

問題自動生成システムを利用した作問演習の実践と評価

福坂 祥基^{*1}, 高木 正則^{*1}, 山田 敬三^{*1}, 佐々木 淳^{*1}

^{*1} 岩手県立大学大学院ソフトウェア情報学研究科

Practice and Evaluation of a Problem Producing Exercise by Using an Automatic Question Generation System

Fukusaka Shoki^{*1}, Masanori Takagi^{*1}, Keizo Yamada^{*1}, Jun Sasaki^{*1}

Graduate School of Software and Information Science, Iwate Prefectural University^{*1}

近年、情報処理技術者試験や TOEIC など様々な検定試験が行われ、出題問題の作成に大きな負担がかかっている。一方で、教育現場では必要となる多様な演習問題が不足している。これに対し、演習問題を自動生成する研究が多数存在する。しかし、問題生成の際に事前準備作業が求められ、作問者にかかる負担は未だ大きい。我々はこれまで作問負担の軽減を目的とし、過去問題をリソースとした問題自動生成システムを提案してきた。本稿では、地域学習を目的とした授業において、本システムを活用した作問演習を実施し、作問負担に関するアンケート結果と、作成された問題の分析結果から本システムの有効性を評価した。

キーワード: コンテンツ自動生成, 作問支援, 問題自動生成, 過去問題, 知識ベース

1. はじめに

近年、TOEIC や情報処理技術者試験など様々な検定試験が行われている。これらの検定試験では、出題する問題の作成に多くの負担がかかっている。一方で、教育現場で必要となる多様な演習問題が不足している。これに対し、演習問題を自動生成する研究が多数存在する⁽¹⁾⁽²⁾。しかし、問題を生成するためのリソース情報に XML のタグを付与することや、出題する知識のカテゴリに対応した語尾を定義するなど、問題生成の事前準備に何らかの作業が必要であり、作問者への負担は未だ大きい。著者らの先行研究では、過去問題をリソースとした問題自動生成用の知識ベースを定義し、この知識ベースを自動構築する手法⁽³⁾を提案した。しかし、過去問題をリソースとした知識ベースの活用可能性については検証されていなかった。そこで、我々は先行研究で古舘らが定義した知識ベースを手動で構築し、これを情報源とした問題自動生成システムを開発した。そして、地域の知識を問う検定試験の作問現

場で本システムを利用してもらい、本システムと過去問題をリソースとした知識ベースの活用可能性を検証した⁽⁴⁾。

本研究では、教育現場である岩手県立大学の地域学習を目的とする授業で問題自動生成システムを利用した作問演習を行い、学生が本システムを利用して作成した問題を定量的な指標で評価した結果から、システムの有効性を検証する。以下、2 章では、知識ベースの構成と問題自動生成システムの概要を述べ、3 章では手動で構築した知識ベースについて述べる。4 章では開発したシステムの利用の流れを述べ、5 章、6 章で導入した演習の概要と評価についてそれぞれ述べる。7 章で本研究のまとめと今後の課題について述べる。

2. 問題自動生成システム

2.1 生成する問題の形式

本研究では、専門用語や固有名詞を問う一問一答形式の問題を生成対象とし、過去に出題された問題の知識を活用して元の問題とは異なる出題形式や、異なる

知識の組み合わせを持つ新しい問題を生成する。なお、本研究では共同研究を行っている盛岡商工会議所が主催するご当地検定の盛岡もの識り検定（以下、もりけん）で出題される盛岡市の地理・歴史・人物などの固有名詞に関する知識を問う問題を活用する。もりけんの過去問題（問題文と正答）の例を表 1 に示す。

表 1 過去問題の例

問題文	正答
新渡戸稲造が設立時に事務次長をつとめた国際組織の名称は何ですか。	国際連盟
新渡戸稲造の後に、国際連盟事務次長になった、盛岡出身の外交官は誰ですか。	杉村陽太郎
今年、盛岡駅前滝の広場で胸像が除幕された先人は誰ですか。	新渡戸稲造
平成 24 年、生誕 150 年を迎える盛岡ゆかりの人物は誰ですか。	新渡戸稲造

