

学習対象の情報構造に基づくシステムデザインにおける 操作データからのラーニングアナリティクス

Learning Analytics of Log Data in Information-Structure Based Learning Environment Design

林 雄介^{*1}
Yusuke HAYSAHI

平嶋 宗^{*1}
Tsukasa HIRASHIMA

^{*1} 広島大学大学院工学研究科
Graduate School of Engineering, Hiroshima University

Learning analytics show promise for the basis of learning support. This helps us to find correlations among data we can get from learning activities. This study proposes an approach to analysis of learners' thinking in learning activities based on information-structure-oriented learning support system design. The key approaches to learning analytics of thinking are information structure of learning content and operation of the structure. We define learners' thinking as the operations, and design and develop learning environments where learners conduct the operations. This paper report examples of such environments and analysis of the log data as operations of information structure.

1. はじめに

教育分野においても、ビッグデータの利活用(教育ビッグデータ)も注目され、ラーニングアナリティクスや Educational Data Mining として、学習に関わるデータの測定、収集、分析、報告の方法について研究が多くなされているが、まだ議論が始まったばかりと言える[Romero 2013][Siemens 2012]. 教育ビッグデータへの期待は、これまで取得してきた顕在的データをより大規模に取得できることに加えて、これまで取得してこなかった、できなかった顕在的データを取得できるようになったこともある。山田は教育ビッグデータの処理の流れを「測定-解析-評価」としてまとめている[山田 2015]. 「測定」の部分では、比喩的に「センサ」と称して学習活動やその結果である事象のデータを収集する。その特徴の一つとしては、非常に大量のデータを対象とした相関性分析が挙げられる。立証が難しい因果関係ではなく、現象として観測できる相関関係に注目して処理することによって、従来は発見の難しい、人が想定しにくい相関性といった価値のある内容を見出すことが可能となる。

ICT を活用することによって、様々な学習活動データを大量に取得することが可能になってきている[緒方 2016]. その一方で、どんなデータを取得し、どのような観点で分析し、その結果をどのように学習支援に利用するかを明らかにすることが課題といえる[山田 2015].

本稿では、モンサクンと Error-based Simulation という二つの学習支援システムで得られたログデータからの分析を例として、学習対象の情報構造に基づいて設計された学習支援システムにおける操作データからのラーニングアナリティクスの可能性について議論する。

2. 情報構造指向アプローチからのラーニングアナリティクス

学習支援システム研究は、(1)利用すべき、あるいは開発すべき情報技術、(2)それによって設計できる、あるいは設計すべきシステム、(3)目指すべき、あるいは目指すことのできる教育・学習活動の 3 つの融合として定義でき、それぞれを中心とする(1)情

報技術シーズベース、(2)情報構造指向、(3)教育ニーズベースの3つに学習支援システム的设计・開発アプローチを分類している[平嶋 2013].

現状のラーニングアナリティクスをこの分類で位置付けると、(1)情報技術シーズベースの研究が多い状況であると考えられる。つまり、ビッグデータという情報技術シーズを主体に現状の技術で集めているデータの量の増加、新しいセンシング技術の利用による取得データの種類の拡大をし、適用事例を増やしながら、どのような教育ニーズに対応できるかを探究しているフェーズであるといえる。

このような情報技術シーズのアプローチも重要であると言えるが、本研究では、(2)情報構造指向のアプローチによって、これまで取得または推定が難しかったデータを対象としてのラーニングアナリティクスを提案したい。

学習の記録としてどのようなデータを蓄積することができるかについては、e ポートフォリオ研究において分類されており、そこでは学習に関して記録できるデータを学習記録と学習履歴、顕在的と潜在的という軸が定義されている[森本 2015]. 学習記録は学習者自らの入力を伴うもの、学習履歴はコンピュータシステムなどで自動的に取得可能なものである。顕在的データは学習成果物や授業風景などの記録、潜在的データは学習者の意識や意図、認知プロセス・思考を外化した記述データ、学習者の自己評価の省察や相互評価等のアセスメントの記述データなどのである。この中で、現状で得られる潜在的データは主に学習記録に当たり、学習者によって外化されるため、学習者特性や文脈に対する依存性が高いことが特徴である。

