

教育ビッグデータ解析による要ケア学生の早期抽出の実践

安藤 公彦[†] 宮坂 秋津^{††}

東京工科大学 先進教育支援センター[†]

1. はじめに

近年教育分野においても Learning Management System(LMS)の導入や各種情報の IT 化により、ビッグデータの活用が進んできている。例えば Learning Analytics(LA)の分野においては、LMSの課題の内容・進行状況や各種ログから、各学生の学習状況を把握することで、講義進行や講義内容の動的な変更や、グループ学習におけるグループ状態の可視化などが可能となっている[1]。また Institutional Research(IR)の分野においては、各種データを統合的に分析することで、大学全体の状況の可視化や、各学部の違いなど多くの分析が可能となってきた。

しかし、IR においては各種情報の統合が課題となっており、どのような仕組みで大学中の情報を取得し分析できる環境を構築するかが重要であると考えられる。

また、一方で大学における重要な課題の一つとして、何らかの問題を抱えているため、授業への不参加、ひいては退学へと追い込まれてしまう学生の発見そののケアが挙げられる。

本研究では、大学内における情報統合の実践とその活用である要ケア学生の早期抽出の手法について考えていきたい。

2. 組織内 IT サービス

大学に限らず組織内の IT サービスは、外部ベンダーが提供するサービスを利用する方式(外注)と、組織内でサービスを自前で構築する方式(自前)が考えられる。一般的には外注が安定性の観点から推奨される。自作する場合は予算の都合によるものがほとんどである。しかし、外部仕様がまだ確定していなかったり、仕様が頻繁に変更されたりする場合には、自作でサー

ビスを構築することのメリットが生まれる。

大学での IT サービスを考えると、要求仕様ははっきりしていてかつ安定性と安全性が必要とされる業務系のサービスは外注が望ましい。例えば教務システムや図書館システムである。一方、教員や学生が直接操作する LMS などの教育系サービスは、授業形態や利用のされ方により、常に要求が変わるため、自作が望ましいといえる。

そこで本学では IT サービスを業務系と教育系に分け、業務系は外注のサービス、特にメンテナンス性と安定性の観点からクラウドサービスを導入した。一方、教育系はオープンソース LMS である Moodle に自作のモジュールを追加して利用し、出席登録システムを兼ねる学生ポータルを自作した[2]。図 1 にその概略を示す。教育系に関しては応答性や通信料を踏まえ、オンプレミスのサーバーで運用をしている。

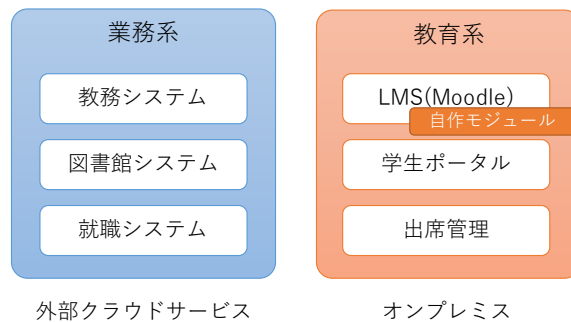


図1 学内 IT サービス環境

3. 情報統合手法

外部の各種サービスを導入している組織において、すべてのサービスのデータを統合することは、すべての外部サービスを同一ベンダーにする以外は非常に困難である。本学では、業務系の各種サービスをクラウド化するとともに、異なるベンダー間のサービスにおいてもデータの統合を可能とするための中核 DB を設置した。

図 2 に中核 DB の動作を示す。中核 DB はクラウド上にある DB サーバーであり、各ベンダーに業務系各種サービスの DB のレプリカをこの中核

Practice of early extraction of students in need of care by analyzing educational big data

