

# Eduinformatics をもとにした IR の学生データ比較解析 における新クライテリアの提案

中田康夫（神戸常盤大学） 村上勝彦（東京大学） 小崎恭弘（大阪教育大学）  
桐村豪文（弘前大学） 杉浦あおい（神戸市立西神戸医療センター）  
伴仲謙欣（神戸常盤大学） 高松邦彦（神戸常盤大学）

## 1. 緒言

われわれは、ここ数年来、“Education”と“Informatics”を結合させた Eduinformatics という領域を提唱している[1]。これは、Bioinformatics という分野が、生物学（Biology）と情報学（Informatics）からなる学際・学融領域であるのと似ている。われわれは、文献[1]のなかで、Eduinformatics は、教育上の課題を解決するために、情報学を従来以上に活用したり、課題解決のための新たな情報学の方法を構築することが重要だということを報告してきた。

近年、各大学では、学生データを取り扱うために、IR（Institutional research）と名がつく組織の新設が相次いで行われている。その理由は、文部科学省が2008年と2012年にIRに関する重要なレポートを報告したからであり[2][3]、現にこれらのレポートが報告された後、わが国の大学におけるIR室の設置数は急速に増加している[4]。さらに、2016年より文部科学省は、わが国の大学でのIR室設置の必要性をこれまで以上に強調し、とくに私立大学においては、私立大学等改革総合支援事業による影響を大きく受けている[4]。

DSIR（Data Science and Institutional Research）は唯一わが国で開催されているIRをテーマとする国際学会である。このDSIRにおいて、2016年よりわれわれは、教職員の協働にとどまらず、現役の学生に加え、大学サポーターとしての卒業生を加えた4者の混成チームというuniversity developmentの新しい形で、高等教育にinformaticsを応用した研究をいくつか報告している[5]–[11]。そのなかで、学生データの活用において、これまで取り扱ってきたデータを用いた解析だけでは発見できないものがいくつもあることが、これまでの本学におけるIR活動のなかで浮かび上がってきた。

そこで本研究では、大学内でのIRにおいて、学生データを使用するための新しいクライテリアを提案する。この新クライテリアには2つの水準があり、1つを「プライマリデータ」と呼び、もう1つを「セカンダリデータ」と呼ぶ。われわれは、プライマリデータを線形結合データではないもの、セカンダリデータをプライマリデータの線形結合として定義する。従来のわが国のIRにおいては、その大部分において、ここで提案する新クライテリアの2つ目の水準であるセカンダリデータを主に扱ってきたように見受けられるが、本稿では1つ目の水準を新たな水準として用いたIR解析の重要性・必要性について検討する。

## 2. 新しいクライテリア

UNESCO (The United Nations Educational, Scientific and Cultural Organization) は、2012年の Learning Analytics に関するレポートにおいて、「学修解析の三層」を提唱している[12] (図1)。以下の引用は、その解析様式の三層の定義である[9]。

マクロレベルの解析は、組織間の解析を可能にする。それは、たとえば、組織実践の「成熟度」の調査[13]や、学生の生涯にわたる標準化された評価データへの州全体のデータアクセスの改善[14]などで実現されている。マクロレベルの解析は、今後よりリアルタイムに解析されるようになり、そしてメソ/ミクロレベルからより詳細にかつより多くのデータを取り込み、非教育セクターで開発されたベンチマークとデータ統合する方法から恩恵を受ける可能性がある。

メソレベルの解析は、機関レベルで機能する。教育機関が、ビジネスインテリジェンス (BI) の恩恵をすでに受けているセクターと共通のビジネスプロセスを共有していれば、会社のウェアハウスにデータサイロ (企業内の1つの部署や人の管理下において、ほかの部署からは隔絶されているデータ) を統合し、ワークフローを最適化し、ダッシュボードを生成するための便利なツールを備えた新しい BI 市場セクターとみなすことができる。例としては、非構造化データのマイニング、「顧客離れ」および将来の市場の予測などである。機関レベルの「アカデミック解析」[15]を構築する努力を部分的に動機づけるビジネスプロセスを最適化することは、BIの必須事項である。

マイクロレベルの解析は、個々の学修者 (さらにはグループ) のプロセスレベルのデータの追跡と解釈をサポートする。このデータは、理想的には可能な限り迅速に最高レベルの詳細なデータを提供できるため、学修者自身ならびに教師にとって最も重要なものである。プラットフォームによって異なるが、このデータはクリックごとのオンライン・アクティビティ、地理情報・位置情報、図書館の貸し出し記録、購入などの履歴、ソーシャルネットワークなどの対人関係データを開示できるため、最も個人的なデータを取り扱うことになる。研究者は、ゲーミフィケーション、自動採点、教育データマイニング、コンピューターでサポートされる協働学修、レコメンダーシステム、インテリジェント・チューター・システム/アダプティブ・ハイパーメディア、情報の可視化、計算言語学と論証、ソーシャルネットワーク解析などの分野の手法を採用している。

