

LMS等を利用したオンライン科目の学習ログの 抽出及び分析環境の検討

中野 裕司^{1,2,3,4,a)} 久保田 真一郎^{5,2} 松葉 龍一^{4,2,1} 戸田 真志^{1,2,3} 永井 孝幸^{1,3,2,4}
右田 雅裕^{1,2} 武蔵 泰雄^{1,3,2} 喜多 敏博^{4,2,1} 杉谷 賢一^{1,3,2,4}

概要 : LMS等を用いたオンライン科目は学習者の提出物、掲示板への書き込み、クイズ受験等、その学習成果物が蓄積されるだけでなく、いつどこからなにをアクセスした等の詳細な情報が、LMS中やそのベースとなるWebサーバのログとして記録される。これらの情報は、学習者の学習動向を分析しそれをフィードバックや科目設計に活かすための貴重な情報となり得るが、様々な形式でいくつかの場所に保存されていることが多く、また、その量も場合によっては大変大きくなることもある。本発表では、具体的な事例に基づき、学習ログの抽出及び分析環境を検討したので、その結果を報告するとともに議論したい。

1. はじめに

オンライン学習が一般化してきた今日、テキストやクイズ、レポート、掲示板等の学習コンテンツがオンライン化されるだけでなく、個々の学習者の学習活動の履歴やその成果物、成績もデジタルデータとして保存されるようになってきた。これらのデータから、学習者がどのような環境で、どのように学習を進め、どのような成果を残したか等、多くの情報を得ることが可能で、学習につまずいた箇所からコンテンツや教授設計の見直しを行ったり、落伍しそうな学習者を発見して個別指導を行ったり、効率的な学習を行っている学習者を調べてその方法を知ったり、様々な活用が考えられる。これらの学習データの収集・分析を Learning Analytics (LA) と呼ぶ [1]。

LMSのログ解析との相違点について、安武は「従来のログ・データの活用に期待されていたこととLAという概

念とで大きく異なっているのは、後者が膨大なデータの解析から見えてくる(であろうと期待されている)学習メカニズムやそれに関する様々な特性の理解に基づいて、real timeな実践的活用を目指している点である。」と指摘している [2]。

本報告では、LMSを活用して受講者1000名規模で実施している2年次開講の情報処理科目「情報処理概論」を対象として、その学習データ収集を中心に報告する。

2. 対象科目

本科目は、情報処理技術に関する知識習得を目的としており、その実施形態は、図1に示すように、教室でのガイダンスに続き、7週間のオンラインの学習を経て、PC教室でのCBT型期末試験で構成される。2013年度、一部学部が必修から選択になったこともあり、実質的な受講者は、2012年度1084名、2013年度859名であった [3], [4]。

最初の7週間は、毎週、オンラインで、テキストや関連記事による学習、小レポートの提出、確認テストの受験を繰り返す。小テストは少なくとも1回は確認テストを受けなければ提出できない仕組みで、締切が毎週日曜日の夜中に設定されている。図2に2013年度の確認テストや小レポートのべ受験者数を示すが、締切日は、10月6日にはじまり、7週目の締め切りが11月17日である。確認テストは何度でも受験でき、1人平均4回程度受験している。7週目の締切までは、1-7週の全確認テストの成績は成績の一部としてカウントされるため、7週目の締切数日前から全確認テストの受験が増加している。

7週目の締切以降は、確認テストは受験できないかわり

¹ 熊本大学 総合情報統括センター
Center for Management of Information Technologies, Kumamoto University, Kumamoto 860-8555, Japan
² 熊本大学 大学院社会文化研究科 教授システム学専攻
Graduate School of Instructional Systems, Kumamoto University, Kumamoto 860-8555, Japan
³ 熊本大学 大学院自然科学研究科研究科 情報電気電子工学専攻
Graduate School of Computer Science and Electrical Engineering, Kumamoto University, Kumamoto 860-8555, Japan
⁴ 熊本大学 eラーニング推進機構
Institute for e-Learning Development, Kumamoto University, Kumamoto 860-8555, Japan
⁵ 宮崎大学 工学部 情報システム工学科
Department of Computer Science and Systems Engineering, Faculty of Engineering, University of Miyazaki, Miyazaki 889-2192, Japan
a) nakano@cc.kumamoto-u.ac.jp

