

最先端の Learning Analytics 研究を目指して — 一周遅れのわが国の LA 研究を打破するために —

Toward the Advanced Learning Analytics Research - Breaking the Mold and Trying Something New -

安武公一^{*1}, 中村泰之^{*2}, 山地一禎^{*3}, 古川雅子^{*3}, 梶田将司^{*4}, 山川修^{*5}, 多川孝央^{*6}
Koichi YASUTAKE^{*1}, Yasuyuki NAKAMURA^{*2}, Kazutsuna YAMAJI^{*3}, Masako FURUKAWA^{*3},
Shoji KAJITA^{*4}, Osamu YAMAKAWA^{*5}, Takahiro TAGAWA^{*6}

^{*1} 広島大学社会科学部

^{*1} Graduate School of Social Sciences, Hiroshima University

^{*2} 名古屋大学情報科学研究科

^{*2} Graduate School of Information Science, Nagoya University

^{*3} 国立情報学研究所

^{*3} National Institute of Informatics

^{*4} 京都大学学術情報メディアセンター

^{*4} Academic Center for Computing and Multimedia Studies, Kyoto University

^{*5} 福井県立大学学術教養センター

^{*5} Center for Arts and Sciences, Fukui Prefectural University

^{*6} 九州大学情報基盤研究開発センター

^{*6} Research Institute for Information Technology, Kyushu University

Email: ^{*1}ystake@hiroshima-u.ac.jp, ^{*2}nakamura@nagoya-u.jp, ^{*3}{yamaji,furukawa}@nii.ac.jp,
^{*4}kajita.shoji.5z@kyoto-u.ac.jp, ^{*5}yamakawa@fpu.ac.jp, ^{*6}tagawat@cc.kyushu-u.ac.jp

あらまし：現在のところ日本の「教育ビッグデータ」研究は、先行する世界の Learning Analytics (LA) 研究の現在から見て、閉塞的で周回遅れなものとなっている。ビッグデータ分析でもない。本研究では、世界の最新の研究動向（われわれの一部は LAK2016, L@S16 に参加した）を踏まえ、わが国の LA 研究において解決すべき問題点を指摘する。その上で、関連するシステム開発研究、環境整備、さらには LA の可能性を広げ真に最先端のデータ分析科学となるために必要な「進むべき方向性」について提案する。

キーワード：Learning Analytics, データサイエンス, ビッグデータ, 数理科学

1. はじめに

今年 (2016 年) 4 月, 学習分析 (Learning Analytics: LA) 分野における代表的な国際会議のひとつである Learning Analytics and Knowledge Conference (LAK) が UK のエディンバラ大学で開催された。さらに今年は学習の大規模データ分析に関する国際会議である Learning at Scale (L@S) も同時期同大学において開催された。発表者ら⁽¹⁾はこれらの国際会議に参加し、最先端の学習分析研究の世界的動向について知見を得てきた。そこから見えてきたのは、国際的な LA 研究は常に進化しており、前回まではカッティングエッジだった話題が今回はすでに実践的で成熟したデータ解析へとダイナミックに展開しているという、当たり前の事実である。すでに海外では LMS 上に学習履歴データ分析を盛んに集めその分析だけを行なう段階から、LA の未知なる可能性に挑戦する段階に入っているとも言える。

こうした世界の現状に比べわが国の LA の現状は、大きく後れを取っているといわざるを得ない。もちろん先端的な研究も進められている。しかしその横展開や層の厚さは国際的なレベルからは程遠い。

そこで本研究では、わが国の LA 研究が現段階かかえている課題について議論するとともに、今後の

研究の展開において必要になるとと思われる、新しい LA 研究のフレームワーク、さらには、関連するシステム開発と基盤情報整備について言及する。なお本稿が対象とするのは高等教育である。初等・中等教育の LA 研究は対象としない。

2. わが国の LA 研究の課題あるいは問題点

わが国の LA 研究がかかえる課題（あるいは問題点）として指摘できるのは、分析対象も分析手法も狭い範囲に閉じてしまっているという現状である。

まず第 1 に、LA 研究の多くはクラス単位、授業単位を対象を限定して行なわれているのが実情である。しかもそこで分析対象として収集される学習データは、主に（授業単位の）LMS 上の履歴データ、あるいはせいぜい対象を広げてデジタル端末（デジタル化された教科書など）へのアクセス記録が中心となっている。

こうした手法の最大の問題点は、特殊性と一般性を厳密に区分して抽出することができないことである。ある特定の環境、条件下で実施された授業内容の、特定の学習デバイスを通して得たデータだけを分析することで、どれだけ「学習の本質的なメカニズム」や「有効な学習プロセスを導くための十分条

件」を明らかにすることが可能なのか、疑問の余地がある（多くは事後的に収集したデータなので必要条件はある程度抽出できるかもしれないが）。

第2に、わが国ではLAとIR (Institutional Research: 大学全体などを対象とした機関データの分析) は別物として扱うことが多い。しかしながら、高等教育環境が様々なレイヤーで重層的に構成されている以上、このような研究の縦割りのな区分を（分析対象としてではなく）最初から頑なに守ることにどれだけの意味があるのか、再考する必要があるだろう（LAKでもL@SでもLAのみならずIRに関するセッションが当たり前のように設けられている）。

