

情報構造指向アプローチをベースとした教育ビッグデータ における相関関係から因果関係の導出

林 雄介[†] 平嶋 宗[†]

概要：近年注目を集めているビッグデータの特徴は非常に大量のデータを対象とした相関性分析であり、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を可能にしている。教育分野においても、ビッグデータの利活用が注目され、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。本稿では、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教育ビッグデータ研究を提案する。そして、その具体例として、本稿では第2言語としての英文読解支援を対象とした研究と単文統合型作問学習環境モンサクンでのログデータ分析の二つの研究を紹介する。

Causal Relations from Correlations in Information-structure Oriented Approach Based Educational Big data

YUAUKE HAYASHI[†] TSUKASA HIRASHIMA[†]

1. はじめに

ビッグデータが様々な分野で注目を集めている[5]。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。

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(Learning Analytics: LA)や Educational Data Mining (EDM)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[19][21]。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[23]。

平嶋は情報構造指向アプローチは学習支援システム研究を(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し、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の設計・開発アプローチを分類している[13]。

現状の教育ビッグデータ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教育ビッグデータ研究を提案したい。

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[8]。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。

既存の教育ビッグデータ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

[†] 広島大学大学院工学研究科
Graduate School of Engineering, Hiroshima University

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(8)。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。

現状の多くの教育ビッグデータ研究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。

2. 学習支援システムと教育ビッグデータ

2.1 学習と教授の理論

人の学習には、いくつかのパラダイムの変遷がある。学習を外部からの刺激に対する人間の観察可能な行動の変容と捉える「行動主義」、内的な情報処理過程に注目し学習を認知構造の変化として捉える「認知主義」、学習者の主体的な知識獲得過程として学習を捉える「構成主義」、そして主体的な知識獲得過程における他者や環境との相互作用に注目した「社会的構成主義」といったものである[3]。

しかし、これらの中にまったく共通点が無いわけではなく、学習の成果として個人の内部もしくは環境との相互作用の変容があり、それによって何らかの物理的な状態や行為が発揮されると考えられる。本研究での解釈として、これを図式化したものを図1に示す。

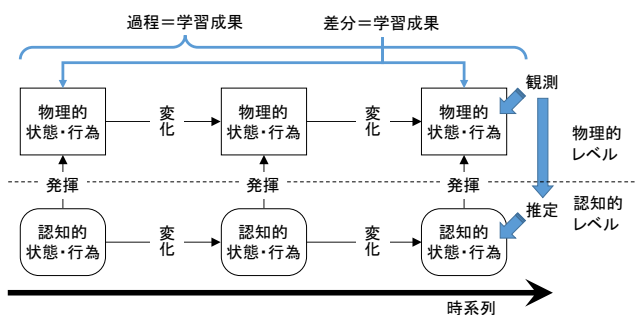


図1 学習活動

Figure 1 Learning activity

これは2層に分かれており、観測可能な物理的な状態や行為を表す物理的レベル、その観測から推定できる認知的状態や行為を表す認知的レベルとする。

行動主義、認知主義では物理的な状態や行為をサンプリングして観測し、その差分を学習成果とする。また、行動主義では、認知的レベルを想定しないが、認知主義では認知レベルまで想定する。そして、学習目標として行為や状態を設定しており、学習成果がそれに到達しているかどうかを評価基準とする。

一方、構成主義や社会的構成主義では、学習成果の目標

を設定することはせず、一連の状態や行為の変遷として学習成果を定義する。これらのパラダイムでは、学習の結果は各学習者が構成するものであり、統一された目標を目指すわけではない。

いずれのパラダイムにしても、学習者に促すことができること、観測できることは物理的レベルであり、それを通じて認知レベルに影響を与えたり、推測したりすることになる。

2.2 学習支援システム

学習活動が何らかの認知的状態・行為に基づき物理的な状態や行為の変化を発揮し、その結果をもって認知的な状態を変化させることとすると、そのような学習を支援するものとしての教授に関して、これまで様々な分析や試みが行われ、その知見として学習理論や教授理論と呼ばれるものが提案されている。これらの理論の特徴としては、学習理論は記述的理論であり、教授理論は処方的理論として、両方とも学習が成立するときの条件(状況)と成果、そのときに行われる活動(手法)の3つの組み合わせという同じものに対して、それをどう読むかの違いとなっていることである[18]。