に、練習用確認テストが受験可能となる。このテストは成績にはカウントされないが、最終成績の7割としてカウントされる期末テストのテスト勉強のために公開される。最終成績の7割の得点が配分される期末テストは、4回に分けてPC教室でCBT型で行われ、問題プールから一定数ランダムに出題される問題を、管理されたPC上で時間内に何度でも受験でき、そのうち最高点が期末テストの成績となる。

3. データ取得

3.1 シェルスクリプト

図3にWebアプリケーションサーバのログから学習者毎のデータを抽出するシェルスクリプトの抜粋を示す。また、図4に、日や時間毎に抽出する場合の例を示す。これらは、Linuxのbashシェル上で実行したもので、cat, grep, sed, sort, uniq, wc -l, date 等のコマンドを組み合わせたものである。何れの場合も、非常に煩雑な処理を、コマンドの組み合わせで実施しており、作成も大変であるし、可読性もよいとは言えない。

3.2 データベース

3.1に示すようにシェルスクリプトでデータ抽出を行う場合、複雑な処理が必要になるだけでなく、一般的に、ファイルを毎回走査することになる。特に、対象が巨大なデータの場合、処理時間も多くなる。このような場合、一般的に、データベースに入れることで、検索、抽出の速度を上げることが可能になる。

例として、チェックテスト等の受験記録を扱うために作成したデータベースのあるテーブルの構造を、表1に示す。今回使用したデータベースは、MySQL互換で、フリーでオープンソースのMariaDB[5]を用いた。表1の例では、元データがLMSで表示した表(HTML)と、別のCSVデータであり、その値をテーブルにインポートする必要がある。Web経由で管理できるツールであるphpMyAdmin[6]を利用すると、GUIでCSVファイルのインポートも可能であるが、大きなファイルの場合は、ウェブサーバ上のアップロードディレクトリから直接アップロードしないとタイムアウトしてインポートできない場合もあるし、時間もかかるので注意が必要である。速度面を考えれば、データベース中のコマンド load data コマンドを用いたほうが得策と考える。

HTMLで表示した表をデータベースに入れるには、なんらかの変換処理を行わねばならない。今回は、スプレッドシートへの貼り付けがうまくいかなかったこと、HTMLのテーブル構造が複雑であり、他のファイルとの組み合わせが必要であったため、プログラミング言語で処理を行った。具体的には、Java言語と、データベース接続用のAPIであるJDBCコネクタを用いて処理したが、perl等でCSV

表1 チェックテスト等の受験情報を保存するためのテーブル構造
Table 1 Table structure for storing data of check-test etc.

カラム	データ型	NULL	デフォルト値	コメント
courseName	varchar(80)	いいえ		
year	int(11)	いいえ		
uid	varchar(8)	いいえ		
scoreName	varchar(40)	いいえ		
scorePoint	double	はい	NULL	
trial	int(11)	はい	NULL	
doneAt	datetime	はい	NULL	
duration	time	はい	NULL	