2.2 システム概要

本システムでは、作問者が入力したキーワード（専門用語や固有名詞）に関連する問題を自動生成した後、作問者が生成された問題を確認・修正、または生成された問題を参考にして新規問題を作成できる作問支援環境を提供する。本システムの概要図を図 1 に示す。作問者はまずシステムへ作問したい問題のキーワードを入力する（図 1①）。システムは入力されたキーワードに関連する知識を知識ベースから取得する。そして、取得した知識量に応じて利用できる出題テンプレートを取捨選択し、各出題テンプレートに取得した知識を挿入することで問題を自動生成する。生成した問題の一覧は作問者へ提示し、作問者は一覧の中から作問に利用したい問題を選択し（図 1②）、任意に修正を加えて（図 1③）、システムへ登録する。

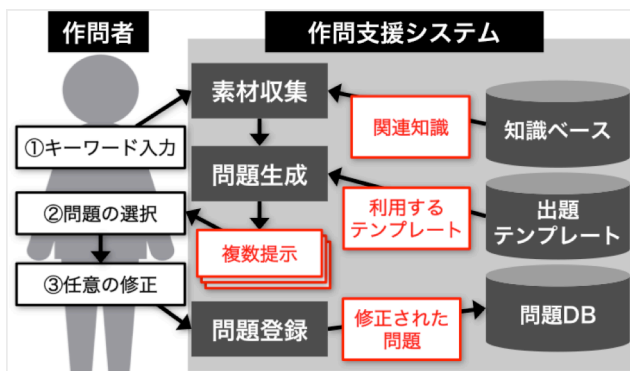


図 1 システム構成図

2.3 知識ベースの構造

先行研究で古舘らが定義した知識ベースは、メッシュ型の構造であったが、本研究で提案する知識ベースは問題生成の容易さを考慮し、木構造で定義した。本知識ベースの構成要素は、「対象知識」、「カテゴリ」、「プロパティ」、「オブジェクト」の 4 つであり、過去問題からこれらの構成要素を抽出する。また、この 4 つの構成要素が木構造で管理される。表 1 の過去問題から構築される知識ベースを図 2 に示す。本知識ベースに登録されている知識は、4 つの構成要素のどれに該当するのかを識別する記号が割り振られており、出題テンプレートに知識を挿入する際に利用される。記号の KL I と KL II は知識ベースのルートからの知識階層を意味し、これらに付随する [O] はオブジェクトを、[O][P] はそのオブジェクトに紐付くプロパティを意味する。オブジェクトに続く番号は、その知識階層の全オブジェクトの中で何番目に登録されているかを表す。また、プロパティに続く番号も同様に、オブジェクトに紐付けられる全プロパティの中で、何番目に登録されているかを示す。変換方法は「対象知識」と「オブジェクト」が一致するかどうかなどによって 3 種類ある。ここで「対象知識」とは、問題で問われている、または解決の中心となる知識を指す。「カテゴリ」