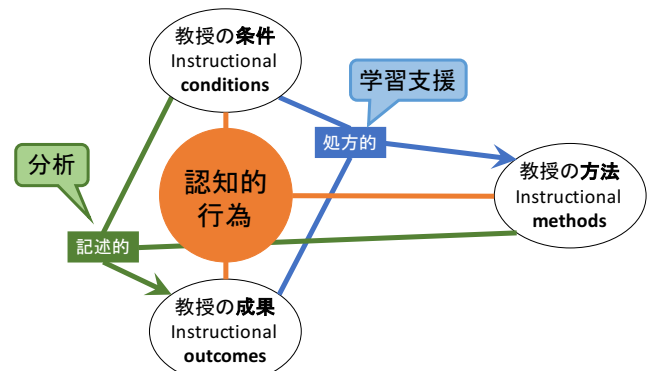


図2 記述的・処方的理論

Figure 2 Descriptive/Prescriptive theory

学習理論は、ある状況において、ある手法を適用したときに、どのような成果が得られたかを記述したものである。一方、教授理論は、ある状況で、ある成果を期待しているときに、どのような手法が効果的かという処方を示すものである。実践的、実験的に行われた数多くの事象から一般化することで学習理論が構成され、それを元に意図的に手法を適用して実験、実践を行うことで教授理論が形成される。両者は相補的なもので、この二つで形成されるサイクルによって学習とその支援についての研究が発展してきたと言える。

ここでは、学習支援についての研究を処方的な理論を構成すること、それを実現するための具体的なガイドラインやツールを開発することとして考える。もう一つ、前節で述べたように、学習活動のある状況と求める成果の組み合わせに対して行われる何らかの活動と定義すると、学習支

援の一つの形はある状況と求める成果の組み合わせに対して必要とされる活動を実施するための環境を整備することと考えることができる。

2.3 教育ビッグデータへの期待

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[23]。「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。

LMS (Learning Management System)は、これまで取得してきた顕在的データをより効率的に収集できるシステムであるといえる。テキストやビデオ講義の視聴、テストへの解答、フォーラムへの書き込み、資料のダウンロード、制作物の登録といったデータを大規模に収集できる。

一方、例えば、学習時の手書きデータは、以前は取得が難しかったが、近年のタブレット端末の普及によって比較的取得しやすくなっており、手書き情報自体をデータ化して処理できると共に、筆記の進行情報についても処理の対象にすることができる。そして、これを使って学習のつまづきの検出なども行われている[2][4]。

3. 情報構造指向の教育ビッグデータ研究

3.1 情報構造指向の学習支援

現状の教育ビッグデータの取り組みは、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[7][14]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[15]。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^[15]。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ることで「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しか言えないが、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無く、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な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、本稿では第2言語としての英文読解支援を対象とした研究と単文統合型作問学習環境モンサクンでのログデータ分析の二つの研究を紹介する。

3.2 キットビルド概念マップによる読解支援

第2言語としての英文読解支援を対象とした研究[1]では、英文を読むだけでなく、その内容を概念マップとして表現することで読解の支援とすることを目的としている。読解において概念マップを作成することの有用性は示されており[20]、そのような研究とキットビルド概念マップ (KB マップ) の違いは、学習者が概念マップを自由に作成するのではなく、教授者が文章に書いてある内容を概念マップとして作成し、それを分解して学習者に概念マップを作成する部品として提供することにある。図3に教授者が作成する概念マップ (ゴールマップ)、図4にゴールマップを分解したキット、図5にキットから作られた学習者マップを示す。これにより、理解の定着に効果があることが示されている[16]が、ここでは概念マップの作成順序に注目する。