既存のラーニングアナリティクスに関する研究事例では、主に顕在的データレベル、もしくは学習記録としての潜在的データ、つまり評価活動に関するものについて行われている。例えば、緒方らは電子教科書の操作を対象として、操作ログ間や教員評価との相関関係分析などを行っている。さらに、学習ログを自己調整学習の観点から分析しており、ある種の学習者特性モデルを想定することで因果を想定し、予測を提案している。現状の多くのラーニングアナリティクスは顕在的データの学習履歴を扱っており、顕在的データの学習履歴、潜在的データの学習記録を扱う研究もある。これらに対して、本研究では潜在的データの学習履歴を扱うことに焦点を当て、そのために必要な情報技術を生み出すことを目標とする。

3. 情報構造指向アプローチによる組み立てることによる学習環境

現状のラーニングアナリティクスの取り組みは、あくまでセンサに入ってくる顕在的データの量を増やす、センサで測定する物理的なデータの範囲を広げると言える。本研究で提案するのは、それに対して、センサで測定するデータを潜在的データとして扱われる内容と関連付けやすいものに変更することである。現状の多くの潜在的データは学習記録として学習者や教授者が自ら入力するものであり、学習者に負担がかかったり、学習者のバイアスがかかった情報となっている。これを学習者に負担をかけず、学習活動を行っている中で学習履歴として自然にデータを取得できるようにすることが必要である。そこで、本研究では組み立てることによる学習[三輪 2014][平嶋 2015a]という学習形態に注目する。

「組み立てることによる学習」という学習形態は学習対象の構造を分析し、その構造で可能な操作として学習活動を考案し、それを可能にする環境を学習支援システムとして設計・開発するアプローチである[平嶋 2015b]。組み立てることによる学習で学習者に求められることは、学習対象の構成要素とその構成原理を認識し、何らかの条件の下でそれを満たす構造を組み立てることである。一般的な問題を解く課題では、(1)解法が決まっている、(2)確実に問題が解ける、(3)構造を部分的にしか利用しないといったことから、構造まで十分に考えなくても、総体的に見ずに局所的に解法を適用していけばある程度の確率で正答を導くこともできる。一方、組み立てることによる学習では、構成原理をコンパイルしたものといえる解法をベースにするのではなく、学習対象の構成原理をベースとして総体的に構造を見て、一貫性、整合性を持つものを作り上げることを要求されるために、より原理的な知識を身につけることができると考えられる。

組み立てることによる学習は、原理的な知識を要求すること、組み立てる、つまり、構成要素を組み合わせて構造を作る中で「組み合わせる」という行為を行うこと、この2つの関連が顕在的データからの潜在的データの推定を行う重要な要素となる。何と何を組み合わせることができるかは、学習対象の構成原理によって規定される。従って、学習者がある2つの要素を組み合わせた際、それが妥当であるかを判断できる。さらに、学習者の組み合わせ操作を記録していくことによって、どの段階でどの構成原理を用いたのかの記録になっていると同時に、誤った構成原理の適用をしているのか、ある構成原理の理解が誤っているのか、といった診断によって思考に関する潜在的データの推定が可能になる。

組み立てることによる学習を通じた潜在的データの推定は、比喩としては、学習者が数学の計算問題で途中の式変形を逐次書いていき、それを元に学習者の理解を推定することに似ている。数式を変形していく作業は、代数的構造という規定となる制約の下で、妥当な式変形を行うことであると言える。そして、妥当な式変形ができるのは、代数的構造を理解しているという前提が必要である。だから、計算結果が間違っていた場合、答えだけでは理解が不十分であると言えないが、途中の式変形が残されていれば、どこに誤りがあり、それがどのような代数的構造の理解の不十分さから起こったかを推定できる。

本研究では目指しているのは、このような、思考を学習対象の根底にある規定を理解し、その制約の下で対象を操作することと定義し、誤りがあった場合は規定の理解の不十分さによるものであると考えることができるようにすることを数学だけではなく、他の学習対象でも行えるようにすることである。数学では、例えば、代数的構造として明確に構成要素と構成原理が定義され、