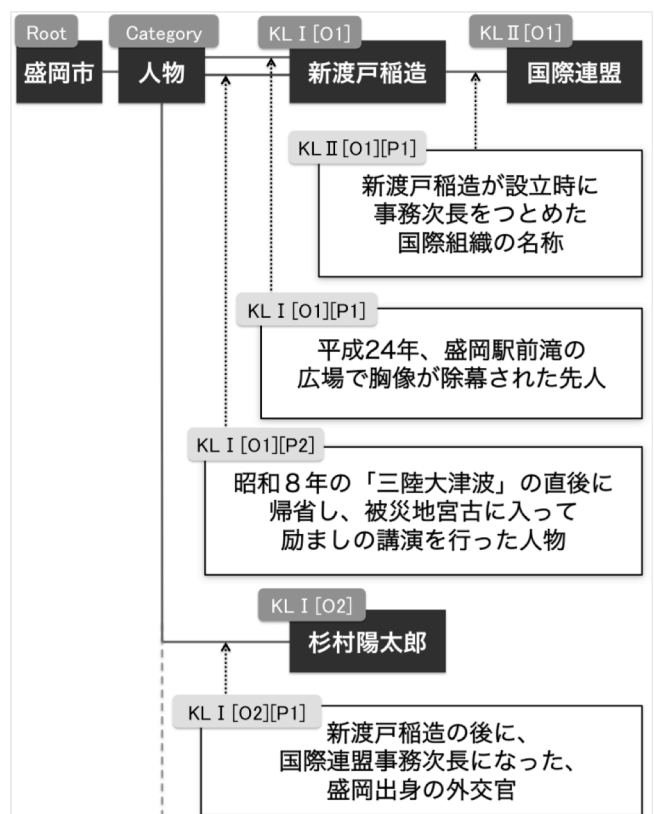


図 2 過去問題（表 1）から構築される知識ベース

はこの「対象知識」の上位概念として位置付ける。例えば、「新渡戸稲造」が対象知識の場合、カテゴリは「人物」に分類される。「プロパティ」は、「対象知識」を説明する情報であり、問題文から「～は誰ですか」のような不要語を取り除いた情報である。「オブジェクト」は、「プロパティ」が指し示す実際の値となり、問題の答えに該当する。

また、もりけんには問題文の条件に適合しない単語や説明を選択させる問題がある(2.5章で後述)。この出題形式の問題から知識を抽出する際は、問題の誤答選択肢をそれぞれオブジェクトとして抽出する。例えば、図3の場合、3つの誤答選択肢がオブジェクトとして抽出される。

2.4 知識ベースの構築手順

(1) 対象となる検定試験で扱う知識の最上位の概念を知識ベースの最上位ノード(ルート)に置く。ここでは、もりけんの過去問題を利用しているため、盛岡市とする。

(2) 各問題の対象知識を抽出し、カテゴリを決定する。

(3) 各問題の対象知識とオブジェクトを比較する

(a) 対象知識とオブジェクトが同値である場合(図4(1))、問題のカテゴリ(対象知識の上位概念)が既に知識ベースに登録されているか検索する。既にカテゴリが登録されている場合、そのカテゴリの子ノードとして対象知識を配置し、カテゴリと対象知識の間にこの問題から抽出したプロパティ

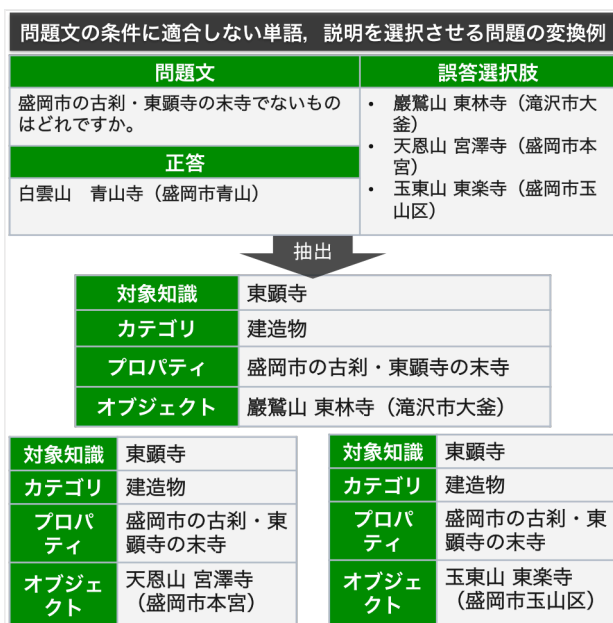


図3 誤答選択肢からオブジェクト抽出の抽出例

を紐付ける。カテゴリが登録されていない場合は、新規のカテゴリとして、ルートの子ノードに配置し、対象知識とプロパティを同様に配置する。また、カテゴリと対象知識の両方が既に登録されている場合はプロパティの紐付けのみを行う。