そして第3にわが国のLA研究の問題点として指摘されるのは、分析手法がこれまでの教育学・学習科学研究に沿ったものがほとんどであることである。これについては次のセクションで説明する。

3. 新しい研究のフレームワークとデータ情報基盤整備の必要性

教育環境は幾層にもレイヤーが重なった重層的な構造のサブシステムである。それは神経科学が対象とするミクロのレベルから Complex Networks Science が対象とするマクロのレベルまで、各層が複雑に相互作用するシステムである。残念ながらいまの（特にわが国の）LAは、学習環境をこうした複雑系のひとつのレイヤーとしてとらえる、ほんとうの意味での「データサイエンス」としてのフレームワークは持っていない。だがこれまでの教育学・学習科学研究の枠組みを越え他分野の領域に目を向けたとき、LAはその可能性を広げることができる⁽²⁾⁽³⁾。

たとえば、認知神経科学 (cognitive neuroscience) レベルまで降りていくなれば、近年の研究によって抽象的な数学的概念の取得（の困難さ）には、生物学的・進化論的な制約を受けた脳の機能と構造が関係していることが分かっている⁽⁴⁾。あるいは同様に cognitive neuroscience の領域で、「デジタルネイティブ」と呼ばれる世代であっても、記憶や学習にはデジタル媒体よりも紙媒体の方が有効であるという研究もある⁽⁵⁾。一方視野を教室を越えて社会的レベルにまで広げるならば、膨大な breadcrumb データの解析によって、集団の生産性と人々のコミュニケーションの在り方（その内容は無関係！）の間には密接な関係があることも分かっている⁽⁶⁾。今後はこれまで対象外とされていたこうした領域を含めた研究もLAの射程に入ってくるだろう（事実、センサデータを使った生態学的研究に関するセッションが今年のLAKには設けられていた）。

ただし、現在のLA研究の方法論に依拠する限り、アプローチの視野をただ広げただけではデータ間の相関性が分かるだけで、データを生成した学習のメカニズムまでは知ることはできない（他領域の教育科目を担当する教員にとって価値があるのは学習のメカニズムに関する情報、特に効果的な学習を実現

させるための十分条件である。必要条件ではない）。これは統計分析に付随する本質的な限界である⁽⁷⁾。

データ間の因果関係を解明する上で有効となるひとつの戦略は「観測によるデータの収集-仮説（モデル）の提唱-仮説の検証」という数理学のアプローチである。このアプローチは極少数の例外を除いてこれまでの教育学・学習科学研究では、ほとんど行われてこなかった。しかしながら、重層的なレイヤーによって構成された複雑系のサブシステムとして学習環境をとらえ、この環境が生成するデータを解析し、その意味をとらえようとしようとするならば、これまでの方法論はほぼ無力である。認知神経科学が対象とする「ミクロ・レベル」と社会的ネットワークといった「マクロ・レベル」の相互関係において高等教育環境を分析しようすれば、それに適しているのはたとえば統計物理学のような数理的なアプローチである（実際、Complex Networks Science が採用しているのは統計物理学的アプローチである）。

最後に、上で述べた新しいフレームワークでのLAとその成果を実践的に活用するために必要となる、大学間でのデータ連携のためのシステム環境ならびに基盤整備について触れる。学習現象の一般的な特性を解明しようとするならば、単に教室単位、授業単位で閉じた研究は不十分である。特に統計データから何らかの一般的法則を得ようとするれば膨大なデータの収集と処理が不可欠である（それこそが「教育ビッグデータ」である）。しかしそうしたデータを収集し共有する仕組みも得られたデータを共同で解析する研究も、わが国ではまったくと言っていいほど行われていない。もちろん学習データの収集と解析には「個人情報」という問題が付いてまわる。しかしながら、この問題が解決しないからと言って入口で足踏みをしていたのでは、わが国のLAはいつまで経っても現状を打破することはできないだろう（停滞するだけである）。こうした、個別の大学の枠を越えたデータ共有ならびに解析を可能とするシステムの開発とそのための基盤整備も、周回遅れとなってしまっているわが国のLA研究、ひいては高等教育改革のためには、是非必要である。

参考文献

- (1) 古川雅子ほか：“第6回 Learning Analytics and Knowledge (LAK16) 参加報告”，mimeo (2016)
- (2) MIT Online Education Initiative: “Online Education: A Catalyst for Higher Education Reforms” (2016)
- (3) 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，IPSIJ-CLE8 (2012)
- (4) S・ドゥアンヌ：“数覚とは何か？”，早川書房 (2010)
- (5) Ferris Jabr: “Why Brain Prefers Paper”，*Scientific American*, vol. 309, pp. 48 - 53 (2013)
- (6) Alex Pentland: “Social Physics”，Penguin Press (2014)
- (7) 新井紀子：“ほんとうにいいの？デジタル教科書”，岩波書店 (2012)