

知識ベースに基づいた学習者個別演習問題生成手法について

高野 敦子 橋本 淳
兵庫大学 経済情報学部 経済情報学科

本研究は、各学習者の理解度と興味に合わせて算数の文章問題を自動生成し、その解法の習得を支援する、WBT (Web Based Training) システムの実現を目指す。本稿では、代表的な文章問題である鶴亀算を取り上げて、WBT システムの実現方法を述べる。まず、その構造を分析し問題の基本構造と難易度の記述を試みる。次に、その基本構造から具体的な問題を自動生成するために必要な知識ベースを構築する。そして、その基本構造と知識ベースを用いて、有名な基本問題の単純なオブジェクトの置き換えによる類似問題はもとより、表面的にはかなり違った応用問題までが自動生成可能であることを示す。問題生成時に難易度とテーマを指定することにより、学習者の理解度と興味を考慮できる仕組みになっている。

Drill Exercise Generation based on the knowledge base

Atsuko Takano Jun Hashimoto
Dept. of Economics & Information Science, Hyogo University

Students will heighten their thinking faculty by solving various arithmetical word problems repeatedly. In this research, we aim at designing a drill type WBT that generates exercise problems considering the progress and interest of learners. In this paper, we collected arithmetical word problems, and extracted structures of them. Then we propose a model of drill exercise generation using the knowledge database. The experimental evaluation of this model is also reported.

1. はじめに

近年、大学生の思考力の低下はますます問題となっており、小学生の時期から思考力を鍛えることの重要性が認識されている。これに対して、方程式などを用いずに、鶴亀算などの各種の算数の文章題を解くことは、小学生としての思考力を鍛えるよい題材であることが知られている。そのような文章題を繰り返し納得いくまで練習することが、将来高度な学習に役立つ思考力を養うと考えられる。^[1]

そのためには、それぞれの生徒に対して理解力や思考速度に合わせた問題を順次与える必要がある。問題は次々違った問題でなければならないし、生徒への動機付けの点からも、各生徒の興味に合わせた題材を扱った問題であることが望ましい。しかし、多人数を一斉に教える授業では、そのようなことは難しい。

そこで、本研究では、各学習者の理解度と興味に合わせて算数の文章問題を自動生成し、その解法の習得を支援する、WBT (Web Based Training) システムの実現を目指す。WBT は自学自習のツールとして有効であることが広く認められている^[2]。また、一斉授業の中に個別指導を組み入れるための手段ともなり得る。本稿では、代表的な文章問題である鶴亀算を取り上げて、WBT システムの実現方法を述べる。まず、その構造を分析し問題の基本構造の記述を試みる。次に、その基本構造から具体的な問題を自動生成するために必要な知識ベースを構築する。そして、その基本構造と知識ベースを用いて、有名な基本問題の単純なオブジェクトの置き換えによる類似問題はもとより、表面的にはかなり違った応用問題までが自動生成可能であることを示す。例えば、SPI 試験の代表的な問題である食塩水の濃度の問題もこの鶴亀算の基本構造から同様の方法で生成することができる。

2. 鶴亀算の問題の分析

2. 1 問題の基本構造とその記述

算数の文章問題はしばしばいくつかのパターンに分けられて、そのパターン毎に解法が説明されるが、その中でも最も代表的なものが次のような鶴亀算と呼ばれる問題のパターンである。

【基本問題】

「足の数が2本のツルと足の数が4本のカメが合わせて6匹います。足の数は全部で20本になります。ツルは何匹いるでしょう。」

ここで、ツルとカメをボールペンと鉛筆に置き換え、足の本数を値段に置き換えることによって、次のような類似問題が容易に生成できる。

【類似問題】

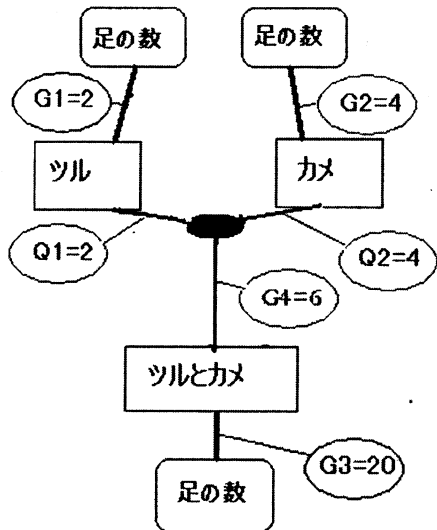
「値段が100円のボールペンと値段が50円の鉛筆を合わせて8本買います。

値段は全部で700円になります。ボールペンを何本買うでしょう。」

平島らはこのような問題の構造を、ネットワーク構造で表現している^[3]。本研究でも、問題構造の記述にネットワーク構造を利用する。ただし、平島らの研究では、文章問題をシステムの利用者が作り、システムはその理解や解法を導くために、問題の構造を利用する。一方、本研究では、システムが文章問題を自動生成するため、システムが最初から、文章問題の解法や答えを知っている。このような違いから、平島らの記述とは異なる図1のような表現方法を用いる。