それに基づいて行える操作として演算が決められているが、どのような対象でもそのようになっているとは限らない。しかし、多くの対象では、それは決められていないわけではなく、決めていない、決めるのが難しいだけであると考えられる。その規定を、その学習対象を構成する要素と要素間の関係、それに対する操作やそれによって起こる作用を整理することによって、様々な学習対象において組み立てることによる学習を成立させることができる。そして、その組み立てることによる学習を実現する学習環境を構築することで、要素の組み合わせ操作を学習環境上のユーザ操作という顕在的データとして記録し、潜在的データである思考を推定するために利用できるようになる。その具体例として、以下では、単文統合型作問学習環境モンサクン、初等力学を対象とした Error-based simulation の2つの学習支援システムでのログデータ分析から見えることを紹介する。

4. 算数文章題の作問学習における事例

4.1 単文統合型作問学習環境モンサクン

モンサクンは学習者に単文を提供し、それを使って作問することによって算数文章題を学習するシステムである[Hirashima 2007]。問題を作ることは、問題を解くことよりも学習効果があるとされ、モンサクンの実践の利用を通じた成果として、算数文章題の問題解決能力の向上が報告されている[山元 2013]。

モンサクンでは、図1に示すように左上に示される条件に合わせて、右側に示される単文を組み合わせることで算数文章題を学習者が作成する。この構成は算数文章題の三文構成モデル[]に基づいて設計されている。三文構成モデルでは、一つの四則演算子しか持たない数式で表すことができる算数文章題は3つの文で構成できることを示している。さらに、その3つの文のうち、2つは存在文という何らかのオブジェクトの数量を表す文であり、残り1つは関係文という2つの存在文で示される2つの数量の関係を表す文でなければならない。

加減算に関する三文構成モデルの特徴は、物語の種類に応じて、関係文の種類が異なることである。算数文章題には「合併」「増加」「減少」「比較」4つの種類の物語があるとされる。この4つの違いは三文構成モデルでは、関係文の違いとして表される[林 2013]。よって、物語の種類が指定された状況で作問するにあたり、学習者が考えるべき重要なポイントとして関係文の選択が挙げられる。

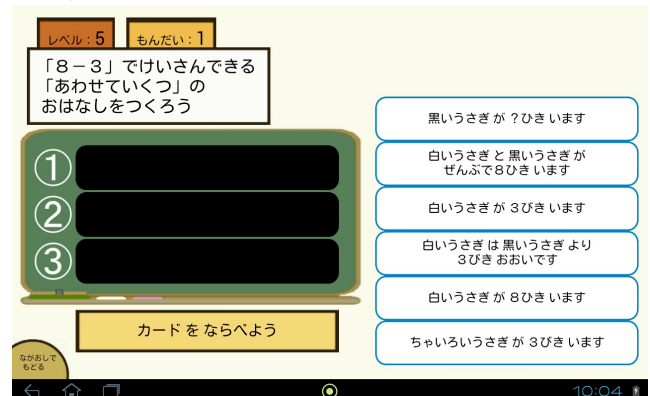


図1 モンサクン

4.2 課題設定

ここでは、理系学部の大学生による利用データの分析を紹介する。作問の条件としての未知数の設定と数式の提示の仕方を変えた3つのレベルを設定し、各レベル 24 課題、合計 72 課題

の作問結果を分析した。レベル1と2は物語の流れに合わせた数式を提示するのは同じで、未知数が右辺にあるか左辺にあるのかの違いである。例えば、りんごを3個持っていて、2個もらって、5個になったという増加の物語があった場合に、物語に沿うと数式は $3+2=5$ で表され、レベル1は $3+2=?$ 、レベル2は $3+?=5$ のように条件とする数式を提示する。一方、レベル3はレベル2と未知数の設定は同じであるが、物語に沿った数式ではなく、未知数を求めるための数式として $5-3$ を提示する。このとき、作成される文章題は、物語に沿った立式では加算となる増加の物語であるが、未知数を求めるためには減算をしなければならず、物語の種類から直接的に想起される演算と未知数を求めるための演算が異なるために一般的に逆思考問題とよばれ、算数文章題を学習する段階では難しいとされている。

