

データ・サイエンスとしての Learning Analytics の 方向性と研究上のフレームワークについて

安武 公一^{1,a)} 山川 修^{2,b)} 中村 泰之^{3,c)} 隅谷 孝洋^{4,d)} 多川 孝央^{5,e)} 井上 仁^{5,f)}

概要: Learning Analytics (LA) に対する関心がわが国でも高まってきている。しかしながら現在の LA で主流となっているのは、これまでの教育工学/学習科学の分野で標準的とされる方法論や分析視点に基づいたアプローチである。LA が本質的な意味においてビッグ・データを対象とするデータ・サイエンスを志向し、LA そのものの可能性を広げるためには、こうした従来の枠組みにとらわれないオルタナティブなアプローチを採用する必要があると、われわれは考えている。本稿ではこの点について議論する。

1. はじめに

おそらく国内で最初の Learning Analytics (LA) に関するシンポジウムを情報処理学会 CLE 研究会が 2011 年に開催^{*1}して以来、次第に LA に対する国内研究者の関心が高まってきている^{*2}。LA とは、

- (1) 学習者が日常的に残すあらゆる種類の情報 (ログ) を大規模に集積し (Volume),
- (2) そうした多種多様な行動履歴データをうまく解析することによって (Variety),

(3) これまででは見出すことが困難であった、学習戦略、授業設計の原則なども含め、効果的な学習環境や教授戦略等を (できれば、リアル・タイムに) 実現させる (Velocity),

ことを目的とした、主として高等教育改善のための包括的な研究コンセプト、ならびにこれに関連する研究領域の名称である [14], [15].

ビッグ・データの時代を背景として提唱されている LA において主流とされているアプローチは、現在のところ、従来の教育工学/学習科学の方法論に基づいたものである。たとえばそれは、伝統的な社会ネットワーク分析であり、ディスコース分析である。このようなアプローチは本来は小規模データを分析の対象としてきた、旧タイプの方法論といえる。だがこうした旧タイプのアプローチでは、MOOC のようなこれまでに想定されていなかった学習環境の分析や、認知神経科学などの学習に関連する諸科学の知見を導入することはできない。本稿では、従来型のアプローチに依拠ばかりではなく、関連諸領域のアプローチを取り込むことによって「データ・サイエンスとしての LA」という、新たな方向性が見出されることについて議論する。

本稿の構成は次の通りである。第 2 章において現在の LA から欠落している、学習に対する Microscopic と Macroscopic な視点について言及する。第 3 章では、ある学習現象を説明する (かもしれない) 数理モデルを例として取り上げ、データ解析と数理モデルによる分析によって従来の方法論によるアプローチと何がどう違ってくるのかということについて触れる。最後の第 4 章では (これまでのアプローチにはとらわれないという意味での) 新しい LA のフレーム・ワークについて述べる。

¹ 広島大学大学院社会科学部研究科
1-2-1 Kagamiyama, Higashihiroshima, Hiroshima, 739-8525 Japan

² 福井県立大学学術教養センター
4-1-1 Matsuoka-Kenjojimama, Eiheiji-cho, Fukui, 910-1195 Japan

³ 名古屋大学大学院情報科学研究科
A4-2(780) Furo-cho, Chikkusa-ku, Nagoya, 464-8601 Japan

⁴ 広島大学情報メディア教育研究センター
1-4-2 Kagamiyama, Higashihiroshima, Hiroshima, 739-8511 Japan

⁵ 九州大学情報基盤研究開発センター
6-10-1 Hakozaki, Higashi-ku, Fukuoka, 812-8581 Japan

a) ystake@hiroshima-u.ac.jp

b) yamakawa@fpu.ac.jp

c) nakamura@nagoya-u.jp

d) sumi@riise.hiroshima-u.ac.jp

e) tagawat@cc.kyushu-u.ac.jp

f) jin@cc.kyushu-u.ac.jp

^{*1} 2011 年 12 月 1-2 日, 第 6 回研究会を福井市 JR 福井駅前 AOSSA にて開催した。

^{*2} CLE 研究会では、第 6 回研究会 (2011 年), 第 8 回研究会 (2012 年), FIT2013 での企画セッション (2013 年) などで LA を取り上げてきた。日本教育工学会 (JSET) では第 29 回全国大会 (2013 年) において LA に関する課題研究のセッションが企画された。教育システム情報学会でも 2014 年の第 39 回全国大会において LA に関する企画セッションが設けられている (http://www.jsise.org/taikai/2014/pls_session.html)。