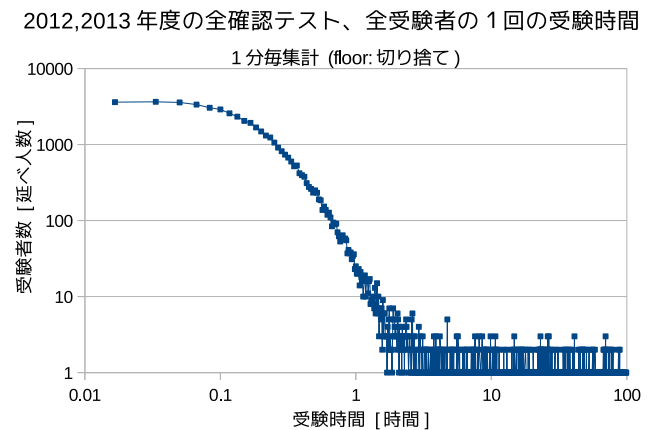


図5 2012,2013年度の全受験者、全確認テスト(含練習用)に関する1回の受験時間

Fig. 5 Testing time of all check-test (including free) for all learners in 2012 and 2013.

に変換して、DBに読み込ませる方法もあろう。今回の例では、1つの科目の2012年度と2013年度の2回分の全テストに関するデータを読み込んだが、全部で、174,823件のデータであった。

データベース上のテーブルを利用して、確認テスト等の受験時間を調べたものが図5で、全ての確認テストや期末試験準備用の確認テストに関して、受験時間とのべ人数の関係を見ている。ここで、100時間までデータが続いているが、これは、LMS内部のタイムアウトが100時間に設定されていたことによる。1時間で100分の1以下に減少することを考え、1時間以降を全て1時間に切り上げるという判断に利用した。

図5のデータ抽出に用いたのが、図6に示すSQLである。このように、時間のスコープを決めたり除権を細かく設定できるが、SQLの知識が必要となる。

3.3 データベースとプログラミング言語の組み合わせ

図7に、2013年度の情報処理概論の練習用第3回確認テストにおける成績が70点以上80点未満の者の受験回とそれまでの最高得点の分布を示す。この場合、第3回のものであるが、同様の処理を1-7回について行いたいし、成

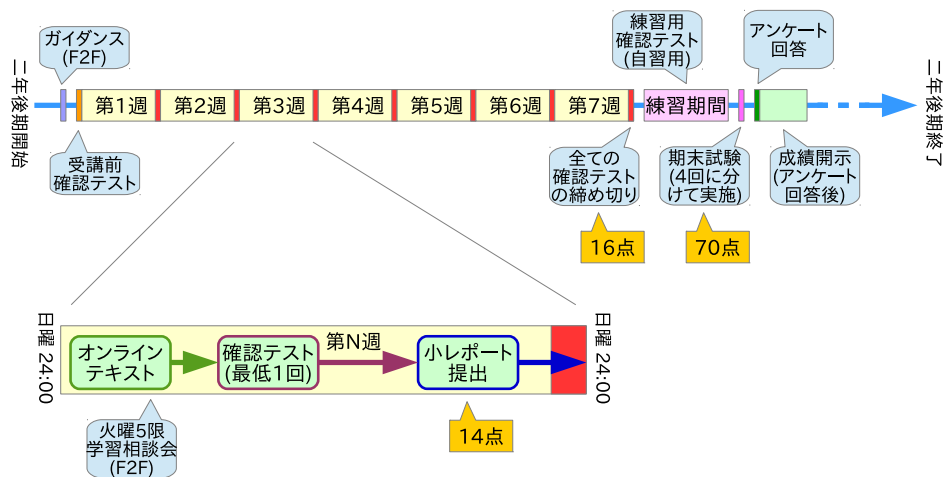


図 1 授業のスケジュールと流れ
 Fig. 1 The schedule and flow of the course.