4.3 データ分析

ここで分析対象とするのは、全72課題を作問した23名分のデータである。これを総操作数で k-means 法で3つに分けた結果を表1に示す。クラスタ1は1課題あたりの操作数の平均と標準偏差が最も小さいグループであり、クラスタ2, 3と大きくなっていく。それぞれのクラスタの被験者について、今回は各課題で最初に存在文と関係文のどちらを決定したかについて集計したものが図2である。これによると、レベル1では3つのクラスタに違いはないが、クラスタ1はレベル2から関係文の割合が増加しているのに対して、クラスタ2, 3はレベル2までは変わらず、レベル3で増加している。

表1 クラスタ毎の操作数

クラスタ	人数	操作数の平均	標準偏差	最大	最小
1	5	4.77	4.26	33	3
2	14	6.56	9.34	129	3
3	4	9.30	20.94	317	3

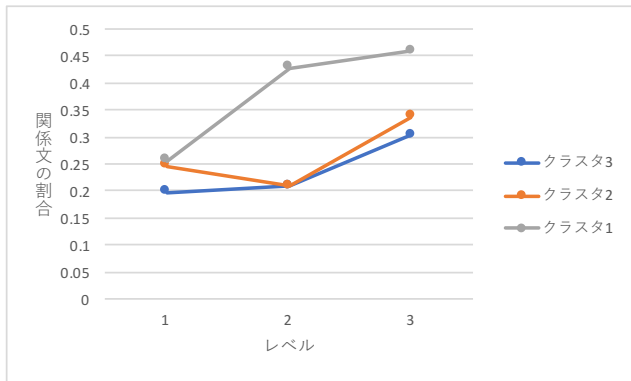


図2 関係文の割合の推移

レベル1, 2については、式の数値の順序と文の数値の順序が合うように物語を組み立てることができるが、レベル3では式の数値の順序と文の数値の順序が特に増加と減少では一致しないので数式と物語の関係の構造を意識しないとイケない。そのために、全てクラスタで関係文の割合が高くなっているが、クラスタ3の被験者はレベル1の段階で構造に気がつき、その構造を使って作問しようとしていると考えられる。

このように、表1に示すように単に結果としての操作数の多さといった指標だけではなく、プロセスに着目することによって、結果に関する因果を考察するためのデータが得られると考えられる。ここでは、最初に存在文と関係文のどちらを選ぶかだけに注目しているが、その後の選択も最終的にどのくらいの操作で正解にたどり着くかに影響すると考えられる。これについては、今後、詳細な分析をしていきたい。

5. 初等力学における力・速度・加速度の関係の学習における事例

5.1 Error-based Simulation

Error-based Simulation (EBS)は物体の運動について、例えば誤っていたとしても、学習者が記述した力・速度・加速度に基づいて、力学系における物体の挙動を提示するものである[Hirashima 1998]. 学習者の記述が誤っていた場合には、その誤りに基づいて物体の不自然な挙動を学習者に提示することで、学習者に正しい挙動との差異を認識させるものである。このシステムによる誤りの修正効果は実践的利用を通じて検証されている[篠原 2016].

また、初等力学分野においては Motion Implies a Force(MIF) 素朴概念とよばれるものが広く見られることが指摘されている[]。これは物体が運動している場合には、速度の変化の有無にかかわらず力が働いていると考えてしまうものであり、力学に関して一通り学習した後も、MIF 素朴概念を持つことが多いと指摘されているものである。EBSはこのMIF素朴概念の解消に対しても効果があることが示唆されており、今回の分析は、大学生においてもMIF素朴概念が発生するか、EBSの利用を通じて、発生したMIF素朴概念を学習者が解消できるかを調べた。