(b) 対象知識とオブジェクトが異なる場合(図4(2)), (2)で問題から抽出した対象知識が既に知識ベースに登録されているか検索する。既に登録されている場合、知識ベースの対象知識の子ノードにオブジェクトを配置し、そのノード間にプロパティを紐付ける。対象知識が登録されていない場合、問題の対象知識をカテゴリの子ノードとして新規に登録し、対象知識の子ノードにオブジェクトを配置し、そのノード間にプロパティを紐付ける。また、(2)で決定した問題のカテゴリが登録されていない場合、問題のカテゴリを知識ベースの新規のカテゴリとしてルートの子ノードに配置し、このカテゴリの子ノードとして対象知識を配置する。このとき、プロパティの紐付けは行わず対象知識の子ノードとしてオブジェクトを配置し、このノード間においてプロパティを紐付ける。

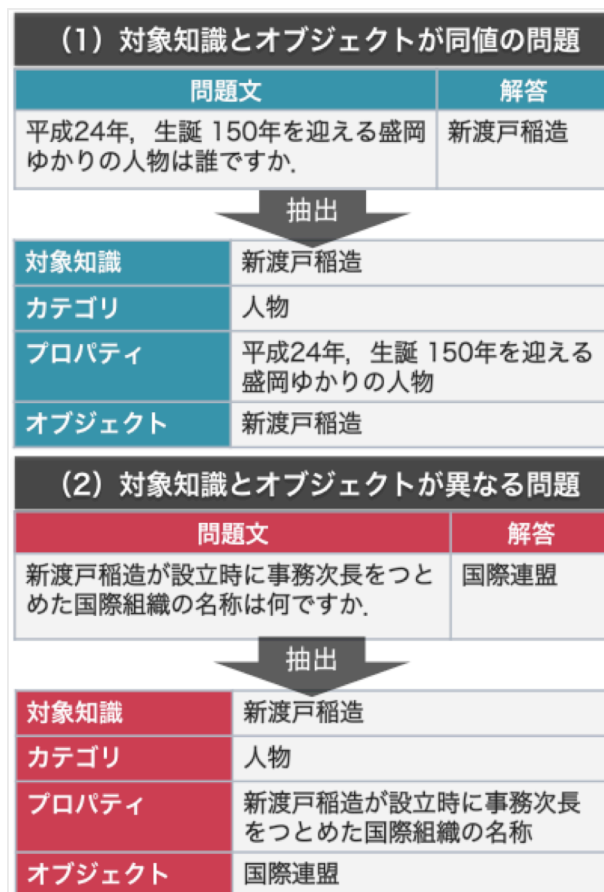


図4 過去問題から構成要素の抽出例

表 2 出題形式一覧

No.	出題形式	割合
①	問題文の条件に適合する単語を選ぶ、または記入する形式	87.6%
②	問題文に示された知識に適した説明を選ぶ形式	2.4%
③	問題文の条件に適合しない単語、または説明を選択する形式	3.6%
④	提示した画像に適した説明又は単語を選ぶ形式	2.4%
⑤	問題文の条件に適合する単語が複数存在し、その中の1つを正答とする形式	2.0%
⑥	列挙された項目を問題文の条件通りに並び替える、または順番選択する形式	1.2%
⑦	問題文に含まれる伏せ時に当たる文字を選択、または記入する形式	0.8%

2.5 出題テンプレートの概要

出題テンプレートを作成するために、もりけんの過去問題 250 問の出題形式や知識の出現箇所を分析した。その結果、画像を利用する問題や、穴埋め問題など 7 種類の形式の問題が存在することが分かった。表 2 に 7 種類の形式を示す。

表 2 の出題形式のうち、No.③の出題形式は平成 28 年度から非推奨の形式となり、平成 27 年度を最後に出題されていないため出題テンプレートとして採用しないこととした。また、出題形式④～⑦は出題される割合が少なく、問題に含まれる知識を知識ベースの各構成要素に変換することが困難であったため、知識抽出の対象外とした。そのため、これらの形式もテンプレートとして採用しないこととした。以上より、本研究では出題形式①と②をモデルとして出題テンプレートを作成した。出題テンプレートの一覧を表 3 に示す。テンプレート①②③は出題形式①をモデルにしており、テンプレート④⑤は出題形式②をモデルにしている。

