

過去問題をリソースとした知識ベース自動構築手法の提案

A Proposal of Method for Automatic Construction of Knowledge Base by Using Resource of Past Exam Question

福坂祥基⁺¹ 高木正則⁺¹ 山田敬三⁺¹ 佐々木淳⁺¹
SHOKI FUKUSAKA⁺¹ MASANORI TAKAGI⁺¹ KEIZO YAMADA⁺¹ JUN SASAKI⁺¹

岩手県立大学ソフトウェア情報学研究科

⁺¹ Graduate School of Software and Information Science, Iwate Prefectural University
Email: g231o026@s.iwate-pu.ac.jp

あらまし：教育現場では多様な演習問題が不足しており，演習問題を自動生成する研究が多数存在する．しかし，問題作成者は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．これに対し先行研究では，過去問題をリソースとし，問題自動生成のための知識ベース自動構築手法を提案したが，対象となる出題形式が限定されていた．本稿では，先行研究の自動構築手法における過去問題の出題形式の拡大を目的に，画像を利用する出題形式や穴埋めの出題形式を含む7種類の出題形式の知識抽出を可能とする手法と自動構築機能を提案する．

キーワード：問題自動生成，知識ベース，過去問題

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とに対し，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，作問者にかかる負担は未だ大きい．これに対し古舘ら⁽³⁾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を提案した．古舘らの手法により，問題自動生成のための事前準備に生じる負担を軽減した問題の自動生成が期待される．しかし，知識の抽出可能な過去問題の形式に制限があり，画像を用いる形式の問題や穴埋め形式の問題は抽出の対象外となっていた．本稿では，これまで対象外とされてきた出題形式から知識の抽出が可能な知識ベースの構造と自動構築の手法を提案する．

2. 知識ベースの概要

2.1 知識ベースの構成要素

本研究で提案する知識ベースは，古舘らによる知識ベースのものを参考に定義し，「対象知識」「カテゴリ」「プロパティ」「オブジェクト」の4つの構成要素から構成される．「対象知識」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．具体的には，対象領域の盛岡に存在する施設名や地域名など盛岡独自の専門用語，または人名や氏族の名前などになることが多い．「カテゴリ」はこの「対象知識」の上位概念として位置付ける．例えば，「盛岡劇場」が対象知識の場合，カテゴリは「劇場」に分類される．「プロパティ」は，「対象知識」を説明する

情報である．「オブジェクト」は，「プロパティ」が指し示す実際の値となる．

2.2 抽出対象となる過去問題

古舘らが定義した知識ベースでは，岩手県盛岡市のご当地検定試験である「盛岡もの識り検定（以下もりけん）」の過去問題からの知識を抽出対象とした．本研究においてももりけんを対象に検討したあと，その他の検定試験への応用可能性を検討する．もりけんでは，1級から3級まで存在しており，主に，多肢選択形式と一問一答形式の二種類の問題が出題されている．しかし，知識を抽出する際は，一問一答形式問題と多肢選択形式問題ともに問題文と解答から知識を抽出し，多肢選択形式問題の誤答選択肢は知識抽出の対象外としている．また，古舘らはこれらの中でも以下の出題形式を抽出の対象外としてきた．

- 提示した画像に適した単語，または問題文の条件に適合する画像を選択する形式
- 問題文に含まれる伏せ字に当たる文字を選択，または記入する形式（穴埋め問題）

本稿ではこの2つの出題形式から知識ベースの自動構築が可能な手法を提案する．

3. 知識ベース自動構築の手法

本研究では，画像利用問題と穴埋め問題を7つの形式に判別したあと，判別された形式に応じて対象知識の出現箇所を割り出して対象知識を抽出する．

3.1 画像利用問題と穴埋め問題の形式自動判別

平成22年度から27年度のもりけんの過去問題（合計1500問）に含まれる出題形式を分析した．分析結果を表1に示す．分析の結果，画像を用いる問題と穴埋め問題は7つの形式に分類された．また，各形式と対象知識の出現箇所を分析し，対象知識の出現箇所の自動決定方法を検討した．出題形式と対象知識と出現箇所の自動決定手順を図1示す．

表 1 画像を用いる問題及び穴埋め問題の形式

No	形式	問題数
1	問題文の伏せ字に当てはまる単語や文字を選択, または記述する形式	20
2	地図で指定された箇所の中から, 問題文の条件に適合するものを選択する形式	4
3	地図で指定した箇所にある物の名称を選択する形式	7
4	問題文で挙げた条件に一致する画像を選択する形式	3
5	写真(または絵)に含まれる物の名称を選択, または記入する形式	15
6	写真(または絵)に含まれる物に関連する単語を記入または選択する形式	9
7	写真(または絵)に含まれる物の用途や状態を選択する形式	3