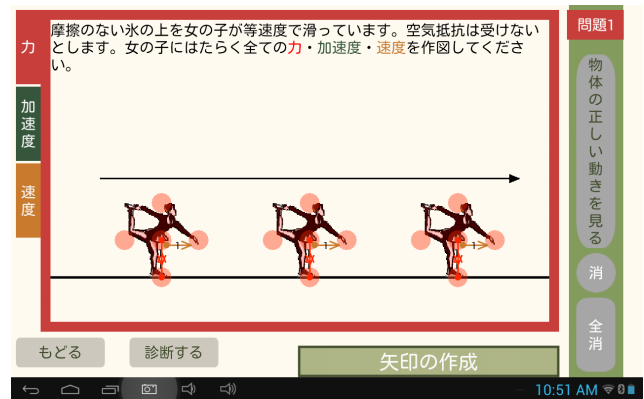


図2 Error-based Simulation (EBS)

5.2 課題設定

今回の分析では、工学系の大学生1年生を対象として収集したデータを利用した。実験の流れは、プレテストの実施、システム利用、ポストテストの実施となっている。ポストテストはプレテストと同じ問題を利用している。ここで扱っている課題は、平面上での等速直線運動、重力方向への等速直線運動、投げ上げの3つである。この種類の問題を各2問ずつでテストを構成しており、力だけを記述する。また、システム上の課題も同じようにこれらの種類でテストとは条件が異なる課題を設定しているが、こちらでは力・速度・加速度すべてを記述する。今回の分析の主眼は、MIF素朴概念が現れるか、そして、それがシステムのフィードバックによって解消されるかである。そして、その結果で学習効果が説明できるかを検討する。

5.3 データ分析

ここでは、まずプレテストとポストテスト結果からシステムによる学習効果を判断した。k-means法で3つのクラスタに分割したところ、c1~c3の3つのクラスタに分かれた。c1はプレとポストの両方でスコアが高かった被験者、c2はプレとポストの両方でスコアが低かった被験者、c3はプレからポストでスコアが上昇し

た被験者である。ここではこれらの 3 つのクラスタでシステム利用の中で違いがあるかを調査した。

まず、それぞれのクラスタの人数と回答数、回答毎の正解率をまとめたものを表 2 に示す。

さらに各クラスタについて、以下を調査する。

(1) クラスタ間で MIF 素朴概念の発生率に違いがあるか

(2) クラスタ間で MIF 素朴概念による誤りの修正にかかる回答数に違いがあるか

(1)について各クラスタでの MIF 素朴概念の割合を調べた結果を表 3 に示す。この結果から、c2, c3 は c1 に対して、有意に MIF 素朴概念の発生率が大きいことが分かる。

(2)について、各クラスタで MIF 素朴概念による誤りの次の回答でその誤りを修正できたかを確認した結果を表 4 に示す。この結果からは c1 と c2 に有意な差があり、c1 はすぐに修正できている割合が c1 の方が c3 より多いことが分かった。

以上の結果から、被験者全体において、MIF 素朴概念が発生しているが、c1 は EBS のフィードバックにより、すぐに誤りを修正できるのに対して、c3 はすぐに修正できないことが分かった。このことから、EBS のフィードバックは学習者が自身の誤りに対して気づかせることができること、その効果は学習者の理解に応じて違いがあると考えられることが分かった。

表 2 各クラスタの取り組み情報

	全体	c1 上位・ 変化無	c2 下位・ 変化無	c3 上昇
取組人数	135	96	6	33
総回答数	812	534	57	221
平均回答数	6.01	5.56	9.5	6.7
正解率	0.15	0.16	0.11	0.12

表 3 クラスタ毎の正解数・不正解数

	全体	c1 上位・ 変化無	c2 下位・ 変化無	c3 上昇
正解	601	446▲	12▽	143▽
不正解	2194	1491▽	83▲	620▲
解答の合計	2795	1937	95	763
正解率	0.23	0.13	0.19	0.22

▲有意に多い、▽有意に少ない、 $p < 0.05$

表 4 クラスタ毎の修正率

	全体	c1 上位・ 変化無	c2 下位・ 変化無	c3 上昇
すぐに修正 できた	970	700▲	37	233▽
すぐに修正 できなかった	1771	1162▽	82	527▲
誤りの合計	2741	1862	119	760

