

大規模学修データからのライフログ抽出とその活用

The Lifelog Extraction and Application from the Big Data of Learning

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

^{*1} 大手前大学 CELL 教育研究所

^{*1} CELL Institute for Educational Development, Otemae 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: nkondo7@otemae.ac.jp

あらまし：近年、教学上の改善・発展に関して、学内外の大規模データに基づいた問題発見／解決への注目が高まっている。本研究では、入学前から4年間の学士課程を通してのあらゆるデータを時系列に整理してこれを学修ライフログ化し、これと個人特性データを含めた大規模データから学修成果の評価や学修支援のための知識発見を行うことについて検討する。

キーワード：大規模学修データ、ライフログ、教学 IR、学修支援、数理モデル、データマイニング

1. はじめに

近年、ビッグデータ、データサイエンス等の用語に代表されるように、大規模なデータに基づくエビデンスベースでの意思決定が重要視され、さまざまな分野において議論、研究、実行されている。

高等教育においてもそれは例外になく、教育の質保証において、データに基づく教育プログラムの評価や施策の立案、成果の可視化等が喫緊の課題となっている。これに関連して、IR (Institutional Research), EM (Enrollment Management), LA (Learning Analytics) などさまざまな観点から、データに基づく教育改善のアプローチが研究されている。

本稿では、大手前大学（以下、本学という。）における入学前から4年間の学士課程を通じた学修データの収集と活用の事例や計画について報告し、大規模データから学修成果の評価や学修支援のための知識発見とその活用を行うことについて検討する。

2. 大規模学修データの活用

2.1 教学上の KPI 策定

教育改善のための教学 IR が担うべき重要な機能として、KPI（重要業績評価指標）の策定と、これに基づく具体的方策の提案がある⁽¹⁾。本学では、ディプロマポリシーに基づいた卒業生の質保証を実現するための「学士課程教育のグランドデザイン」を制定し、全学的にすべての教育活動がこれを参照してその実現に向けて活動することとしている。教学 IR の立場においては、あらゆる教育活動がこのグランドデザインに沿って設計されているかどうか、また学生がその学修成果としてグランドデザインに示された目標値を達成しているかどうかを明らかにすることが求められる。グランドデザインに基づく活動の実施状況や学生の目標達成状況などについて適切な指標と測定方法を定め、これを KPI として PDCA サイクルを回すことが必要である。

2.2 学修支援

学修支援の観点からは、入学前から学士課程を通してのあらゆる学生データを紐付け、その特性を詳細に分析することで広い意味での学修支援を行うことが重要であり、これも教学 IR の重要な役割である。なかでも、学習困難に陥っている学生や、中途退学の可能性がある学生をいち早く発見し、適切な支援を行うことが、質保証の意味でもリテンション率向上の意味でも求められる。

2.3 学生データの収集とライフログ化

本学において、学生に関するデータは従来から表1のように各部署や教員個別に管理されてきた。システムに関しても、教務システムや LMS (Learning Management System) など、複数のシステムが独自に開発され運用されてきたが、学内のさまざまなシステムを一元化するプロジェクトが発足し、数年計画

表1 学生に関するデータと所管部署

部署	主なデータ
教学運営室	入学前学習取組状況、出欠席（全授業・オリエンテーション等）、全学実施の試験結果、学習アンケート、インタビュー調査
学習支援センター	課題提出状況、学習支援センター利用状況、チューター報告書
教務課	成績、履修状況、三者面談記録、学生異動
学生課	学籍情報、アルバイト・サークル状況、奨学金、学生アンケート、特殊支援報告書
アドミッションズ・オフィス	入試種別、出身校課程、出身校評定値、入試成績
情報メディアセンター	LMS 上の学習データ、LMS 利用ログ、その他システム利用ログ
キャリアサポート室	進路決定状況、就職活動履歴、面談結果
就業力育成支援室	学修振り返り、プレゼン映像、自己アピール
資格サポートセンター	資格講座受講状況、資格取得状況
図書館	図書貸出履歴、OPAC 利用ログ
各教員	各授業の課題、テスト結果等

でこれを実行している最中である。

最終的にはシステム上で学生のさまざまなデータが一元化されることが望まれるが、まずは学内における教学 IR や教学評価研究の担当部署により各部署のデータを集約し、各種のデータを横断した学生データを作成しつつある。入学前から 4 年間の学士課程を通してさまざまなタイプのデータが存在するが、全ての授業の出欠席状況や、LMS 等のシステムログ、課題のデジタルデータなどを含め、大規模なデータを時系列に整理し、学修ライフログとしてこれを活用することを目指す。その際には、先述のグランドデザインに基づく KPI や、学修支援のための情報の可視化を十分に意識することが必要である。

