

教学 IR における LMS ログデータ活用の試み

A Trial for Application of LMS Log Data on Institutional Research

近藤 伸彦^{*1}, 畠中 利治^{*2}
Nobuhiko KONDO^{*1}, Toshiharu HATANAKA^{*2}

^{*1} 首都大学東京大学教育センター

^{*1} University Education Center, Tokyo Metropolitan University

^{*2} 大阪大学大学院情報科学研究科

^{*2} Graduate School of Information Science and Technology, Osaka University
Email: kondo@tmu.ac.jp

あらまし：近年，教育機関において大規模なデータがリアルタイムに蓄積されるようになるとともに，ラーニングアナリティクスなどの分野が大きく発展している．同時に教育の質保証の観点から教学 IR の重要性が認識されてきているが，こうした大規模な教育データ分析をいかに教学 IR に融合させるかは大きな課題である．本発表では，LMS ログデータを用いた教学 IR へのアプローチについてまとめ，教育機関レベルでの修学支援や教育改善へ大規模データを活用する可能性と今後のありかたについて議論する．

キーワード：教学 IR，大規模教育データ，ラーニングアナリティクス，LMS，ログデータ

1. はじめに

大学教育の質保証の文脈から，教学 IR (Institutional Research) の重要性が認識されるようになって久しい．教学 IR はエンrollment・マネジメントと関連が深く，学習成果の可視化や適応的な修学支援は教学 IR における重要なトピックとなっている．

一方，テクノロジーの発展にともなって，教育に関する大規模でマルチモーダルなデータがリアルタイムに蓄積できるようになり，ラーニングアナリティクスやエデュケーション・データマイニングなどの分野が大きく発展している．

本稿では，こうした大規模な教育データ分析と教学 IR の統合という課題に焦点をあてる．とくに LMS ログデータを用いた予測モデルを教学 IR の一環として活用することについて考察し，教育機関レベルでの修学支援や教育改善へ大規模データを活用する可能性について議論する．

2. 教学 IR における LMS ログデータ活用

2.1 ささまざまな粒度のデータの統合的な活用

ラーニングアナリティクスでは，学習者への介入等のアクションを行うために，学習者に関する予測モデル (predictive model) の構築が行われることが多い⁽¹⁾．予測モデルは統計的手法や機械学習によりデータに基いて構築される．ここで用いられるデータは，主に授業や e ラーニングコースなどの単位で収集された粒度の細かいデータであることが多い．

これに対して，教学 IR で用いられるデータは，個人属性や，出席などの行動データ，成績などの学習成果データといった，定期・不定期に得られる粒度の荒いデータが多い．

ラーニングアナリティクスで用いられるような粒度の細かいデータをマイクロレベル，教学 IR で用いられるような粒度の荒いデータをマクロレベルとして，

これらを統合したマルチレベルな教育ビッグデータ分析が必要であることも言及されつつある⁽²⁾，教学 IR からみれば，教育機関においてさまざまな形で蓄積されつつある粒度の細かいマイクロレベルデータを用いた予測モデルを活用することで，リアルタイムな学習支援や修学状態のモニタリング・予測などが可能になり，学生個々に応じたきめ細かな分析や修学支援を行うことにつながると考えられる．

2.2 教学 IR における LMS ログデータ活用の例

著者らはこれまで，教学 IR において学生のデータを整理するための概念として，学内に散在する大規模なデータを一元集約し学生ごとに時系列に整理した「学修ライフログ」を考えてきた⁽³⁾．これにより，時間進行にしたがってライフログのように教学データが蓄積されていくさまを模式的に表わし，教育機関の修学プロセス全体にわたる修学状態やその変化を把握しやすくしようとしている．従来の教学 IR においては粒度の荒いデータが扱われることが多いが，学修ライフログではデータの粒度は問わず，いかに粒度の異なるデータを統合的に扱うかが重要となる．

著者らはこれまでに，LMS ログデータを用いて修学状態を予測する試みを行ってきた．たとえば，隠れマルコフモデルを用いて学生の潜在的な修学状態を確率的に推定する手法⁽⁴⁾や，LMS ログデータから抽出した特徴量から機械学習を用いて GPA を予測する手法⁽⁵⁾などを提案してきた．前者においては，日ごとの LMS 操作回数について隠れマルコフモデルを構築し，修学へのコミットメントに関する潜在的な状態を推定しており，実際の授業への出席状況との関連が示唆されている．後者においては，X 大学における初年次前期の LMS ログデータを用いて同学期の成績を 2 値予測する問題を取り扱っている．目的変数を GPA レベル (全学生の GPA 平均から標