ここで、G1からG4は問題文の中で、明にあるいは暗に与えられる値であり、Q1とQ2が学習者に求めさせる値である。

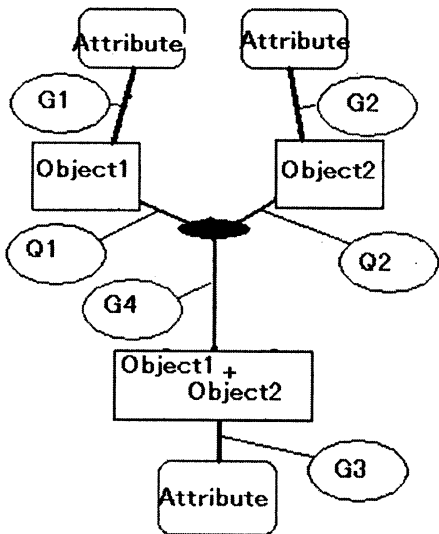
そして図1の構造をモデル化したのが図2である。図1におけるツルとカメを2つのオブジェクト (Object1, Object2) に、足の数をそれに対する属性 (Attribute) に、そして、オブジェクトの個数と属性の値を6個の変数 (G1からG4, Q1とQ2) に、というように一般化している。



成立する関係: $G4=Q1+Q2$
 $G3=G1 \times Q1+G2 \times Q2$

図1. 鶴亀算基本問題の構造表現

図2の構造を鶴亀算タイプ問題の基本問題構造と呼ぶことにする。



成立する関係: $G4=Q1+Q2$
 $G3=Q1 \times G1+Q2 \times G2$

図2 鶴亀算タイプ問題の構造表現

図2において、オブジェクトと属性を表す内容未定ノードと6個の変数に適当な概念や数値を代入することによって類似問題などの問題を生成す

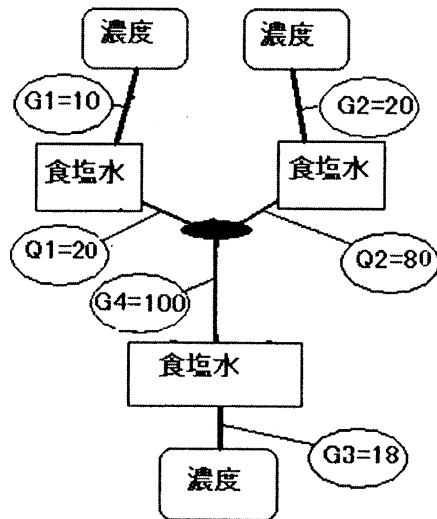
ることができる。

ここでさらに、次の問題もほぼ同様の構造で表現できることを図3に示す。

【発展問題（食塩水の濃度の問題）】

「濃度が10%の食塩水と濃度が20%の食塩水を合わせて100gの食塩水を作ります。

全体の濃度は18%になります。10%の食塩水は何gでしょう。」



成立する関係: $G4=Q1+Q2$
 $G3=(Q1 \times G1+Q2 \times G2)/G4$

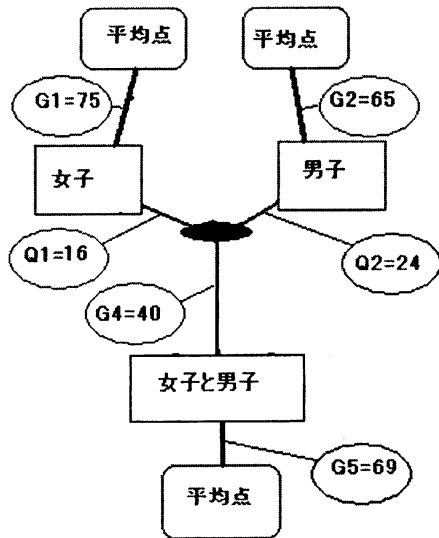
図3. 食塩水の濃度の問題の構造表現

図3が、基本問題構造と違うのは変数G3の式が図2におけるG3の式をG4で割ったものになっているという点のみである。G3がこのような式になるのは、属性が濃度、平均値、速さといった割合として表される場合である。