2. Microscopic と Macroscopic な視点

[15]でも指摘したように、現在の LA [2]は文化人類学的な状況的学習論を理論基盤としているため、学習の「進化論／生物学的／物理学的制約と機能」に対する Microscopic な視点と、学習の Sociodynamics に関する Macroscopic な視点が欠落している。

2.1 学習に対する Microscopic な分析視点の重要性

[15]では学部専門課程以上のレベルの数学教育を例にとり、現在の教育学／学習科学研究ではこのレベルでの問題に十分に解を出すことができていないことを指摘した^{*3}。それ以外にもたとえば次のような問題に対して、状況的学習論とこれに依る LA は解を提示することができない。

- (1) 学習者が学習する内容は、長期記憶として脳の側頭葉などにコード化され再活用されるようになってこそ、意味がある。このメカニズムと学習共同体の環境が学習者に刺激として与える影響との関係や「コミュニケーションを介した学習の効果」の関係などについては、ほとんど議論されていない。
- (2) 人間の認知過程に対しては、内部自由度を持った要素間の相互作用が系の状態に不安定性をもたらし、そうした不安定さがカオス遍歴などのダイナミクスを生成することによって、認知過程での「シンボル化」といった現象がもたらされるのではないかという見解がある [6]。こうした脳機能に対する力学系的観点から見たメカニズムと状況的学習論が主張する現象論的解釈は、ほぼ接合できていない。
- (3) 教材や学習コンテンツについて従来の媒体とデジタルなものとの比較すると、後者による学習は必ずしも優位であるとは主張できない（場合によっては余計な認知的負荷を脳に与えることになり、マイナスの効果を与えてしまう）とする研究結果が少なからず発表されている^{*4}。デジタル以前のコンテンツが持つ具体性・身体性、あるいは人間の認識機能の発達が進化論的には何ら必然的なものではなく、ありあわせの機能を使った結果でしかないという、生物学上の制約がこの結果に関係している^{*5}。こうした研究結果に対する反

論は、おそらく現在の LA のアプローチからはなかなか出すことは難しいだろう^{*6}。

われわれは状況論的学習論に依る LA をまったく否定しようというのではない。そうではなくここでわれわれが主張したいのは、教育学／学習科学が自らの方法論に固執することなく、関連する諸科学の成果を積極的に取り入れ、理論の整合性を精緻化することの重要性である。この方向においてこそ「学習」に対するわれわれの理解は深まる。

データ収集と解析の技術が飛躍的に進歩した現在、認知神経科学、分子生物学や力学系的な生命科学も含めた理論生物学、人工生命科学等、これまではまったく異なる領域であると思われていた関連諸科学の知見は、「学習」に関するデータ解析という共通項を媒介として結び合わせることが可能となる。LA が革新的な視野の拡大と学習に対する本質的な理解のもとでの高等教育の改革（要するに、ほんとうに効果のある改革）を狙おうとするのであれば、教室環境を対象とした旧タイプのデータ分析だけではなく、こうした Microscopic な視点からのデータ分析も必要であるとわれわれは考えている。

2.2 Macroscopic な分析視点の重要性

この節では学習に対する Macroscopic な分析視点の重要性について、[14], [15]では論じなかった、*ceteris paribus* な分析手法の限界について議論する。

状況論的学習論に基づく LA が対象とするのは、ほとんどの場合、教室という「閉じた学習環境」である。この点は、「さまざまな変数がダイナミックに変動するときどのような相互作用が生み出されるかといった問題」を重視し「学習研究から得られた知見はそうした生態学的妥当性の高い環境で機能しなければ意味がない」[8]とするデザイン原則アプローチ研究でも同様である。

こうした方法論はともすれば、学習空間をある特定の限定された孤立系として扱ってしまうことになりがちである（つまり、そこで起きたことだけを分析や観察の対象としてしまい、ほかの要素を考慮の対象外に置いてしまう）。特定の「学習環境」や「教室」だけを分析の対象とする現在の教育学／学習科学研究からは、一般的な妥当性ある知見を導き出すことはおそらくほとんどできない。なぜならそれは、経済学における部分均衡分析のみをやっているようなものだからである。部分均衡分析ではある特定の市場だけを考察の対象とし、その他の市場や社会システムは完全に考察の対象から外す。したがってそこから得られる結論は *ceteris paribus* (with other things the same) というきわめて限定的な条件付のものでしかない^{*7}。