この定義にもとづくと、われわれはマクロレベルの解析において、次に述べる2つのタイプのデータを取り扱うことの必要性和重要性を認識することができる。1つはプライマリデータであり、もう1つはセカンダリデータである。

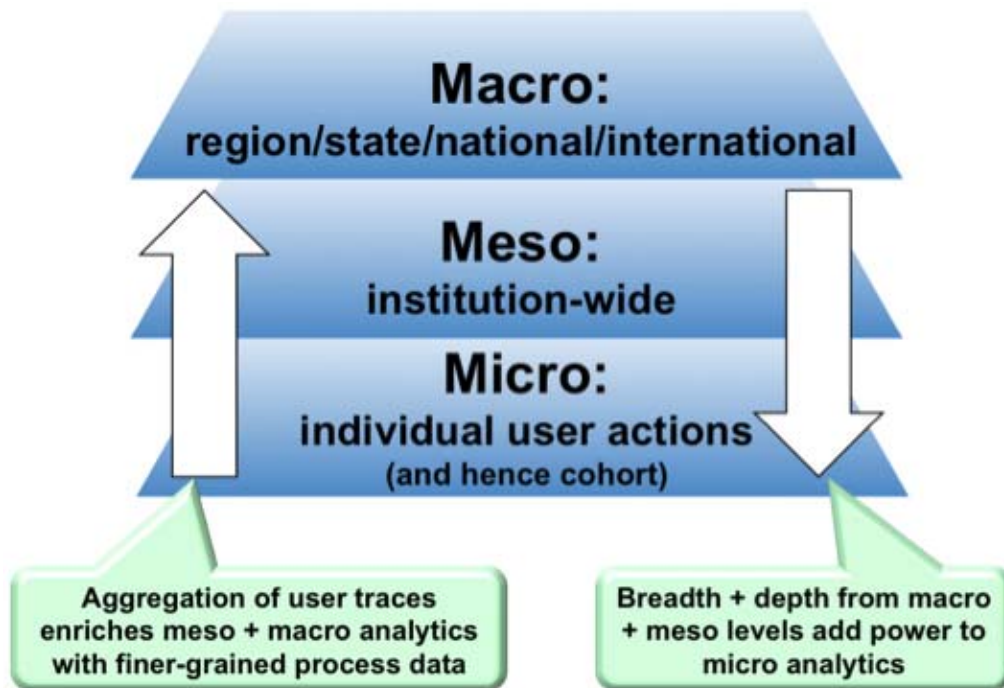


図1 学修解析の三層（from UNESCO IITE, Learning Analytics, 2012）

われわれはここで、次のような線形関数の組み合わせを定義する。関数  $f_1, f_2, \dots, f_n$  およびスカラー  $a_1, a_2, \dots, a_n$  の場合、 $f_1, f_2, \dots, f_n$  の線形結合を次のように定義する。

$$a_1 f_1 + a_2 f_2 + \dots + a_n f_n$$

したがって、線形結合は、関数  $f$  と重み  $a_1, a_2, \dots, a_n$  の和となる。

さらに、線形の写像または関数を次のように定義する。

関数または写像  $x$  および  $y$  の場合、線形写像または関数は次の2つの性質を満たす。

$$\text{加法性: } f(x + y) = f(x) + f(y)$$

$$\text{斉次性: すべての } a \text{ に対し、 } f(ax) = a f(x)$$

明らかに、写像や関数が線形の場合、それらを線形結合しても、また線形の性質があることに注意する。

ここでわれわれは、プライマリデータを線形結合していないデータ、セカンダリデータをプライマリデータの線形結合と定義する。

たとえば、マクロレベルでは、プライマリデータは試験の各問題に対する正解と不正解、または講義における学生の出席と欠席のようなデータとなる。マクロレベルでは、セカンダリデータは、試験の合計点や、講義受講者の合計出席者数などとなる。

メソレベルでは、セカンダリデータは講義の学生記録および成績平均点（GPA）、または大学の年間順位などである。GPA と大学の年間順位は、線形関数の組み合わせとなっており、セカンダリデータである学生の成績（Grade）を使用して算出される。したがって、セカンダリデータは、プライマリデータの線形結合だけではなく、セカンダリデータ自身の線形結合になっている点に注意が必要である。