本テンプレートでは、知識ベースの知識を挿入する箇所や内容が記号で表現されている。知識階層やオブジェクト、プロパティの記号は知識ベース (図 2) と対応しており、 i, j, k はオブジェクトやプロパティの識別番号を示す。また、 n, p, o はオブジェクトやブ

表 3 出題テンプレート

No.	問題文	正答
①	KL I [O][Pi]($1 \leq i \leq n$)であり、 KL I [O][Pj]($1 \leq j \leq n, i \neq j$)のは何か	KL I [O]
②	KL I [O][Pi]($1 \leq i \leq n$)である KL I [O] + KL II [Oj][Pk]($1 \leq j \leq o$)($1 \leq k \leq p$)のは何か	KL II [Oj] ($1 \leq j \leq o$)
③	KL I [Oi][Pj]($1 \leq i \leq n$)($1 \leq j \leq o$)である KL I [Oi]($1 \leq i \leq n$)+KL I [O][Pk]($1 \leq k \leq p$)のは何か	KL I [O]
④	KL I [O]の	KL I [O][Pi]($1 \leq i \leq n$)
⑤	説明として正しいものはどれか	KL II [Oi][Pj] ($1 \leq i \leq n$)($1 \leq j \leq n$) は KL II [Oi]($1 \leq i \leq n$) である

ロパティの総数を示している。作問者が入力したキーワードが KL I [O]のいずれかに一致するとき問題が生成される。生成時は、テンプレート中の記号に該当する知識ベースの情報を 1 つずつ挿入する。

3. 知識ベースの構築

本研究では、もりけんで過去に出題された 2000 問を分析し、各問題を 4 つの構成要素に変換した上で、知識ベースを手動で構築した。なお、もりけんで出題された問題の出題形式は、一問一答形式と多肢選択形式があるが、いずれも問題文と正答から知識を抽出した。対象知識は盛岡の専門用語となり得る単語を手動で抽出した。また、本稿のシステムでは問題の自動生成に構成要素のカテゴリの内容は関与しない。そのため、カテゴリの決定は筆者の主観で行った。なお、画像を利用する問題や選択肢に列挙した項目の順番を問う問題は、抽出の対象外とした。

4. 問題自動生成システムの開発

3 章で構築した知識ベースと表 3 の出題テンプレートを活用した問題自動生成システムを開発した。表 4 に開発環境を示す。本研究で開発した問題自動生成シ

表 4 開発環境

Web サーバ	PHP 5.3(CakePHP 2.5.2)
開発言語	Apache 2.2.3
DBMS	MySQL 5.0.95

システムの主な利用手順を以下に示す。なお、認証のためのログイン機能については説明を割愛し、ログイン直後の手順を示す。

(1) 作問者は「新渡戸稲造」や「盛岡城」を始めとする任意のキーワードを入力する。図 5 にキーワード入力画面を示す。入力フォームにキーワードを入力後、システムでは知識ベースの KL[O]を検索し、入力されたキーワードが含まれる対象知識を一覧表示する。図 6 に対象知識の一覧画面を示す。例えば、「盛岡」と入力すると、「盛岡城」などが一覧表示される。また、入力されたキーワードが含まれる対象知識が知識ベースに登録されていない場合は何も表示せずに、再度入力を促す。



図 6 キーワード入力画面



図 7 入力された文字列が含まれる対象知識一覧

(2) 手順①の一覧の中から、作問者は任意にキーワードを選択する。キーワードが選択されたらシステムでは知識ベースからキーワードに関連する知識を収集し、出題テンプレートに挿入することで問題を自動生成する。図 7 に生成された問題の一覧



図 5 生成された問題の一覧画面

画面を示す。

(3) 作問者は手順②で表示された問題の一覧から作問に利用する問題を選択し、任意に修正する。図 8 に選択した問題情報の修正画面を示す。



図 8 自動生成された問題の修正

(4) 手順③で問題修正後、システムに問題を登録する。登録された問題は登録問題一覧画面で確認できる。図 9 に登録問題一覧画面を示す。再度作問を行う場合は画面の「作問する」ボタンを押下することで、手順①に移行する。