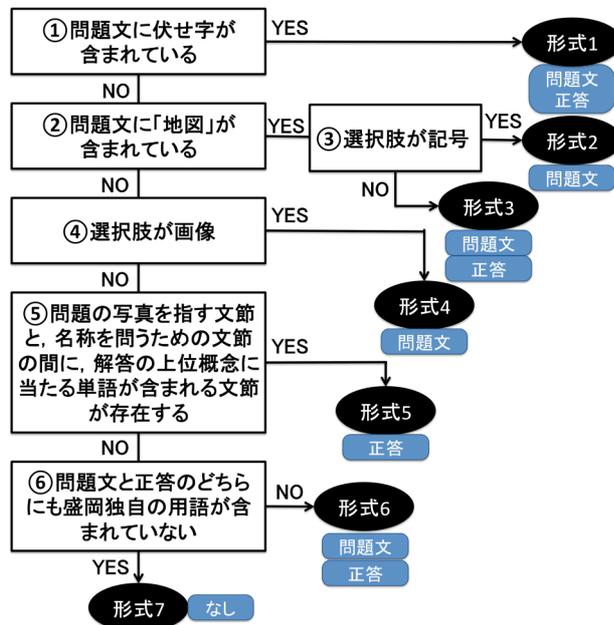


図 1 出題形式の判別と対象知識の出現箇所の決定手順

3.2 構成要素の自動抽出

知識ベースの構成要素は図 1 の「対象知識」の出現箇所の特定後に抽出する。「カテゴリ」の決定は古館らによる手法同様、日本語 Wikipedia オントロジー (JWO : Japanese Wikipedia Ontology) (4) を利用して行う。

形式 1 では、対象知識となり得る専門用語の中に伏せ字が挿入されている可能性がある。そのため、問題文の伏せ字に正答を挿入した文章から対象知識を抽出する。プロパティとオブジェクトは、それぞれ問題文と正答から抽出する。

形式 2 は、問題文中に「地図」が含まれ、尚且つ選択肢がアルファベットのような記号のみであれば分類される。そのため、対象知識は問題文から抽出し、プロパティ、オブジェクトはそれぞれ地図情報(画像)が含まれた問題文と正答から抽出する。

形式 3 は、問題文に「地図」が含まれるが、選択肢が記号ではなく用語であれば分類される。要素の抽出では、対象知識は問題文と正答の両方を対象に抽出し、プロパティとオブジェクトは形式 2 と同様に抽出する。

形式 4 は、問題の選択肢が画像であれば分類されるため、問題文から対象知識を抽出し、オブジェクトを正答の写真から抽出する。また、プロパティは自動的に「写真」とする。

形式 5 には、「この写真の」のような問題で上げた写真を指す文節と、「の名前は何か」のように名称を問うための文節が存在する。これらの文節の間に正答の上位概念となり得る名詞が存在する場合、写真に含まれている物の名称が問われていると判断する。この形式では、正答から対象知識の抽出を行い、オブジェクトとプロパティの抽出は形式 4 と同様に行う。

形式 1~5 以外の問題を形式 6, 7 とし、この中でも問題文と正答のどちらにも独自用語(岩手や盛岡が含まれる複合名詞, 人物名, 建築物名など)が含まれていないものを形式 7 として決定する。形式 7 は、知識ベースを構築する上で情報が少ないと判断し、本稿では抽出の対象外とした。形式 6 では、問題文と正答から対象知識を抽出する。それぞれ写真情報(画像)が含まれた問題文と正答から抽出する。

4. まとめと今後の課題

本稿では、先行研究による知識ベース自動構築機能において、これまで抽出の対象外とされてきた問題の形式の内、画像を用いる問題と穴埋め問題の形式から知識の抽出が可能になる手法を新たに提案した。今後は、本提案による自動構築機能を開発し、これを情報源として自動生成した問題を評価することで、本知識ベースの妥当性を検証する。

謝辞

本研究に協力して頂いた盛岡商工会議所ととりけん作問委員会の皆様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) 古館昌伸, 高木正則, 高木輝彦: “試験問題の自動生成を可能とする知識ベース自動構築手法の提案と評価”, 研究報告コンピュータと教育 (2015)
- (4) 玉川奨, 香川宏介, 森田武史, 山口高平: “日本語 Wikipedia オントロジーの構築と利用”, 人工知能学会 第 29 回セマンティック ウェブとオントロジー研究会 (2013)