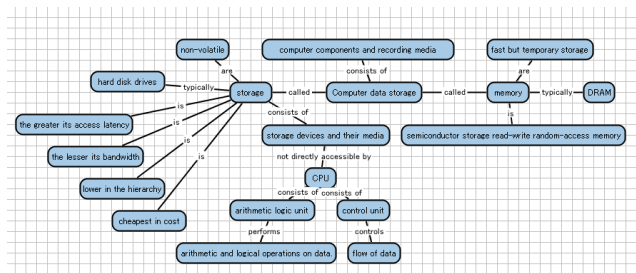


図3 ゴールマップ
 Figure 3 Goal map

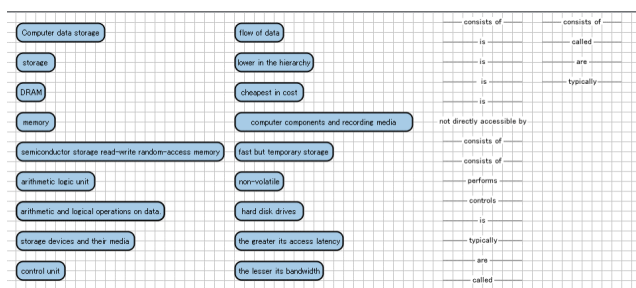


図4 キット
 Figure 4 Kit

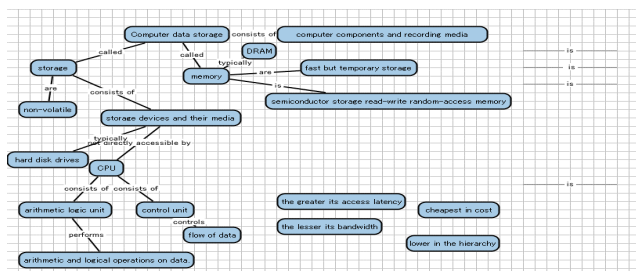


図5 学習者マップ
 Figure 5 Learner map

図6に二人の学習者がKBマップを作成した際のデータを示している。このとき、学習者はまず英文だけを一定時間内で読み、その内容を把握した上で、KBマップを作成している。このデータにおいて横軸は時間、縦軸は作成した命題の文章中の位置を示す。横軸はKBマップ作成開始が原点であり、右に行くにつれて時間が経過していることを示している。縦軸は下が文章の先頭、上が末尾となっている。従って、文章を先頭から順番に概念マップで正しく表していくと、右上がりの直線となる。学習者Aはそのように右上がりになっており、ほぼ英文の並び通りの順番で英文の内容をKBマップで作成している。一方、もう一人の学習者の方は一部に英文の順序から大きく逸脱している場所があることが分かる。これを他の学習者のデータ、他の英文でのデータを合わせて分析すると、KBマップでマップを作成すると、一般的な概念マップで作成することを比べると、一般的な概念マップの方が作成順序と対応する英文の並びが合っており、KBマップの方が順序のずれが大きかった。

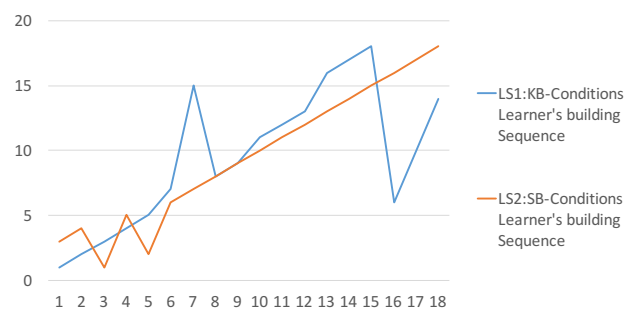


図6 マップ作成箇所と文章の対応

Figure 6 Correspondence between propositions and sentences

英文の読解は単語、文、段落といったように認識した内容を積み上げていくボトムアップ処理と一部から他の内容を推定して内容を認識していくトップダウン処理の2つがあり、その二つを組み合わせることが重要であると言われている[10]。マップ作成の順序が英文の流れに近いということはボトムアップ処理に対応していると考えられる。

3.3 モンサクンによる作問学習

2つ目の事例は単文統合型作問学習環境モンサクン[11][12][17][24]でのログデータ分析である。モンサクンでは、学習者はシステムから提示される単文を3つ組み合わせることで算数文章題を作成する。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、モンサクンでは、単文の組み合わせで問題を作成するため、与えられた課題の中で作成できる組み合わせが決まっている。現状のモンサクンでは5つまたは6つの単文が提示され、6つのときは、例えば、図8や9に示すような木構造ですべての組み合わせとそれらの間の遷移関係を表すことができる。ルートが何も選択されていない状態、展開して

いくと共に1つ、2つと選択された状態で、葉で3つ組み合わせさせた状態を表している。学習者はシステム上で3つの文を組み合わせて答え合わせを行い、正しく作問できるまで繰り返して取り組む。

