら「悪問」を減少させることが今後の課題である。

本論文では、時事問題を対象として作問システムを作成したが、日本語の文章であれば、時事問題に限らず様々な分野の問題を作成することができる。そのため、本提案手法を用いて、新たな作問システムが作成されることが期待される。

参 考 文 献

- [1] 野崎 輝, 寛 宗徳, 二石 芳裕, 渡辺 一衛: 珠算における自動作問システムの構築, 成蹊大学理工学研究報告 Vol.48, pp.69-74, 2011
- [2] 山崎 隆介: 連鎖型パズルゲームにおけるパズル問題の創作, ゲームプログラミングワークショップ 2013 論文集, pp.181-121, 2013
- [3] 工藤 拓: CaboCha/南瓜: Yet Another Japanese Dependency Structure Analyzer
<http://taku910.github.io/cabocha/> (2017年1月8日閲覧)
- [4] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space, arXiv preprint arXiv: 1301.3781, 2013.
- [5] 毎日新聞社: CD-毎日新聞データ集
<http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html> (2017年1月16日閲覧)
- [6] 日本ニュース時事能力検定協会 監修: 2015 ニュース検定 公式テキスト&問題集 基礎編 3・4級, 毎日新聞社, 2015
- [7] 鈴木 正敏, 松田 耕史, 関根 聡, 岡崎 直観, 乾 健太郎: Wikipedia 記事に対する拡張固有表現ラベルの多重付与, 言語処理学会 第22回年次大会 発表論文集, pp.797-800, 2016