### 3. まとめと将来展望

従来の大学 IR に関する報告を俯瞰すると、研究者は、大学のマイクロレベルのプライマリデータとセカンダリデータのみで解析を行い、IR スタッフはメソレベルのセカンダリデータでのみ解析をしていることがほとんどである。なぜなら、通常、マイクロレベルのプライマリデータは、大学教員個人によってのみ取得および保管される一方、メソレベルのセカンダリデータは、大学の IR 室や教務課といった組織や部署が保管しているからである。このため、教員はセカンダリデータセット全体にアクセスできず、逆に IR 室や教務課はプライマリデータにアクセスすることができない状況になっている。このことから、大学 IR といっても、それぞれの解析者が取り扱っているデータのクライテリアが異なっているため、IR の本来の目的・目標が十全に達成しえない状況になっていることが推察される。この課題を解決するために、IR 関係者はどのようにすればよいのであろうか。

1 つ目として、大学のシステムとして、誰がどのようにどのタイミングでプライマリデータとセカンダリデータを収集・格納するのかについて検討する必要がある。

近年、わが国の大学では、教員は従来のように単に最終試験による成績評価だけでなく、講義に参加する学生の「主体的に学修に取り組む態度」に関する評価などを含めて、多面的に評価することが強く求められるようになってきている。しかし、このような個々の学生に関する詳細なデータはプライマリデータであり、現在においてはマイクロレベルの主要データではないことに注意が必要である。大学に保管される学生の成績データはセカンダリデータであり、試験の点数や態度に関する点数などを線形関数の組み合わせとして計算されたものが主要データとして保管されている。

さらに、試験の点数についても考慮する必要がある。通常、試験はいくつかの問の組み合わせで構成されている。ほとんどの教員は、各問に対する学生の正解または不正解は記録・保管せず、合計点のみが最終的に記録・保管される。この場合、各問に対する学生の正解または不正解がプライマリデータとなる。つまり、試験の成績は、マイクロレベルのセカンダリデータとなっている。容易に想像できるように、ほとんどの教員はそのようなプライマリデータを記録・保管しない。このようなプライマリデータを記録・保管するには、教員はオンライン試験またはマークシート方式の試験を実施する必要がある。

しかし、これだけでは不十分で、プライマリデータとセカンダリデータを IR において使い分けていくためには、教員はこれらのデータを IR 部門に提供する必要性が生じる。また同時に、IR 部門においては教員が作成したプライマリデータを保存する必要性が生じる。しかし、

これまでの役割分担ままではこのことはなし得ない。われわれの研究[5]–[11]が示すように、プライマリデータとセカンダリデータを保管・格納・活用するにあたっては、大学における教職協働が肝要になる。

2つ目として、大学教育における ICT (Information and Communication Technology) の有効活用である。ICT の活用は、大学のマイクロレベルでのプライマリデータとメソレベルでのプライマリまたはセカンダリデータを容易に比較することを可能としてきている。実際、学習管理システム (LMS : Learning Management System) や MOOC は、マイクロレベルで多くの主要なデータを簡単に取得できる。たとえば、ログインデータと学習の時間データは、マイクロレベルのプライマリデータとなっている。実際、近年、近藤らによって LMS ログデータ、つまり、プライマリデータと他のセカンダリデータを用いた研究が行われている[16]。

われわれは 2012 年以降、新入生に対して行った数学の能力に関する調査（以下、調査）において、各問についての正解、不正解のデータをすでに取得している。将来、これらのプライマリデータと大学内にある種々のセカンダリデータを使ってさまざまな解析を行う予定である。しかし、プライマリデータとセカンダリデータを用いて解析するにあたっては、さまざまな問題・課題が多数存在することが予測される。したがって、この調査結果を使って解析する場合には、プライマリデータの要素数を減らすために、項目反応理論を利用する必要性が生じる可能性が考えられる。

今回提案した学生データを使用するための新しいクライテリア、すなわちプライマリデータとセカンダリデータを活用することによって、とくに大学における教育の質保証に直接的につながる具体的な調査・解析結果を得ることが期待できる。

本論文は、2019 年の International Institute of Applied Informatics, Advanced Applied Informatics (IIAI-AAI) の第 8 回 Data Science and Institutional Research (DSIR) において発表した内容に加筆修正を加えたものである[10]。