図 9 登録された問題の一覧

5. 本システムを活用した作問演習

5.1 演習概要

本稿では、本学の科目「地域と情報」(履修者 63 名)で本システムを利用した作問演習を行い、学生により作成された問題を評価することで本システムの有効性を測る。評価の手法としては、問題自動生成機能から作成された問題と、手動で作成された問題を作成してもらう。これらを実際に解いてもらい、得られた結果に信頼性・難易度・識別度・注意係数の指標を用いて、比較することで検証を行う。なお、作問は、「もりけん」の過去問題を参考に行ってもらった。

本演習の流れを表 5 に示す。本演習は、2 度の授業に分けて行った。第 1 回目 (7 月 21 日) は、もりけん

の過去問題を使用した事前テストを行い、その結果に基づいたグループ分けを行った。その後、それぞれのグループは作問作業を行ってもらった。演習の終了後、筆者がそれぞれのグループで作成された問題で構成されるテストを作成し、第2回目（7月28日）ではグループ間でお互いのテストを解いてもらった。以下に各工程の詳細を述べる。

表 5 演習の流れ

日時	実施内容
7月21日	事前テスト
	受講者のグループ分け
	グループごとの作問作業
7月28日	本テスト

5.2 事前テストとグループ分け

事前テストは、過去問題（平成28年度3級）の中から30問出題した。4択形式の問題となっており、システム上で図10のように出題した。履修者が解答を終了するとシステムは採点を始める。全履修者の採点が終了すると、グループ分けを行う。

グループは、問題自動生成機能を利用して作問を行うグループAと、問題自動生成機能は利用せずに問題登録機能のみを利用した作問を行うグループBに分かれる。システム側で履修者を得点順に並び替え、得点の高い学生から交互にA、Bグループへ割り振った。これにより、グループ間での知識の偏りを抑制し、事前知識の差異が実験結果に影響しないよう考慮した。その結果、グループAには32名、グループBには31名が振分けられ、グループごとの平均点はそれぞれ14.6点、15.5点、標準偏差は4.285と3.963となった。

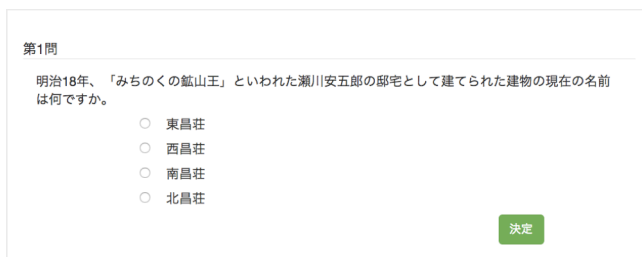


図 10 事前テストの出題画面

5.3 グループごとの作問作業

所属するグループによって、システムで表示される

画面や作問作業の流れは異なる。グループAに所属する場合は、図5の画面が表示され自動生成機能を利用して作問する。グループBは、図8の入力画面が何も入力されていない状態で表示され、学生がゼロから作問する。また、グループ分けの完了後、受講者には3つのキーワードがランダムに割振られてシステム画面に表示される。本演習では、このキーワードをテーマに作問してもらう（グループAは図5にキーワードを入力して自動生成）。受講者には1つのキーワードに対して1問（計3問）作成するように指示した。また、このキーワードは我々が事前に30個用意したもので、割振りの際は1人に同一のキーワードが複数与えられないよう考慮して行った。実際に用意したキーワードの例を5つ表6に示す。なお、両グループともインターネットの活用を許可した。

表 6 作問テーマの例

No.	テーマ
①	姫神山
②	烏帽子岩
③	シダレカツラ
④	盛岡八幡宮
⑤	高松の池

5.4 本テスト作成と実施

先の作問作業により、Aチーム（32名）からは99問、Bチーム（31名）からは97問の計196問の問題が作成された。これらから、本テストとして出題する問題を以下の基準に該当しないものから選抜した。

