

作問学習支援システムのための問題診断機能の評価

中野 明

平嶋 宗

竹内 章

九州工業大学

1. はじめに

本研究では、「問題を作ることによる学習」の実現を目指して、コンピュータ上で学習者に文章問題を作らせ、その文章問題を計算機が診断/評価/作問の誘導/誤りの修正支援/課題の提示などの処理をする知的な作問学習支援環境の開発を行っている。本システムは、文章作成を行うためのツールとフィードバックを提供する GUI、通信機能、文章問題を診断する機能、個別指導のための機能、を持つ。また、本システムはサーバ・クライアント方式を採用しており、クライアントであるインタフェースを Java 言語で実現することで、ブラウザを用いたマルチプラットフォームな利用が可能である。¹

すでに我々は、本システムで行っている作問学習の手順、提供する各機能の説明、本システムの診断メカニズムにより同定可能な誤りの整理、作問指導に用いるために文章問題の難しさ算定方法、本システムを用いた予備的評価実験に関しての報告を行っている[1]、[2]。

本稿では、本システムの問題診断機能の評価について報告する。まず、2.では、問題診断機能が検出している誤りについて述べる。3.では、評価実験の方法とその実験結果、および考察を行う。

2. 問題診断機能により検出される誤り

学習者の作る問題のバリエーションは、本システムのように制限されたインタフェースを用いても極めて多いと考えられ、誤りを含めた全ての場合を予め用意するのは不可能である。本システムは、指定された解法を用いて、学習者が作った問題の解決を試みる。この問題解決には、平嶋らが以前に提唱した問題解決モデル MIPS を用いている[3]。MIPS による問題解決の過程では、問題がその解法で解決できるかどうかだけでなく、どのような情報が不足しているのか等を診断することができる。

本章では、本システムの診断する過程(学習者の作った問題文を解釈し意味ネットワークを生成する過程・解法が適用可能かを調べる問題構造の検証の過程・解を導出し数量の検証を行う過程)に沿って、以下に示す誤りを検出することが可能である。例としては解法がツルカメ算の場合で説明を行う。

(1).意味ネットワークの生成

(1-a).基本関係の誤り

基本関係とは、概念・属性・属性値で構成される意味ネットワークの基本構成単位であり、多くの場合、入力文1文に対応している。

(1-a-i).入力した他の基本関係との比較

E1: 数量に関する曖昧性や矛盾が発生する

W1: 重複している

(1-a-ii).基本関係と対象領域に関する知識との比較

E2: 事実関係の誤り。(例) ツルの魚は～。

E3: 事実値の誤り/事実値の範囲の違反。(例) ツルの足の数は4本/リンゴの値段が-100円。

(1-b).生成した意味ネットワークの検証

E4: 文間のつながりが欠如している。初期状態のネットワークとして結合される基本関係が全く存在しない場合で、数学的な文脈が取れない文章。

W2: 意味ネットワークを構成しない基本関係がある。

(2).問題構造の検証

(2-a).情報の不足

E5: 補完すべき情報をシステムが一意に同定できる場合(例) リンゴとミカンを使った問題の場合、どちらか一方の値段しか記述していない時。

E6: 補完すべき情報をシステムが一意に同定できない場合。つまり、複数の補完候補が存在する場合。

(2-b).解法の適用条件を満たさない

E7: 求める数量の指定がない

E8: 解が一意に決定しない(例) リンゴとミカンを使った問題で、2つの値段が等しい時。

(2-c).

W3: 問題解決に必要なでない文の存在。意味ネットワークとして結合されていても、問題解決自体には必要でない場合。(例) ツルとカメの足と匹数の問題で、ツルの羽の数などの記述がある時。

(3).数量の検証

E9: 答えが現実的な数量として適していない場合

計算結果の提示し答えが的確でない事を指摘、文章中の数値に関しての再考を促す。(例) 計算結果でツルの総匹数が小数や負の整数になる時。

3. 評価実験

本システムの問題診断機能の信頼性を調べるために、家庭教師などの学習指導を経験したことのある大学生・大学院生20人を被験者とする評価実験を行った。実験では、対象となる解法をツルカメ算とし、「ツルカメ算で解決できる問題の作問」を本システムの提供する環境下で行う。

まず実験の事前作業として以下の4つを行った。①被験者はツルカメ算を用いて問題解決演習を行う、②ツルカメ算の適用条件に関する説明を受ける、③本システムの GUI を用いた場合の文の作成方法に関する説明を受ける、④本システムの GUI の操作を慣れるまで利用する。