3. 数理的アプローチとその事例

3.1 数理的アプローチ

大規模な学修ライフログを構築することができたとしても、莫大・多様で複雑に関係しあうデータからいかに意味のある情報を抽出するかが問題である。

効果のある教育プログラムや、高い成果をあげる学生の成長フロー、支援すべき学生などを発見するにあたり、学生・教員個人から教育プログラムまで、教育の場そのものを統合的な大規模システムとして捉え、その非線形で動的な構造のモデルをデータから推定しこれを活用していくことが有効と考えられる。そのためのアプローチとして、データマイニングや機械学習の手法を用いたさまざまな検討が近年提案されつつある。入学前後から 1 年次の種々の修学状況データにより 3 年次前期までの単位取得状況を推定する決定木分析を行った例⁽²⁾、さまざまなデータマイニング手法により、学生の学修成果を予測する数理モデルを構築する例⁽³⁾などがある。

3.2 数値実験

学修データをデータマイニング的に分析する例として、2009 年度入学生について、入学前から初年次までのデータにより経年後の状態を推定する数理モデルの構築を試みた。クラス分類問題と捉え、以下の 2 種類のクラス分類のためのモデル構築を行った。

- (1) 《修学状況》「4 年間のうちに退学・除籍/4 年間で卒業/4 年で卒業できず留年」の 3 クラス
- (2) 《学修成果》「卒業時に進路決定/未決定」の 2 クラス

入力変数（特徴量）としては、表 2 に示す 12 の変数を用いた。数理モデリング手法としては、ニューラルネットワークの一種である多層パーセプトロン（MLP）と RBF ネットワーク（RBFN）、および決定木アルゴリズムの C4.5 を用い、機械学習ツール Weka⁽⁴⁾ により実行した。データ数は(1)については 794、(2)については 621（退学・除籍者を除く）であった。10-fold cross validation の結果を表 3 に示す。

入学前から初年次のデータを用いることでその後の退学・除籍や留年等の修学状況を予測し、支援の必要な学生を早期に発見できることが示唆される。

表 2 入力変数

変数	値
性別	名義尺度, 2 クラス
学部	名義尺度, 3 クラス
入試種別	名義尺度, 7 クラス
出身校課程	名義尺度, 5 クラス
出身校評定値	連続値, [0, 5]
導入教育（入学前学習, 新入生オリエンテーション）参加率	連続値, [0, 1]
必修 4 科目出席率（春学期）	連続値, [0, 1]
必修 4 科目出席率（秋学期）	連続値, [0, 1]
1 年次春学期末修得単位数	離散値（最大値 20）
1 年次秋学期末修得単位数	離散値（最大値 48）
1 年次春学期末通算 GPA	連続値, [0, 4]
1 年次秋学期末通算 GPA	連続値, [0, 4]

表 3 クラス分類実験結果（正答率）

手法	(1)修学状況	(2)学修成果
MLP	80.7%	58.8%
RBFN	82.6%	63.6%
C4.5	82.9%	59.4%

進路決定状況に関しては、今回のようなデータのみでは精度良く予測することは難しい。

今回の実験では入学前から初年次のデータのみを入力変数として扱ったが、より広範囲のデータを結合し、意味のある変数やルールを発見することで、学修成果や学修支援に活きる KPI の策定を推進することが可能と思われる。またこれを実際の施策策定に結びつけるには、現実的に活用できる数理モデルの構築が必要であり、モデルの可解釈性も重要な視点である。また、教育環境等の変化を吸収する動的なモデル構築も求められるであろう。

4. おわりに

本稿では、学修データの集約とライフログ化、それによる教育改善についての要点と本学の事例をまとめ、簡単な数理的モデリングの例を示した。今後はさらに広範囲のデータを結合した大規模なライフログ化を行い、数理的アプローチを援用しながら教育改善のための知識発見の手法を開発していきたい。

参考文献

- (1) 松田岳士: “教学 IR の役割と実践事例”, 教育システム情報学会誌, vol.31, No.1, pp.19-27 (2014)
- (2) 雨森聡, 松田岳士, 森朋子: “教学 IR の一方略: 島根大学の事例を用いて”, 京都大学高等教育研究, 第 18 号, pp.1-10 (2012)
- (3) S. K. Yadav, B. Bharadwaj, S. Pal: “Data Mining Applications: A Comparative Study for Predicting Student's Performance”, International Journal of Innovative Technology & Creative Engineering, Vol.1, No.12, pp.13-19 (2012)
- (4) Weka 3: Data Mining Software in Java: <http://www.cs.waikato.ac.nz/ml/weka/> (2014/6/25 アクセス)