- 疑問文として成り立たない
- 誤答選択肢が不適切で正答を判別しやすい
- 誤字が含まれている
- 問われている知識が重複する

本テストにおいては、作問者自身が解答することを防ぐため、テストAはグループBに、テストBはグループAに出題した。解答結果の平均点として、テストAは21.7点、テストBは15.5点となった。

6. 評価

6.1 評価指標

作成された問題の評価には、「信頼性」、「難易度」、「識別度」、「注意係数」の4つの指標を用いる。以下に各指標の求め方と、両テストの比較結果を述べる。

6.2 信頼性⁽⁵⁾

信頼性は実施したテストが信頼できるかを示す指標である。これを求める手法として内部一貫法を採用し、クロンバックの α 係数を算出することで評価した。 α 係数の求め方は以下の式となる。値は0から1までとり、0.8を超えると高いとされている。 α 係数は、次式で求める：

$$\alpha = \frac{\text{項目数}}{\text{項目数} - 1} \cdot \left(1 - \frac{\text{各項目の分散の合計}}{\text{合計点の分散}} \right)$$

算出の結果、テストAは0.6330、Bは0.6775と、どちらも低い結果となり、テストAの信頼性はテストBの信頼性を下回る結果となった。

6.3 難易度

難易度は、古典的テスト理論を用いて、全回答者数に対する正答した解答者数の割合を求めることで算出した。0から1までの値をとり、高いほど易しく、低いほど難しいとされる。評価基準⁽⁶⁾と本テストにおける問題の該当数を表7に示す。テストAはBと比較し、問題はない項目数が多く、不適切とされる問題数が少なくなった。

表 9 項目難易度の基準とテストの該当項目数

基準	評価	テスト A	テスト B
0.4 未満か 0.8 以上	不適切	6	16
0.3 未満か 0.9 以上	極端に高い、 または低い	8	7
上記以外	問題はない	24	15

6.4 識別度

識別度は、項目がいかにかにテスト受検者の能力を分別できるかを示す指標である。主に0から1までの値をとり、値が高いほどその能力が高いとされている。評価基準⁽⁶⁾と本テストにおける該当数を表8に示す。テ

ストAはBと比較し、問題がない項目が多く含まれ、識別度が低い基準の項目数も少なかった。

表 7 項目識別度の基準とテストの該当項目数

基準	評価	テスト A	テスト B
0.2 未満	極端に低い	13	16
0.3 未満	低く不十分	6	7
上記以外	問題はない	19	15

6.5 注意係数

注意係数は、テストの中で特異な問題を示す指標であり、S-P表⁽⁷⁾（表の説明は割愛する）を用いて算出する。注意係数が高い問題は、高い得点を持つ受講者が誤答し、低い得点の受講者が正答している特異な問題であり注意が必要とされる。項目注意係数の評価基準と本テストにおける該当数を表9に示す。テストAはBと比較して、特に注意すべき項目数が1問だけ多いが、注意すべき項目は少なく、問題はない項目は多い結果となった。

表 8 項目注意係数の基準とテストの該当項目数

基準	評価	テスト A	テスト B
0.75 以上	特に注意が 必要な項目	12	11
0.5 以上 0.75 未満	注意すべき 項目	13	18
上記以外	問題はない	13	9

6.6 各指標の評価

信頼性係数は両テストとも信頼性が低く、更に、テストAの係数はBを下回る。一方で、両テスト各項目の「難易度」、「識別度」、「注意係数」を算出した結果、適切な数値を持つ項目数は全てテストBを上回り、不適切または異常が有る項目数も、概ね下回っていた。

このことから、本研究の問題自動生成機能を活用することで、テスト全体の信頼性の向上は期待できないが、手動による作問に比べて、正常な「難易度」、「識別度」、「注意係数」を持つ問題の作成を支援できることが示唆される。

6.7 アンケート結果と評価

本演習の終了後に実施したアンケートの結果を図 11 に示す。問題作成の難易度について、5 段階評価（1 が易しく 5 が難しい）で質問した結果、グループ A は B に比べ、問題作成が易しいと感じた受講者の割合が高く、難しいと感じた割合は低い結果となった ($p=0.06$)。今回のアンケートの結果と、6 章 6 節の各指標の評価を踏まえると、少ない負担で、適切な指標の値を持つ問題の作成も可能であることが示唆される。