事前作業を経て、被験者は、解法利用と成立条件、解法適用可能な問題に関する作問、本システムを用いた文章表現、を再確認することができる。そこで、被験者に、「子供が作問する際に起こす誤り」を想定し作問してもらい(数学的な演算関係を持たない文章も含む)、本システムの診断メカニズムによる誤り同定の能力を調査した。尚、誤りの指摘方法として、複数検出される場合には、処理の初期に検出される(Error 番号の若い)誤りほど基本的な誤りとし、優先させ提示した。また、Warning は正解の時のみ提示した。診断結果を表1、図1に示す。

評価実験の後、被験者にアンケート調査を行った。項目は、①「ツルカメ算」の理解をする方法として作問は有用か、②作問の有用性を本システムは実現しているか、③作問の際に診断結果(誤りの指摘、修正を示唆するコメント)を受けることは必要か、④誤りの指摘は信頼できるか、の4項目である。アンケート結果を表2に示す。

3.1 考察

表1は、被験者の作った問題を診断した際の「誤りの検出状況」の側面から整理したものである。表1中の「診断数」とは、「被験者の作った問題」全てをシステムが診断した際の結果。表1中の「誤診1」は、被験者が「正しい問題」として入力したにも関わらず「誤りのある問題」とシステムが診断した際の、その診断結果の内訳を示す。表1中の「誤診2」は、「被験者の想定した誤り」と「診断結果の誤り」とが異なった数である。表1中の「誤診3」は、システムが「被験者の作った誤りのある問題」を「正しい問題」と誤診するケースを示している。そこで、被験者からのクレームは全て誤診1、誤診2、誤診3に分類される。表1中の「正診」は、妥当と評価される「システムの診断件数」の内訳である。以下に誤診1、誤診2、誤診3において得られたクレームの分析を行う。

¹本システムはすでに以下の URL で公開している。

<http://www.minnie.ai.kyutech.ac.jp/~nakano/problem-posing.shtml>
サーバの不具合等で利用できない場合・ご意見・感想は
nakano@minnie.ai.kyutech.ac.jp 以下にご連絡下さい。

表 1: 作問した問題の診断結果

誤り	診断数	誤診			正診
		1	2	3	
E1	4				4
W1	7				7
E2	78	5	4		69
E3	16				16
E4	1				1
W2	0				0
E5	39				39
E6	2				2
E7	58	3	3		52
E8	5				5
W3	0				0
E9	38		1		37
なし	125			1	124
合計	377(7)				360(7)

トもしくはイベントの数を答える問題だけがツルカメ算として扱われており、属性を数える問題は、消去算として扱われていた。ツルカメ算の解法のキーポイントは、二つあるオブジェクトのうち、全てを一方のオブジェクトと仮定して考えてみる、という点であり、属性の問題に対してはこのような説明は適切とはいえないためであると考えられる。本機能も現時点では、ツルカメ算の一般的な解釈に沿って実現されているが、解法の一般的な理解を考える上で、属性の数を答える問題を扱うかどうかを検討する必要があるといえる。

(1-c).ツルカメ算の解法を用いて導いた値を用いて答えを導く(E7と診断した3件):ある意味ネットワークの一部分で解法が適用できる場合、その一部分のネットワークを構成する概念を用いることで問題解決が可能である。このクレームを解消するためには、解法適用後、解の情報を再度意味ネットワークに追加し、数量的な整合をとるための別の問題解決手法を準備する必要がある。

(2).誤診2の分析

(2-a).指摘した点の説明が不十分である、または、被験者の直感とが異なっている場合(E2と診断した中の2件,E7と診断した3件,E9と診断した1件):クレーム対象(被験者の入力した問題)の内容確認を行った際、システムの診断は的確であったと思われる。このクレームに対しては、診断のコメント(自然言語表現)を被験者が理解しやすい表現に修正することで解決可能。

(2-b).ツルカメ算においてシステムはオブジェクトの属性値が0となる問題を許容しない(E2と診断した中の2件):システムでは、入力された文中に属性値0が含まれる問題(たとえば、クモとチョウとでのツルカメ算において、クモの羽の数=0を用いる問題)を許容せず、属性値0となる基本関係を誤りとして検出するためE2と診断される。しかし、被験者にとっては、「属性値が0の場合のツルカメ算」と認識しているため、このクレームのような誤解が生じる。被験者への説明不足が原因と考えられ、文章表現を修正することで解決可能。