^{*3} 抽象度の高い数学理論を理解することを認知科学的に分析した文献に [7] がある。ただしこの文献は、認知科学の観点から見て「数学を理解するとはどういうことか」を論じたものであり、高等教育レベルの数学を学び理解する上での十分条件について議論したものではない。

^{*4} たとえば [5] など。[3] の参考文献リストは有益である。

^{*5} たとえば、人類が文字を使用するようになったのはたかだか数千年前ではない [12]。人間の思考と文字は不可分の関係にあるものではなく、文字を持たない文化の方が圧倒的に多い。文字を「読み理解する」という行為は生物としての人間にとって「自然な」行為ではない。そうした「不自然な行為」に対してさらにデジタルなコンテンツを使用することは、人間の認知機能のある面に対して余分な負荷をかけているのかもしれない。

^{*6} [1] は、この点に関する痛烈な批判である。

^{*7} これに対して「社会全体」を一挙に分析の対象とする分析を「一般均衡分析」と呼ぶ。もちろん、位相空間論に基づき高度に抽象化された一般均衡分析ではあるが、非常に多くの問題点がある。ここでは論じない。

ceteris paribus で限定的な方法論・分析手法がなぜ問題であるかと言えば、たとえば MOOC のような大規模な学習環境を考えるとよい。従来のきわめて小規模で閉じた系を対象として得られてきた知見が MOOC に果たして適用可能なものであるかどうか、明確なことはほとんど何も言えない*8。

本来学習環境は、幾重にも重層的／階層的に折り重なった社会システムのサブシステムであり、そこで発生する現象はほかのサブシステムに影響を与え、かつほかのサブシステムから影響を受けているはずである。こうした相互作用のフィードバック構造を念頭に置いた分析視点、すなわち学習環境を Macroscopic な視点からとらえ、「そこでの」相互作用、「そこからの」相互作用、そして「そこへの」相互作用を分析の対象とする Sociodynamics な枠組みを、現在の LA は持っていない。

3. ある数理モデルによる学習現象の理解

Microscopic あるいは Macroscopic，いずれのアプローチにおいても、ただ単に膨大で多種多様なデータ群を集めさえすれば「これまで見ていなかった学習現象の本質とメカニズム」がただちに分かるわけではない。ここで必要になるのが、データ群を処理して見えてくる現象の本質とそうした現象を生成するメカニズムを説明するための「理論」である。ただし、[15]でも指摘したように、従来の自然言語による会話分析やアクター・ネットワーク分析の枠組みは、Microscopic あるいは Macroscopic な「ビッグ・データ」の解析に対しては絶望的に不利であろう。そこでわれわれが注目しているのが、近年急速に発展してきている Socio-Econophysics 的な数理モデルによるアプローチである。学習現象が一連の（しかも膨大な）データ群によって記述されるのであれば、そうしたデータ群の背後にある現象のメカニズムを数値データと親和性のある数理モデルによって説明できる可能性も出てくる。以下、こうしたアプローチの例として MOOC の顕著なドロップ・アウト率をとりあげる。この現象を非常に単純な出生死滅モデルによって説明できる（かもしれない）ことを示す。

いま MOOC のような、次々に学習コースが生成される学習環境を考える。この環境では学習者に課題が提示され、それらはインデックス $i = 1, 2, \dots, n, \dots$ によって識別されるものとする*9。このインデックスは課題内容の

難易度にしたがって順序付けられているものとし、各学習課題に関する知識の包含関係は $I_1 \subset I_2 \subset \dots \subset I_n \subset \dots$ であるものとする。ここで I_i は第 i 番目の学習課題（以下、課題 i ）で修得することのできる知識や技能の集合を示す。課題 i に取り組んでいる学習者の数を k_i とし、これを学習課題 i のサイズと呼ぶ。課題 i に取り組んだ学習者が次の課題 $i+1$ へと進む確率は 1 ではなく、一部はそれ以前の課題 $i-1$ に戻らなければならないとしよう。この想定のもとで時間間隔 Δt において課題 i のサイズ k_i がサイズ k_j へと推移する推移確率 p_{k_i, k_j} は、次の式によって定められているものとする。

$$p_{k_i, k_j} = \begin{cases} \omega_+(k_i)\Delta t + O(\Delta t) & \text{if } k_j = k_i + 1 \\ \omega_-(k_i)\Delta t + O(\Delta t) & \text{if } k_j = k_i - 1 \\ O(\Delta t) & \text{otherwise} \end{cases}$$