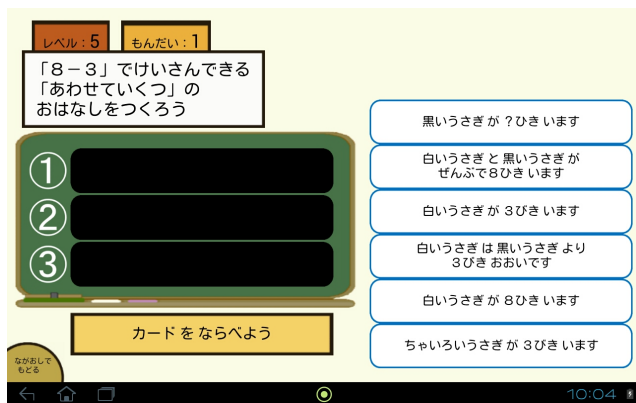


図7 モンサクン
Figure 7 Monsakun

二人の学習者の回答を途中の状態も含めて示しているのが図8と図9で円の大きさがその文の組み合わせを作った回数に対応している。両者とも最終的には正解に到達しているが、それまでに作ったものが同じではないことが分かる。モンサクンにおける学習者の作問活動の分析としてトラップステートという概念を定義している[22]。これは作問中に作られた1つ以上の文の組み合わせの中で、頻度が多い上に、その状態から正解に到達するまでに多くの遷移を必要としたものであり、ボトルネックになったと考えられるものである。

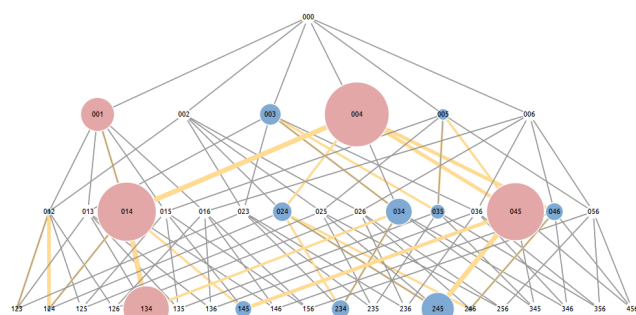


図8 学習者Aの作問プロセス
Figure 8 Problem-posing process of Learner A

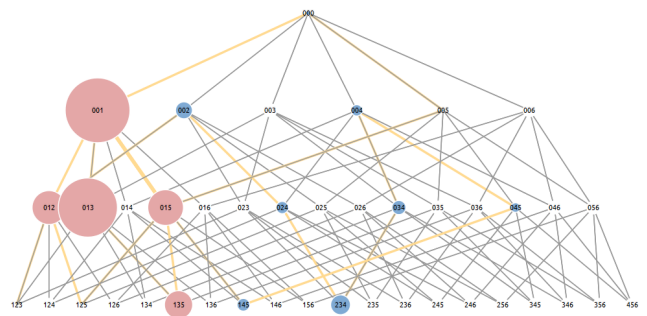


図9 学習者Bの作問プロセス
Figure 9 Problem-posing process of Learner B

図8と図9の学習者の作問プロセスからトラップステートを算出し、そこに含まれる正解には使われない文(ダミー)を抽出すると、図8の学習者では「白いうさぎが8ひきいます」、図8の学習者では「白いうさぎは黒いうさぎより3ひきおおいです」という文になる。図9の学習者では、課題の設定上、「8」は合わさった数になり、「白いうさぎと黒いうさぎはぜんぶで8ひきいます」という文で使われるべきものであるが、作られるべき話の中での「8」の役割が分かっていないために間違っていると考えられる。図?の学習者の方は「あわせていくつ」の話が求められているのにも関わらず、話の種類とそれらを表す文章表現を分かっていないために、「ちがいはいくつ」という違う種類の話を構成するための「白いうさぎは黒いうさぎより3ひきおおいです」という文を使ってしまっていると考えられる。このように、単純に誤りだけではなく、誤りの原因となる学習者の理解の不足や間違いを推定することができる。この例は個人を対象としているが、これを集積することで、特定の個人の間違い、クラス単位で間違いやすいところ、一般に間違いやすいところといった様々なスケールで学習者の理解の分析をすることができる。

3.4 2つの事例の比較

ここで挙げた二つの事例は、どちらもプロセスデータを用いることで、学習者の理解を推定する仕組みの提案である。情報システム上で演習を行い、そこでの操作が認知的行為と対応付けて設計されていることによって、観測・記録できる端末上での物理的操作の履歴から認知的な状態や行為の推定をすることができるようになって考えられる。ただし、この二つはドメインのモデルの深さの違いによって、実行できる分析に違いがある。KB マップによる英語の読解支援については、文章の内容をマップで表しているがノードやリンクのラベルの意味についてはシステムは解釈できない。ただ、教授者が設定した各命題と文章の対応を処理するだけである。一方、算数文章題の単文統合型作問演習では、各文の意味をオブジェクト、数量、述語で管理しており、組み合わせによってどんな式を作れるか、どんな種類の話になるかを判定できるようにしている。よって、

学習者が作成した文の組み合わせでどのような問題ができ、要求に合っているかを判定することができる。そして、その判定結果を使って適応的な支援をすることもできる(文献)。意味記述のコストと判定のリッチさには比例関係があり、より高度な支援をしようとする、その分の意味記述と処理のコストがかかる。