準偏差を減じた値を閾値とする2値)とし、説明変数は LMS ログデータから抽出した、プレイヤー起動数 (player)、夜間操作数 (night)、ログイン総数 (login)、学習活動開始回数 (start)、学習活動完了回数 (submission)、総ログイン秒数 (time) の各変数としている。入学時オリエンテーションの週に前期の授業週 15 週を加えた 16 週間について、各週までの全ログデータをもとに上述の説明変数・目的変数をもつ機械学習モデルを用いて予測を行っている。ランダムフォレストモデルを用いた場合、それぞれの週における「at-risk (成績が良くない)」な学生に関する予測性能は図 1 のようになっている。ここで Precision は予測ラベルが正解している割合、Recall は真のラベルを正しく分類している割合、F-measure は Precision と Recall の調和平均である。またこのとき、モデルにおいて計算される説明変数の相対的重要度の変化は図 2 のようになり、予測性能の時間変化と説明変数の相対的重要度の時間変化には関連がみられると考えられる。

これらの例から、LMS ログデータを用いることで、授業への出席に相当するような学生の修学状態に関する情報が扱えることが期待される。また、粒度の細かい時系列データを扱うことから、修学状態や予測性能の変化を一定の時間分解能で検出することができると考えられる。きめ細かな修学支援や、教育者への教育支援、施策やカリキュラムの検討など、より精緻な教学 IR に向けて LMS ログのようなマイクロレベルのデータが果たす役割は大きいと思われる。

2.3 予測結果の精査

予測モデルが用いられる研究および実践においては、モデルの予測性能に関心をもたれることが多い。高い予測性能が望ましいことは論を俟たないが、精度の向上をめざすとともに、予測が成功・失敗したケースを個別に精査することも重要であると思われる。たとえば予測が外れたデータに該当する学生について、実際の学生に対応している教職員の生の感覚と照らし合わせることで、現存するデータのみでの予測と現実的な印象とのギャップを明らかにするなどである。このようなデータの精査は、今後 IR において取得すべきデータや予測モデルの構造・変数などに対する知見の獲得につながる可能性がある。LMS ログなどの粒度の細かいデータについては、ローデータからの特徴量抽出の妥当性や、データの可視化にもとづく検討が必要である。

3. おわりに

本稿では、粒度の細かいデータを教学 IR に活用することについて、LMS ログデータの活用例を中心に検討した。

他の教学データ同様、LMS は種類も使われ方も機関によってさまざまであり、ある機関での数値例をそのまま参考にすることが困難なことも多いと考えられる。今後は、教学 IR においてマクロとミクロを

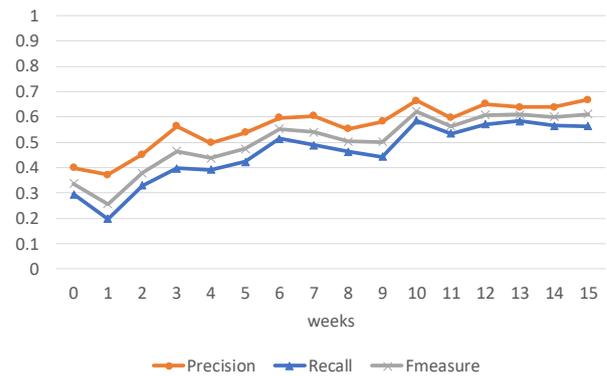


図 1 各週における予測性能 (5)より引用)

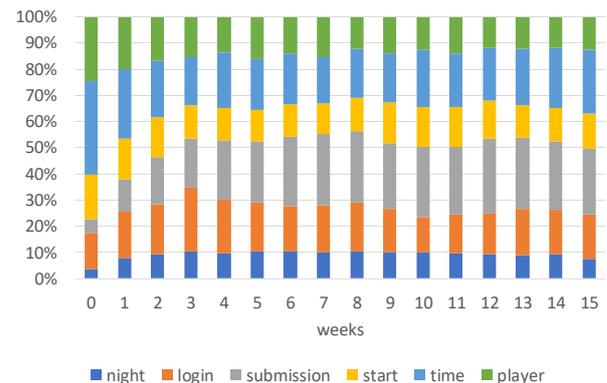


図 2 説明変数の重要度の変化 (5)より引用)

統合した予測モデルの構築と評価に関する汎用的な枠組みを作り、さまざまな機関の事例を共有するしくみが必要と考える。

謝辞

本研究の一部は JSPS 科研費 JP16K16331 および JP16H03082, ならびに首都大学東京傾斜的研究費の助成を受けた。

参考文献

- (1) C. Brooks and C. Thompson: “Predictive Modelling in Teaching and Learning”, Handbook of Learning Analytics, pp. 61-68 (2017)
- (2) 船守美穂: “デジタル技術は高等教育のマス化問題を救えるか?—MOOCs, 教育のビッグデータ, 教学 IR の模索”, 情報知識学会誌, Vol.24, No.4, pp.424-436 (2014)
- (3) 近藤伸彦, 畠中利治: “学士課程における大規模データに基づく学修状態のモデル化”, 教育システム情報学会誌, Vol.33, No.2, pp.94-103 (2016)
- (4) 近藤伸彦, 畠中利治: “LMS ログデータからの修学モデル構築”, 教育システム情報学会第 41 回全国大会講演論文集, pp.401-402 (2016)
- (5) Nobuhiko Kondo, Midori Okubo, and Toshiharu Hatanaka: “Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data”, 6th International Conference on Data Science and Institutional Research (2017, accepted)