例えば、次のような平均値の問題も図3と同様の構造で記述できることを平嶋らが指摘している。

【発展問題（平均値の問題）】

「平均点が75点の女子と平均点が65点の男子が合わせて40人います。全体の平均点は69点になります。女子は何人いるでしょう。」



成立する関係: $G4=Q1+Q2$
 $G3=(G1 \times Q1+G2 \times Q2)/G4$

図4. 平均値の問題の構造表現

そこで、属性を以下の2種類に分類して扱うことにする。

属性タイプ	属性例1	属性例2	属性例3
絶対数	個数	足の数	金額
割合数	平均点	濃度	速さ(分速)

表1. 属性の分類

そして、図2におけるG3の式を、

属性が絶対数の場合：

$$G3=G1 \times Q1+G2 \times Q2$$

属性が割合数の場合：

$$G3=(G1 \times Q1+G2 \times Q2) / Q4$$

と拡張することにより、食塩水の濃度の問題を含めた広い範囲の問題を鶴亀算タイプの問題として扱うことができる。

2. 2 問題の難易度

属性が絶対数の場合と、割合数の場合を比べると、G3を求める式からも分かるように、発展問題の方がより鶴亀算の問題構造の理解が定着している必要があることが予想できる。また、「ツルの足は2本」というような常識として学習者が知っていると考えられる情報は問題文の中では省略し

て明記しない方法もある。これも問題構造を見えにくくする要因になるため、理解が曖昧な場合には、難易度が上がると考えられる。一方、問題文中の数値の観点からは、オブジェクト数などの数値のけた数が増えると間違いを起こしやすいことも予想できる。したがって、問題の難易度レベルを問題構造の点と、数値の桁数の点からの2つの観点から次のように分類する。これによって、学習者の進度に応じた問題を生成できるようにする。

(1) 構造レベル

レベル0：属性が絶対数で必要な情報をすべて問題文に明記

レベル1：属性が絶対数で常識として学習者が持っていると考えられる知識（G1とG2）は問題文中では省略

レベル2：属性が割合数

(2) 数値レベル

レベル0：G1およびG2の桁数が1桁

レベル1：G1またはG2の桁数が2桁以上

3. 知識ベースの構築

3. 1 概念シソーラス

2章で求めた基本構造から実際に問題文を生成するためには基本構造において不定となっているノードや変数に概念や値を当てはめる必要がある。そのために必要な知識ベースを構築する。問題はノードに代入する2つのオブジェクトと属性と3つの最初から値を設定する3つの数値によって決定されるため、それぞれのデータの可能な組み合わせの数だけ問題ができることになる。

そこで、問題集から収集した40題の鶴亀算パターン問題に必要な知識を知識ベース化した。

まず、オブジェクトからシソーラスを生成すると図5のようになった。シソーラス化することにより、個々のオブジェクト毎に対応する属性やそれに付随するデータは、できるだけそれらのオブジェクトの上位の概念に対するデータとして記述できるようにする。つまり、概念に付随するデー

タは下位で再定義されない限り下位の概念にも継承される。

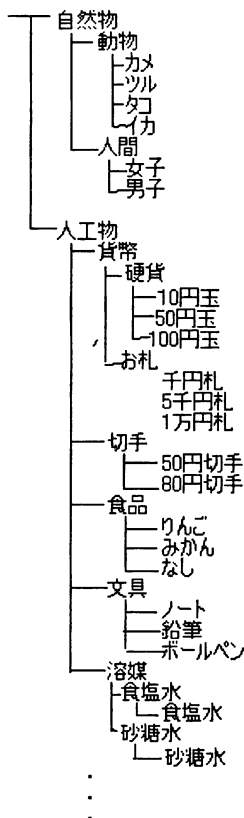


図5. 概念シソーラス

{ノート, 鉛筆, ボールペン} のような一番下位の概念は問題文に現れるオブジェクトとなる概念なので、それらの各集合をオブジェクトレベルの概念群と呼ぶ。

3. 2 概念データ

次に、オブジェクトが取り得る属性など、オブジェクトに対して問題生成上必要な知識をその上位に位置する概念に対して定義する。例えば、カメ、ツル、タコなどが属性として、「足の数」をとる、という知識は「動物」という概念に対して、定義する。

この知識を概念データとし、以下の知識を記述しておく。

- (1) 取りえる属性の集合
- (2) 属するオブジェクトを数える際の単位
- (3) 属するオブジェクトを数える際の基本単位 (オブジェクト数はこの整数倍とする)

【例】食塩水の概念データ

- (1) {濃度}
- (2) グラム
- (1) 10

3. 3 属性データ

次に、属性の値 (属性値) の単位、など属性に対して、問題生成上必要な知識を属性に対して定義する。例えば、「足の数」という属性に対する属性値の集合は {2,4,8,10} とする。