4. おわりに

本稿では、学習観の違い、記述的・処方的理論の考え方をベースに学習支援システム研究を位置付けた上で、情報構造指向アプローチで学習支援システムを設計することで、従来の情報技術シーズベースの教育ビッグデータ研究ではなく、学習支援の課題とより密接に連携させることができる情報構造指向アプローチをベースとした教育ビッグデータを提案した。そして、その具体例として、第2言語としての英文読解支援を対象とした研究と単文統合型作問学習環境モンサクンでのログデータ分析の二つの研究を紹介した。

これらの研究で得られた成果は、まだ情報構造指向アプローチに基づき設計した学習支援環境で得られたログデータの分析の第一歩であり、ビッグデータとしてデータのスケールを大きくすることで、モデルだけでは見えてこない実際の学習者の振るまいとその背後にある認知的状況や行為について調べ、学習支援にフィードバックできるようにしていきたい。

参考文献

- [1] Alkhateeb, M., et al.: "The Effects of KB - mapping Method to Avoid Sentence - by - Sentence Comprehension Style in EFL Reading", Proc. of ICCE2015, pp. 46-55 (2015)
- [2] 浅井洋樹, 野澤明里, 苑田翔吾, 山名早人, : "オンライン手書きデータを用いた学習者のつまづき検出", DEIM2012, 2012.
- [3] Cooper, P.A. (1993) Paradigm Shifts in Designed Instruction: From Behaviorism to Cognitivism to Constructivism, *Educational Technology*, 33(5), 12-19.
- [4] 中塚智尋, 森村吉貴, 橋本敦史, 飯山将晃, 村上正行, 美濃導彦: "ペンストロークの時間間隔を用いた答案の解答停滞箇所の検出", 教育システム情報学会研究報告 30(7), pp71-74, (2016)
- [5] 中野 美由紀: "ビッグデータ統合利活用における課題と技術", 電子情報通信学会誌, 97(5), pp. 343-347 (2014)
- [6] Nur Hasanah, Yusuke Hayashi and Tsukasa Hirashima: Investigation of Students' Performance in Monsakun Problem Posing Activity based on the Triplet Structure Model of Arithmetical Word Problems, Proc. of ICCE2015, pp. 27-36, (2015)
- [7] 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.
- [8] 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)
- [9] 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)
- [10] 樋口晶彦: 科学技術英文読解の一考察-スキーマ理論と言語的特徴-, 鹿児島大学教育学部研究紀要, 65: 19-29, 2014.
- [11] Hirashima, T., Yokoyama, T., Okamoto, M. and Takeuchi, A.: Learning by problem-posing as sentence-integration and experimental use, AIED 2007, pp.254-261 (2007)
- [12] Hirashima, T. and Kurayama, M.: Learning by problem-posing for reverse-thinking problems, *Artificial Intelligence in Education*, pp.123-130 (2011)
- [13] 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).
- [14] 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)
- [15] 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)
- [16] Hirashima, T. et al.: Framework of kit-build concept map for automatic diagnosis and its preliminary use, *RPTEL*, 10(1), pp. 1-21 (2015)
- [17] 倉山めぐみ, 平嶋宗:逆思考型を対象とした算数文章題の作問学習支援システムの設計開発と実践的利用, 人工知能学会論文誌, Vol.27, No.2, pp.82-91 (2012)
- [18] Reigeluth, C. M.: "Instructional-design: What is it and why is it?" In Reigeluth, C. M. (Ed.), *Instructional-design theories and models: An overview of their current status*, pp. 3-36, Lawrence Erlbaum Associates, Inc., 1983.
- [19] Romero, C., & Ventura, S. (2013). "Data mining in education" *Data Mining and Knowledge Discovery*, 3(1), 12-27.
- [20] Saeedi A. et al.: "Comparing Effectiveness of Methods of Presentation and Providing Concept Maps on Reading Comprehension", *Journal of School Psychology*, 2(3), pp. 125-143 (2013)
- [21] Siemens, G., & Baker, R.S. "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration", Proc. of LAK2012 (2012).
- [22] Supianto, A.A., Hayashi, Y. & Hirashima, T.: Visualizations of problem-posing activity sequences toward modeling the thinking process. *Research and Practice in Technology Enhanced Learning* (2016) 11:14. DOI 10.1186/s41039-016-0042-4
- [23] 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.
- [24] 山元翔, 神戸健寛, 吉田祐太, 前田一誠, 平嶋宗: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)