さらに境界条件として Δt において $k_i = 1$ であるような課題が確率 $p \ll 1$ で存在するものとする。時刻 $t \geq 0$ における k_i の学習者全体に占める割合を $0 \leq p(k_i, t) \leq 1$ とすれば、 $p(k_i, t)$ に関する次のマスター方程式を得る。

$$\frac{\partial}{\partial t} p(k_i, t) = \omega_+(k-1)p(k-1, t) + \omega_-(k+1)p(k+1, t) - [\omega_+(k) + \omega_-(k)]p(k, t) + p\delta_{k,1}$$

ここで $\delta_{k,1}$ はクロネッカーのデルタである。

このマスター方程式に対して、限界推移 ω_+ と ω_- を特定化すれば詳細つりあい条件のもとで以下の結論を得ることができる [4], [16].

- (1) 限界推移がサイズ k の非線形関数であるとき、課題に取り組む学習者の数の分布はあるべき乗則 (Power Law) にしたがう。すなわち課題のほとんどを修了することができる学習者はきわめて少数でしかない。圧倒的多数の学習者は学習を貫徹することができていない。限界推移とサイズ k が線形の関係にあるときにはこの結論は妥当しない。このことは、MOOC のようなきわめて大規模な学習環境における学習離脱率が顕著に高い必然的な理由を説明しているのかもしれない。
- (2) 限界推移に対して k が線形と非線形の影響を与えると、非線形の効果が線形の効果に対して優位であれば、課題別学習者の数の分布はあるべき乗則にしたがう。すなわち課題のほとんどを修了することができる学習者はきわめて少数でしかない。圧倒的多数の学習者は学習を貫徹することができていない。このことは、MOOC のようなきわめて大規模な学習環境における学習離脱率が顕著に高い必然的な理由を説明しているのかもしれない。
- (3) 以上のことから、次の課題に進むか否かを決める限界

することが可能なのであれば、その背後には共通するメカニズムが隠されている可能性もある。そうした探求は数理モデルに基づいてこそ明確になる。それがここで述べたいことである。

*8 MOOC は、人間が安定的な関係を築くことのできる認知的上限数である Dunbar Number を越えた学習者が集めた集合体である。このような学習環境がどのような効果をもたらすのか、そのような環境はそもそも「安定的」なものであるのかといった問題に対して、ほとんどまだ何も分かっていない。もしかすると現在のところ MOOC は目新しいだけで、本質的には小規模で閉じた学習環境の単なる集まりでしかないのかもしれない。

*9 以下ここでモデルは [4] に負う。もともとこのモデルは都市の人口分布に関する法則性を説明するために提案されたものである。ここではそのモデルをほぼそのまま読み替えているにすぎない。しかしながら、もしも同じ方程式によって異なる現象を説明

推移がその課題に取り組んでいる学習者数の非線形関数であるとき、すべての課題を学習者が修了するのは必然的に困難なものとなると言える。

もちろん、いまここで示したモデルは非常に単純なものでしかなく、また、実際の学習者による課題の推移がこのマスター方程式にしたがっているものであるかについては検討の余地が残っている。しかもどのような非線形の効果と考えられるかについても、ここではまったく特定化していない。われわれがここで主張したいのは次の点である。

- (1) (大規模データによる検証を前提として) 理論モデルによって学習現象を理解することは、学習のメカニズムを把握する上で有益である。たとえば、上のマスター方程式モデルが妥当するものだとすれば、モデルをより具体的に記述することで、学習から脱落するメカニズムを(自然言語によるメタファーではなく)厳密な論理において(つまりどういう要素がどう効いてくるのか、それは何故かという問題に対して厳密に)とらえることが可能となる。
- (2) 学習のメカニズムを数理モデルの論理によってとらえることができるのであれば、そのメカニズムによってもたらされる、正負を含めた学習への影響を把握することができ、ひいてはそれらを起因とする諸問題を解決するためのシステム開発、授業環境のデザイン設計などに貢献することができる。

4. LA に関する研究のフレーム・ワーク

以上述べてきたように、LA が真にデータ・サイエンスを志向し、「従来の方法論で見えていなかった」学習現象の本質とメカニズムの解明を求めるとすれば、これまでは対象外とされていた Microscopic, Macroscopic 両面における膨大なデータの収集とそこから見えてくる現象を数理的なモデルによって説明するというアプローチが必要不可欠なものであるとわれわれは考えている。つまりわれわれが LA に対して考えているのは、社会現象の解明に対して統計物理学的手法によって迫ろうとすでに先行している、次のような Socio-Econphysics 的アプローチである [10]^{*10}。