この知識を属性データとし、以下の知識を記述しておく。

- (1) 属性値の単位
- (2) 属性値の取りうる集合
- (3) 取りうる動詞
- (4) 属性値に対する助詞

3. 4 個別知識

固有のオブジェクトと属性の組に対して特に規定される値。例えば、「足の数」という属性に対する属性値の集合は {2,4,8,10,} であるが、オブジェクトをツルに特定したとき、{2}に限定されるという知識などである。

4. 練習問題自動生成方法

4. 1 問題レベルとテーマの設定

学習者の進度から、生成する問題のレベル (構造レベル0から2, 数値レベル0から1), さらにテーマとしてシソーラス中の概念を設定する。

4. 2 問題構造の作成

問題基本構造の内容不定ノートと変数部にオブジェクト, 属性, 属性値を代入して問題構造を作成する。

- 1) 設定した概念の下位に位置するオブジェクトレベルの概念群からオブジェクト *OBJ1* を選択して Object1 に代入

概念群が唯一のオブジェクトから成る場合

Object2 と Object3 に *OBJ1* を代入

それ以外の場合

同じ概念群から *OBJ1* 以外のオブジェ

クト *OBJ2* を選択して Object 2 に代入

Object3 に「*OBJ1* と *OBJ2*」を代入

2) 概念群に対応した属性の中から属性 *Attr* を選び、Attribute に代入

3) *OBJ1* の *Attr* に対する属性値群の中から 1 つを選び G1 に代入 (数値レベルの桁数考慮)

4) *OBJ2* の *Attr* に対する属性値群の中から 1 つを選び G2 に代入 (数値レベルの桁数考慮)

5) *OBJ1* の基本数 × 乱数(1 から 10 までの整数)を Q1 に代入

6) *OBJ2* の基本数 × 乱数(1 から 10 までの整数)を Q2 に代入

7) $G4=Q1+Q2$ を代入

8) *Attr* が絶対数の場合 (構造レベル 0,1)

$G3=Q1 \times G1+Q2 \times G2$ を代入

Attr が割合数の場合 (構造レベル 2)

$G3=(Q1 \times G1+Q2 \times G2) / G4$ を代入

4. 3 問題構造から問題文への変換

今回のように、限られた内容の文を生成するためには、文の骨組みを作っておいて、その中の単語を適切に決めることによって文を作成するロット法^[4]が有効である。

問題文は次の 3 つの文から構成される。

第 1 文目 (【】内は概念レベル 1 の場合は省く)

【<Attribute> が <G1><S1> の】<Object1> と

【<Attribute> が <G2><S1> の】<Object2><S2>

あわせて<G3><S4><S3>ます。

第 2 文目

<attribute>は、全部で<G4><S1>になります。

第 3 文目

<Object1>は何<S1>でしょう。

ここで、

<S1> : 属性値の単位

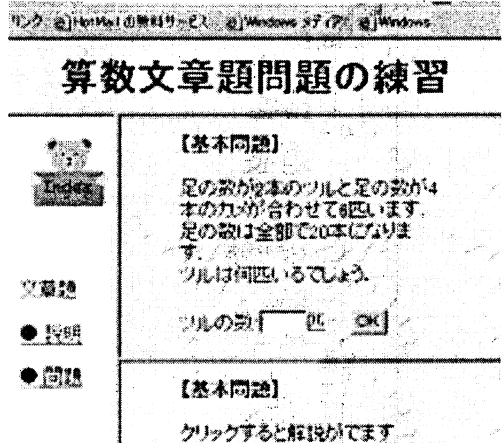
<S2> : 属性に続く助詞

<S3> : 属性に対応する動詞の連用形

<S4> : オブジェクトを数える時の単位

5. システムの動作例

作成したシステムでは次のような画面で中段右側のフレームに問題が出題され、学習者が答えを入力し、その後、正解が表示される。必要に応じて、下の段に解法の説明を表示するようにする。



6. 評価

システムはまだ作成中であるが、新たに問題集から収集した 20 個の問題のうち、18 個はこの方法で生成できることがわかった。さらにランダムに知識を取り出して問題を作成した結果 50 問中 41 問は妥当な問題が作成できたが、一旦作成した後のチェック機能が必要なこと、自然な文にするための機能を追加する必要があることがわかった。

参考文献

[1]認知工学：“つるかめ算差集め算の考え方”
[2]金西,林,光原,矢野：“教材知識に基づき WBT 上で演習問題を生成する機能の実現”,教育システム情報学会誌 Vol.20, No.2, pp.71-81, 2003

[3]平嶋,梅田,志岐,竹内：“XMLを用いた算数の文章問題の作成・共有環境”,教育システム情報学会誌 Vol.18, No.3&4, pp.284-296, 2001

[4]長尾真編：“自然言語処理”,岩波講座ソフトウェア科学,pp.337,1996.