(3).誤診3の分析

「女子の平均点は2.0点です」と明記しているにも関わらず、答の段において「女子の人数は0人」とし「正しい問題」と診断する。この場合、オブジェクトの数量(女子の人数)が存在しないにも関わらず、その属性(平均点)が存在しており矛盾が生じる。このような例は、属性に「統計的な概念」が割り当てられた場合に生じるため、それら特別な概念に対しては知識を追加することで解決可能。

(1).誤診1の分析

(1-a).システム内の知識不足により扱えない(E2と診断した中の2件):知識の追加により解決可能。

(1-b).属性の数を答える問題を扱えない(E2と診断した中の3件):方程式として考えた場合、属性の数を答える問題(たとえば、リング1個の値段とミカン1個の値段の合計とリング3個とミカン4個の値段の合計を与えて、リングとミカンのそれぞれの単価を求める問題)はツルカメ算と同じ形式を持つ。しかしながら、著者等が調査した問題集5冊のうち、4冊では、オブジェク

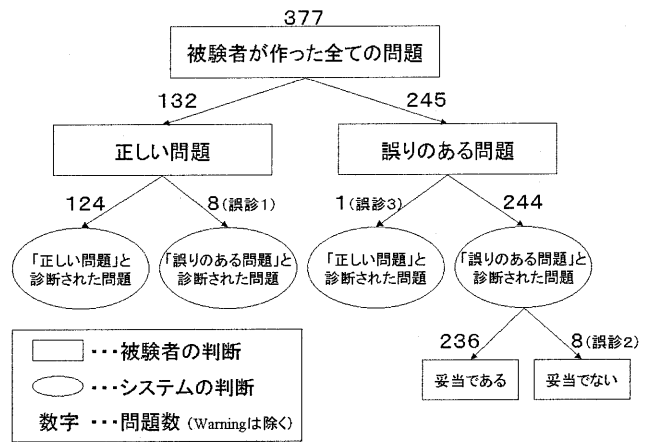


図 1: 作問した問題の診断結果 2

図1は、被験者の作った問題を診断した際の「システムの診断能力」を客観的に観ることを目的とした整理である。図中の誤診1, 誤診2, 誤診3は表1と同様。Warningを除く被験者の作った問題377問のうち360問(「正しい問題」を正しいと診断した124問+「誤りのある問題」の誤りを診断した336問)を的確に診断したことが確認できる。

表 2: アンケート結果

	優	良	可	不 可	わから ない
①	5	11	2	0	2
②	7	10	1	1	1
③	15	4	1	0	0
④	5	11	4	0	0

表2のアンケート結果では、③診断機能の必要性は高く認められているが、④信頼性の評価はさほど高くなかった。このことは、診断機能強化の示唆といえる。

4. おわりに

我々は、コンピュータ上で学習者に文章問題を作らせ、その文章問題を計算機が診断/評価/作問の誘導/誤りの修正支援/課題の提示などの処理をする知的な作問学習支援環境の開発しており、その例として、問題解決モデルMIPSを実装した作問学習支援システムを報告してきた。本稿は、本システムの診断メカニズムにおける信頼性の確認としての調査をまとめたものである。

2.では、本システムの診断機能が検出する誤りの種類について述べ、3.では、診断機能の信頼性を確かめるための評価実験について報告している。評価実験の結果、「MIPSを用いる診断メカニズム(人間の問題解決過程に沿った診断メカニズム)」により問題の誤りを細かく同定・指摘する能力があることが確認された。しかし、利用者の理解状況を反映したコメントにまで指摘点を昇華させることはできなかった。また、改良すべき点も数点得ることができた。

今後は、評価実験で得られた問題点を改良し、また計算機を使い慣れない学習者にでも容易に本システムを使えるようにインタフェースの洗練を行った上で、小学校教諭、小学生、を被験者とするアンケートや学習効果の検証実験を行うことを予定している。

参考文献

[1] Nakano, A., T.Hirashima, A.Takeuchi, "Problem-Making Practice to Master Solution-Methods in Intelligent Learning Environment", Proc of ICCE'99, 1999.
 [2] 中野 明・平嶋 宗・竹内 章:「問題を作ることによる学習」の知的支援環境,電子情報通信学会論文誌, Vol.J83-D-1, No.6, pp.539-549, 2000.
 [3] 平嶋 宗・中村 裕一・池田 満・溝口 理一郎・豊田 順一:「ITSを指向した問題解決モデル MIPS」,人工知能学会誌,Vol.7,No.3,pp475-486, 1992.