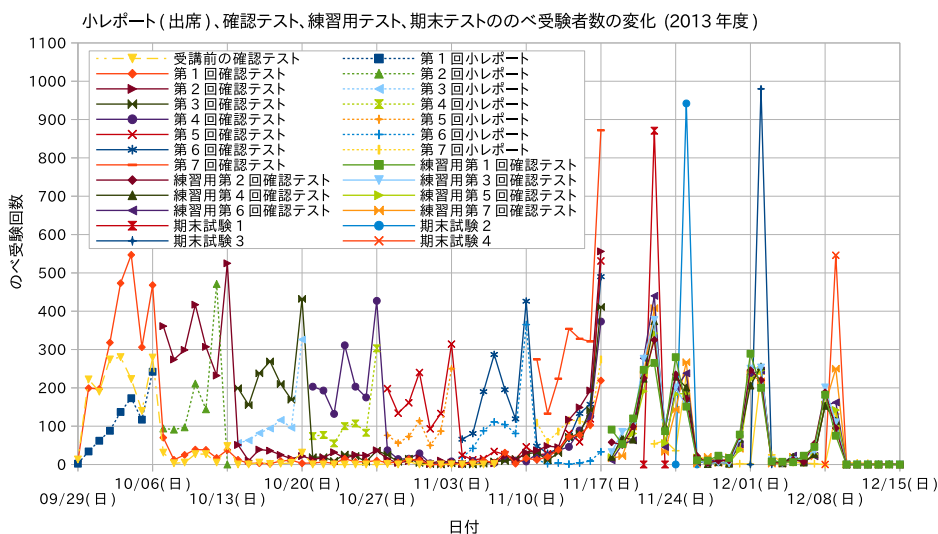


図 2 確認テスト、小レポート、練習用確認テスト、期末テスト等の受験回数 (2013 年度)
 Fig. 2 The total trial times of check-test, report, check-test for preparing final examination, final examination (2013).

```

select d,count(d)
from
  (select 1*floor(time_to_sec(duration)/60)
   as d
  from score
  where courseName='gairon'
        and (year=2012 or year=2013)
        and scoreName like 'No%chkt'
  )
as a group by d order by d asc";
    
```

図 6 図 5 の抽出に用いた SQL
 Fig. 6 SQL to obtain Fig.5.

績も、全てのレンジで行いたい。また、結果は別のファイルに書き出したい。さらに、年度別、練習用かそうでないかなどの分類もしたい。

このような場合は、SQL だけで行うのは無理ではないかもしれないが、かなり複雑になる。プログラミング言語であれば、for ループや条件分岐、文字列の合成等が容易に行えるため、プログラミング言語から SQL を呼び出すのが効率的と考えた。実際には、Java 言語と JDBC コネクタを組み合わせ、全年度、練習用も含めた全ての回のテスト、全ての得点レンジについて別々の CVS ファイルを作成した。

4. 問題点

学習ログの抽出方法に関して、シェルスクリプト、SQL、SQL とプログラミング言語の組み合わせで行った例を示した。しかし、何れの方法も、あまり容易とは思えず、また可読性もあまりよいとは言えない。さらに大きなデータになると、データベースで扱うことも難しくなることが予想

参考文献

- [1] Learning analytics in Wikipedia, 入手先
(http://en.wikipedia.org/wiki/Learning_analytics)
(2014.09.24 確認)
- [2] 安武公一, ライフログの教育活用における海外動向 —
Learning Analytics and Knowledge (LAK) 2012 報告—,
サイエンティフィック・システム研究会第2回会合, 入手先
(http://www.sskn.gr.jp/MAINSITE/download/newsletter/2012/20121024-edu-2/lecture-02/SSKEN_edu-2012-2_yasutake_paper.pdf) (2012.10.24).
- [3] 久保田 他, 情報処理科目におけるオンラインの雑誌記事
の活用, AXIES 2012 年次大会講演論文集, 69-73 (2012).
- [4] 中野他 (2012) オンラインテキストを活用した大規模情報
処理科目の実践, 日本教育工学会第30回全国大会講演集,
533-534 (2014).
- [5] MariaDB ホームページ <https://mariadb.org/> (2014.09.24
確認)
- [6] PhpMyAdmin ホ ム ペ ー ジ
<http://www.phpmyadmin.net/> (2014.09.24 確認)