- (1) 高頻度かつ大量に実際の学習に関連する多様なデータを科学的視点に基づいて解析し、そこから学習現象の本質に関する仮説を見出す。
- (2) その仮説を説明する理論モデルを構築する。
- (3) 理論モデルの妥当性は実際のデータによって検証する。
- (4) 説明力を高めた理論モデルによって現象を予測する。
- (5) 妥当性のテストにパスした理論モデルあるいは理論に

基づいて、より効果的な学習環境の構築、学習支援システムの開発などを行なう。

確かにこうしたアプローチは、従来の教育工学/学習科学研究とは異質のものである。しかしながら、Watts も言っているように「アレグザンダー・ホープが人間の適切な研究課題は天上ではなくわれわれのなかにあると説いてから三百年後、われわれはようやく自分たちの望遠鏡を手に入れた」[11] のであり、学習の本質に迫るアプローチをこれまでの手法に限定する必然性はどこにもない。自分たちが慣れ親しんできた手法や方法論に固執するだけでは、おそらくビッグ・データを活用し、学習現象に対する理解を深めることはできないだろう。現在の LA の限界を越えるために必要なのは、これまでの方法論にとらわれない柔軟なアプローチである。そうした LA の新しい可能性を追求する価値は十分にある。

参考文献

- [1] 新井紀子：ほんとうにいいの？デジタル教科書，岩波書店(2012)。
- [2] Buckingham Shum, S. and Ferguson, R.: Social Learning Analytics, *Journal of Educational Technology & Society*, Vol. 15, No. 3, pp. 3–26 (2012)。
- [3] N.G. カー：ネット・バカ：インターネットがわたしたちの脳にしていること，青土社(2010)。
- [4] 藤原義久：知っておくとよい数学の要素 2, 50 のキーワードで読み解く経済学教室—社会経済物理学とは何か？，東京図書，pp. 208–215 (2011)。
- [5] Jabr, F.: Why the Brain Prefers Paper, Vol. 309, November, pp. 48–53 (online), DOI: 10.1038/scientificamerican1113-48 (2013)。
- [6] 金子邦彦：生命とは何か：複雑系生命科学へ，東京大学出版会，2 edition (2014)。
- [7] Lakoff, G. and Núñez, R. E.: *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics Into Being*, Basic Books (2000)。
- [8] 大島 純：学習科学の展開，教授・学習過程論：学習科学の展開，放送大学教育振興会，pp. 200–213 (2006)。
- [9] Sumiya, T., Inoue, H., Tagawa, T., Yamakawa, O. and Yasutake, K.: Redesign of Course Visualization Platform VisP, *Proceedings of IADIS International Conference e-Society 2010*, pp. 557–559 (2010)。
- [10] 高安美佐子，黒田敏正：ソーシャルメディアの経済物理学—ウェブから読み解く人間行動，日本評論社(2012)。
- [11] ダンカン・ワッツ：偶然の科学，早川書房(2012)。
- [12] Wolf, M.: *Proust and the Squid: the Story and Science of the Reading Brain*, Harper (2007)。
- [13] 山川 修，安武公一，多川孝央，隅谷孝洋，井上 仁：学習科学における新しい研究アプローチの提案，日本教育工学学会第 28 回全国大会講演論文集，pp. 477–478 (2012)。
- [14] 安武公一，中村泰之，山川 修：Social Learning Analytics ver.2 の提案，情報処理学会研究報告(第 8 回 CLE 研究会)，2012-CLE-8 (2012)。
- [15] 安武公一，中村泰之，山川 修：Learning Analytics に関する教育工学/学習科学研究の Missing Link，日本教育工学学会第 29 回全国大会講演論文集，pp. 95–98 (2013)。
- [16] Zanette, D. H. and Manrubia, S. C.: Role of Intermittency in Urban Development: A Model of Large-Scale City Formation, *Phys. Rev. Lett.*, Vol. 79, pp. 523–526

^{*10} 学習環境が生成するデータ解析を主な対象とはしているが、数理モデルによる現象の把握が重要であることについては、すでに [13] で触れている。また、異なる学習環境やシステムから収集される多様なデータ・セットを統一的に解析し、そのデータ解析の結果を実践の場でスマートに提示するシステムのアイデアについては、[9] がある。

(online), DOI: 10.1103/PhysRevLett.79.523 (1997).