また、自由記述のアンケートも実施した。主な回答内容として、問題の自動生成機能に対する意見や、本演習に対する感想が多数を占めていた。問題自動生成機能に関しては、「問題作成の手間が省けて良かった」との声がある一方、「似た問題が生成されている」、「生成される問題が難しい」、「生成数が少ないキーワードがある」など、今後の課題となりえる意見もあった。作問演習に対する感想としては、「盛岡市への理解を深めることができた」、「生まれ育った地域でも知らないことが多くあると実感した」、「盛岡について勉強してみたいと思った」などの声が寄せられた。これらから作問演習による作問の対象への興味や関心、理解を促進する効果も示唆された。

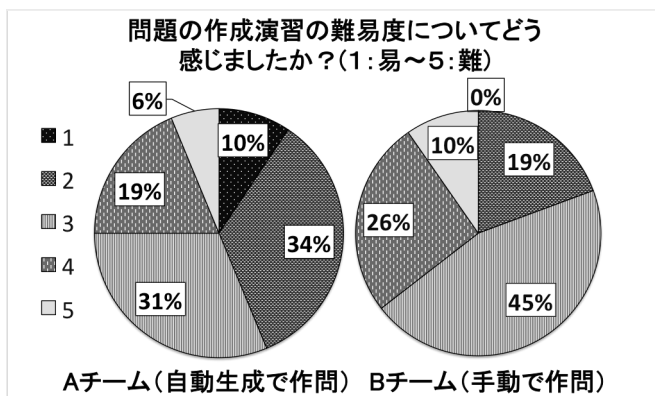


図 11 演習後のアンケート結果

7. まとめと今後の予定

我々はこれまで、もりけんの過去問題をリソースとした知識ベースを構築し、問題自動生成システムの提案と開発してきた。本稿では、大学における授業に作問演習という形でシステムを導入し、作成された問題を実際に解いてもらった。この回答結果と事後のアン

ケートから、少ない負担で、適切な「難易度」、「識別度」、「注意係数」を持つ問題の作成が可能であることや、作問演習による興味や理解の促進が示唆された。

また、今回の演習では、問題や異常がある指標を持つ項目も得ることができた。現段階の構想ではあるが、これらの要因や特徴を分析することで、編集時に注意喚起を促す機能や、問題の自動生成の段階で修正されたものを出力する時点で対策したものを出力する機能を開発することで対策していきたい。今後は、知識ベースを半自動で構築できる方法を検討し、より少ない負担で作問できる方法を検討する。

謝辞

本研究に協力して頂いた、授業「地域と情報」の履修者の皆様と担当の教員に感謝を申し上げます。

参考文献

- (1) 菅沼明, 峯恒憲, 正代隆義: “学生の理解度と問題の難易度を動的に評価する練習問題自動生成システム”, 情報処理学会論文誌, Vol.46, No.7, p. 1810-1818 (2005)
- (2) 田村吉宏, 山内崇資, 林佑樹, 中野有紀子: “Wikipedia を用いた質問応答と多肢選択問題による歴史学習”, 人工知能学会全国大会論文集, 29th, ROMBUNNO.1N2-2, 2015
- (3) 古舘昌伸, 福坂祥基, 高木正則: “試験問題をリソースとした知識ベース自動構築手法の検討”, 教育システム情報学会第 39 回全国大会, p.75-76 (2014)
- (4) 福坂祥基, 高木正則, 山田敬三, 佐々木淳: “過去問題をリソースとする知識ベースを活用した問題自動生成システムの開発と評価”, 情報処理学会情報教育シンポジウム, p. 39-46 (2016)
- (5) 妥当性と信頼性:[Online].
<http://www.u-gakugei.ac.jp/~kishilab/validity-reliability.htm>
- (6) 大友賢二: 『言語テスト・データの新しい分析法 項目応答理論入門』, 大修館書店, 1996.
- (7) 佐藤隆博: S-P 表の入門, 教育実践文庫 3, 明治図書, 1985.