## 【文献】

- [1] K. Takamatsu, K. Murakami, T. Kirimura, K. Bannaka, I. Noda, L. R.-J. Wei, K. Mitsunari, M. Seki, E. Matsumoto, M. Bohgaki, A. Imanishi, M. Omori, R. Adachi, M. Yamasaki, H. Sakamoto, K. Takao, J. Asahi, T. Nakamura, *et al.*, “‘Eduinformatics’: A new education field promotion,” *Bull. kobe Tokiwa Univ.*, vol. 11, pp. 27–44, 2018.
- [2] MEXT, “Towards the qualitative transformation of university education in order to build a new future,” *The Central Council for Education*, 2012. [Online]. Available: [http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo0/toushin/1325047.htm](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1325047.htm). [Accessed: 15-Mar-2018].
- [3] MEXT, “Construction of education for university,” *The Central Council for Education*, 2008. [Online]. Available: [http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo4/houkoku/080410.htm](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/houkoku/080410.htm). [Accessed: 15-Mar-2018].
- [4] 岩崎保道 and 鈴木弘道, “IR 組織の動向,” *関西大学高等教育研究*, vol. 8, pp. 93–101, 2017.
- [5] K. Takamatsu, T. Kirimura, K. Bannaka, I. Noda, M. Omori, R. Adachi, K. Mitsunari, T. Nakamura, and Y. Nakata, “SWOT analysis and complex network analysis to enhance governance

- in universities by collaboration between faculty and staff,” in *Advanced Applied Informatics (IIAI-AAI), 2016 5th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2016, pp. 1188–1189.
- [6] K. Takamatsu, K. Murakami, T. Kirimura, K. Bannaka, I. Noda, M. Yamasaki, L. Rahpael-Joe, I. Wei, K. Mitsunari, T. Nakamura, and Y. Nakata, “A new way of visualizing curricula using competencies: Cosine similarity, multidimensional scaling methods, and scatter plotting,” in *Advanced Applied Informatics (IIAI-AAI), 2017 6th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2017, pp. 192–197.
- [7] M. Seki, Y. Nakata, Y. Kozaki, K. Murakami, E. Matsumoto, K. Bannaka, I. Noda, K. Mitsunari, and K. Takamatsu, “Generic Support System Prototype in Eduinformatics for Students Selecting a Seminar,” in *Advanced Applied Informatics (IIAI-AAI), 2018 7th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2018, pp. 470–473.
- [8] K. Murakami, K. Takamatsu, Y. Kozaki, A. Kishida, K. Bannaka, I. Noda, J. Asahi, K. Takao, K. Mitsunari, T. Nakamura, and Y. Nakata, “Predicting the Probability of Student Dropout through EMIR Using Data from Current and Graduate Students,” in *Advanced Applied Informatics (IIAI-AAI), 2018 7th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2018, pp. 478–481.
- [9] K. Takamatsu, Y. Kozaki, K. Murakami, K. Bannaka, I. Noda, W. Lim, Raphael-Joel, K. Mitsunari, T. Nakamura, and Y. Nakata, “A New Way of Visualizing Curricula Using Competencies : Cosine Similarity and t-SNE,” in *Advanced Applied Informatics (IIAI-AAI), 2018 7th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2018, pp. 390–395.
- [10] Y. Nakata, K. Murakami, Y. Kozaki, T. Kirimura, A. Sugiura, K. Bannak, and K. Takamatsu, “New Proposal to Compare Student Data in Institutional Research,” in *Advanced Applied Informatics (IIAI-AAI), 2019 8th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2019, pp. 404–407.
- [11] K. Takamatsu, K. Murakami, T. Oshiro, A. Sugiura, K. Bannaka, and Y. Nakata, “Predicting the Probability of Student’s Academic Abilities and Progress with EMIR and Data from Current and Graduated Students,” in *Advanced Applied Informatics (IIAI-AAI), 2019 8th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2019, pp. 359–362.
- [12] UNESCO, “Learning Analytics,” 2012.
- [13] D. Norris, L. Baer, J. Leonard, L. Pugliese, and P. Lefrere, “Action analytics: Measuring and improving performance that matters in higher education,” *Educ. Rev.*, vol. 43, no. 1, p. 42, 2008.
- [14] “DQC: Data Quality Campaign.” [Online]. Available: <https://dataqualitycampaign.org/>. [Accessed: 01-Apr-2018].
- [15] J. P. Campbell, P. B. DeBlois, and D. G. Oblinger, “Academic analytics: A new tool for a new era,” *Educ. Rev.*, vol. 42, no. 4, p. 40, 2007.
- [16] N. Kondo, M. Okubo, and T. Hatanaka, “Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data,” in *Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017*, 2017, pp. 198–201.