▲有意に多い、▽有意に少ない、 $p < 0.05$

6. おわりに

本稿では、学習対象の情報構造に基づくシステムデザインにおける操作データからのラーニングアナリティクスの提案として、モンサクンと Error-based Simulation という二つの学習支援システムを例として、システムの操作ログとして得られたデータを学習者による情報構造の操作として分析したものの最初の結果を紹介した。モンサクンと EBS のデータ分析から、システムの利用後のデータ、または利用前後のデータの比較と、学習者のシステム利用中のデータを照らし合わせることで、学習結果とその原

因として因果関係と考えられるデータとしてシステム利用中の振る舞いの違いを抽出できた。しかし、これらはまだ学習プロセスの一端を示すものであり、システム利用前後のデータと学習プロセスのデータを結びつけ、因果関係として確立させるための手法の確立について更なる検討を進めたい。

謝辞

モンサクンの学習データ取得にあたり協力をしていただいた尚美学園大学の山本樹先生に感謝いたします。本研究の成果の一部は科研費(15H02931)の助成を受けたものである。

参考文献

- [林 2013] 林 雄介, 他, 算数文章題における物語の構造分析", 第 68 回人工知能学会先進的学習科学と工学研究会 (SIG-ALST), pp.7-12, 2013.
- [Hirashima 1998] Hirashima, T., et al., Error-based simulation for error-visualization and its management, International Journal of Artificial Intelligence in Education, Vol. 9, pp. 17-31, 1998.
- [Hirashima 2007] Hirashima, T., et al., Learning by problem-posing as sentence-integration and experimental use, AIED 2007, pp.254-261, 2007.
- [平嶋 2013] 平嶋 宗, 学習課題の内容分析とそれに基づく学習支援システムの設計・開発:算数を事例として, 教育システム情報学会誌, Vol.30, No.1, pp.8-19, 2013.
- [平嶋 2015a] 平嶋 宗, 「学習課題」中心の学習研究—情報構造としての学習課題の再定義と構造操作としての学習活動の設計, 人工知能学会誌, 39(3), pp. 277-280, 2015.
- [平嶋 2015b] 平嶋 宗, 学習課題の情報構造としての再定義とその内容に基づいて設計された活動としての組み立てることによる学習, 第 29 回人工知能学会全国大会, 1B3-CS-2, 2015.
- [三輪 2014] 三輪 和久, 認知科学研究において「組み立てることによる学習」を考えることの意義, 第 28 回人工知能学会全国大会, 1B4-OS-12a-1, 2014.
- [森本 2015] 森本 康彦, eポートフォリオとしての教育ビッグデータとラーニングアナリティクス, コンピュータ&エデュケーション, 38, pp. 18-27, 2015.
- [緒方 2016] 緒方広明 他, 教育ビッグデータの利活用に向けた学習ログの蓄積と分析, 教育システム情報学会誌, 33(2), pp. 58-66, 2016.
- [Siemens 2012] Siemens, G., & Baker, R.S. "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration", Proc. of LAK2012, 2012.
- [篠原 2016] 篠原 智哉 他, 運動する物体にはたらく力を対象とした Error-based Simulation の中学校理科における実践利用, 電子情報通信学会論文誌, Vol.J99-D, No.4, pp.439-451, 2016.
- [Romero 2013] Romero, C., & Ventura, S.: "Data mining in education" Data Mining and Knowledge Discovery, 3(1), 12-27, 2013
- [山田 2015] 山田 恒夫: MOOCと学習解析:教育革新のための情報基盤に向けて, 情報処理学会論文誌教育とコンピュータ, 1(4), pp.1-11, 2015.
- [山元 2013] 山元 翔 他, 教室授業との融合を目的とした単文統合型作問学習支援システムモンサクン Touch の開発と実践利用, 電子情報通信学会論文誌 (D), Vol.J96-D, No.10, pp.2440-2451, 2013