[†]Kimihiko Ando

Advanced Education Support Center,
Tokyo University of Technology

^{††}Akitsu Miyasaka

Advanced Education Support Center,
Tokyo University of Technology

DB 上に設置するよう依頼することで、各種サービスのデータを採取できるようにしたものである。どのテーブルのどの列がどのような意味であるか適切に管理することで、サービス間のデータの統合が可能となる。基本的に学生に関するデータは学籍番号をキーとして統一的に扱うことが可能である。

一方、教育系サービスは自作のため、中核 DB へのデータのアップロードは容易であり、中核 DB 内のデータ構造に合わせてシステムを構築することが可能となっている。

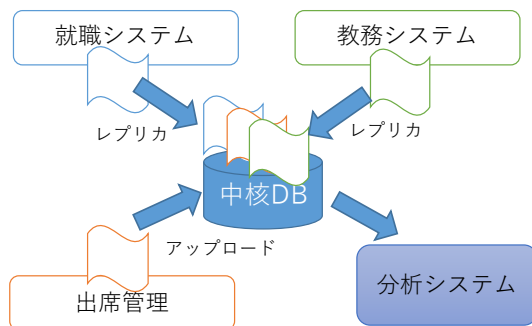


図2 中核DB

4. 要ケア学生の検出

従来、ケアが必要な学生の発見は、各講義の出席状況や成績の取得状況等から、ある程度把握されてきていた。しかし、成績が悪くなった時点ですでにケアするには遅すぎる場合が多く、それは出席状況が悪くなってからでも同じである。理想としては、出席状況が悪くなる前にケアすべき学生の抽出をすべきである。特に、入学時には問題がなく、学生生活を続けていくうちにケアが必要となる学生の検出が必要とされる。そのためには、単なる出席や成績より細かな情報が必要と考えられる。

本学では、教育系システムが自作のため、本学独自のより詳細な出席システムを開発した。どの座席からいつ出席登録したかを記録できるため、より詳細な出席状況の把握が可能である。また、毎学期、全講義の Moodle コースを作成しており、すべての講義での Moodle 利用を推奨している。そのため、学生の学習状況についてもより詳細に分析することが可能であると考えられる。

5. 検出手法

要ケア学生の抽出には、出席情報や LMS のログに加え、授業評価アンケート結果や各種業務系サービスの情報など、多岐にわたるデータ収

集及び分析が必要となる。データの収集は中核 DB を中心とし、各種サービスのデータの統合を行う。また、出席情報からは教壇からの位置関係や、隣接する学生の情報などから、今までにない分析が可能になると期待できる。また、分析手法としてデータマイニングを行うだけでなく、ディープラーニングによる分類器の作成を予定している。過去4年間の出席情報及び Moodle のログから、退学や休学してしまった学生の推測を行う。なおデータの 9/10 を教師データとして使い、残り 1/10 を検証用データとして使う。

6. データ

分析に用いるデータについて表1に示す。本学の入学者数は年間約2000人であり4年間分で8000人となっている。出席データは、出席の有無のデータが530万件であり、座席や時間も含まれた出席データの場合は400万件程度となる。Moodleの行動ログについては学生の行動だけでなく教員やシステム等の行動も含まれている。アンケートは基本的に Moodle 上で行われたものを対象とし、授業評価アンケートのみならず、全学生対象の普段の状況を調べる全学調査アンケートや、各授業で行われたアンケートも含まれている。

表1 分析に用いるデータ

学生数	約8000人
期間	2015年4月~2019年3月
出席データ	約530万件
Moodle行動ログ	約4000万件
アンケート結果	約60万件

謝辞

*本研究は JSPS 科研費 16K01134 の助成を受けたものである。

参考文献

- [1] 安藤公彦、柴田千尋、稲葉竹俊、深層学習技術を用いた自動コーディングによる協調学習のプロセスの分析, コンピュータ&エデュケーション, 43, pp79-84
- [2] 安藤公彦, アクティブラーニング全学導入にむけた ICT 基盤の構築と実践, 平成 29 年度教育改革 ICT 戦略大会