

LMSにおける学習ログを基にした 学習者の振る舞い変化を検知する試み

久保田 真一郎^{1,a)} 岡崎 直宣^{2,b)}

概要: アダプティブな学習支援を行うためのシステムを開発するためには、そのトリガーとなる学習者の変化を検知する仕組みが必要である。オンライン学習を支援するシステムである学習支援システム(LMS)は学習者の操作やアクセスをログ(学習ログ)として記録する機能を持つ。この学習ログは学習者の行動をプログラム処理して扱うことを可能にし、その変化を検知することでアダプティブな学習支援を可能にすると考えている。初年次学生を対象にLMSを利用したブレンディッド型の教育実践を通して、LMSに記録される学習ログを基に時系列特徴ベクトルを生成し、ある窓幅内の時系列特徴ベクトルを学習用データとして主成分分析を行う。主成分分析によって得られる固有ベクトルとその後の学習ログをもとに各主成分の時系列変化を考察することで、学習者の振る舞いの変化を検知することを試みた。学習者の振る舞いの変化を検知する本研究の成果はアダプティブな学習支援を実現する上で有効であると考えている。

Consideration for Detecting Variation of Learners' Behavior based on a Learning Log of LMS

Abstract: In order to develop the system to support learners adaptively on learning context, it is necessary to detect the variation of learners as a trigger. The Learning Management System(LMS), to support learners on the online learning, has a feature to record a learning log, which is data to record learners' manipulation and access. This learning log may enable to process learners' activities as programmable data, and to support learners adaptively through detecting the variation of learners' behavior. In our practice of a blended learning course for the First-Year student, time series characteristic vectors were derived, based on a learning log of LMS. And We applied principal component analysis (PCA) to a data set, which was constructed from time series characteristic vectors in a certain period. Time series data of each principal component were derived from an eigenvector of the PCA and each time series log data. Through observing the variation of the time series data, the learners' variation would be detected. This research outcome would affect to the adaptive learning system.

1. はじめに

学習管理システム(LMS)を利用したオンライン学習が開始された当初からLMSのもつ学習ログを用いた分析は数多く行われ、現在も多くの研究が行われている [1], [2], [3]. 学習ログは、LMS上で学習者が学習活動を行うとその操作やアクセスが記録されるもので、学習者が意識すること

なく学習ログは蓄積される。このことは、学習者にとっても、また状況を把握したい側にとっても負担のないセンサ情報だと考えられる。LMS上のコースに学習活動に関わるコンテンツを配備することは学習者をセンシングすることにつながり、より多くの学習者情報を収集でき、学習者の状況を把握できるようになる。

学習者の状況を把握することは、学習者中心の学習を実践する際の学習者個人に適応したアダプティブな学習環境を提供するために必要となる。状況把握にとどまらず、学習者の行動に対してアダプティブな学習環境をシステムとして提供するためには、システム的に学習者の状況を把握

¹ 熊本大学
Kumamoto University

² 宮崎大学
University of Miyazaki

a) kubota@cc.kumamoto-u.ac.jp

b) okazaki@cs.miyazaki-u.ac.jp

した上で、学習者の変化を検知する仕組みが必要となる。そこで、本研究では、学習者個人の状況に合わせてアダプティブな学習支援を行うために、学習者個人の状況を負担なしに集めることのできる学習ログを用いて、学習者の振る舞いの変化を検知する手法について検討する。

2. 関連研究

Lakhina ら [4] は、巨大なネットワークのネットワークトラフィックのフローに対して主成分分析 (PCA; Principal Component Analysis) を適応し、その主成分を監視することでネットワークトラフィックの異常状態を検知できることを提案し、実際のネットワークトラフィックのログデータを用いて検証を行っている。Lakhina らの手法では、ISP ルータ上を流れるトラフィックの送信元ルータの IP アドレスと宛先ルータの IP アドレスの組み合わせ毎に 10 分間のトラフィックを時系列に 1 週間分並べた行列を構成し、その行列の主成分を求め、いくつかの主成分を考察することでトラフィックの異常状態を検知できることを示している。例えば、考察対象のルータが 10 個であれば 100 の組み合わせがあり、10 分間のトラフィックをそれぞれの組み合わせごとに値を集計し、時系列データの 1 つ目が構成される。これを繰り返すと、長時間データであればあるほど扱う行列の行数が増加し、扱うルータの数が増えると行列の列数が増加するため、ネットワークトラフィックの分析は高次元の分析となることが知られており、このような多変量データを低次元のデータに縮約する方法として PCA はよく用いられる。本研究においても、学習ログは様々な要素を持ち、時系列データになることは明らかである。従って、学習ログをもとに高次元の時系列データが抽出され、それらを低次元にして考察することで、多変量の時の情報をできるだけ失うことなく低次元で考察できると考えている。

3. 研究方法

学習ログをもとに時系列の特徴ベクトルを抽出し、分析の基本とするある期間 T_n から T_{n+1} の特徴ベクトルで行列を構成し、主成分分析を行う。ある時刻 $t_k (T_n \leq t_k < T_{n+1})$ の特徴ベクトルを

$$(x_1(t_k), x_2(t_k), \dots, x_p(t_k))$$

とすると、時系列データによって構成される行列は

$$\begin{pmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_p(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_p(t_2) \\ \vdots & \vdots & \cdots & \vdots \\ x_1(t_k) & x_2(t_k) & \cdots & x_p(t_k) \\ \vdots & \vdots & \cdots & \vdots \end{pmatrix}$$

となる。この行列に対して主成分分析を適応して、時刻 t_k における主成分 $y_i(t_k)$ は

$$y_i(t_k) = \sum_{j=1}^p h_{ij} x_j(t_k), \quad (T_n \leq t_k < T_{n+1})$$

として得られる。ただし、 h_{ij} は i 番目の主成分を与える固有ベクトルであり、行列 h_{ij} を他の区間 (T'_n から T'_{n+1}) における時系列データに適用することで、分析の基本とする区間 (T_n から T_{n+1}) と他の区間 (T'_n から T'_{n+1}) との変化を考察することができる。

$$y_i(t'_k) = \sum_{j=1}^p h_{ij} x_j(t'_k), \quad (T'_n \leq t'_k < T'_{n+1})$$

本研究では、各主成分の値が正規分布するものと仮定し、主成分の値のうち正規分布に従わない変化をホテリング理論 [5] により検知する。本来であれば多変量のホテリング理論により異常な値を検知すべきであるが、今回の研究では簡単のために 1 次元ホテリング理論により異常値を検出する。最後に、異常値が検知された時刻における学習ログを考察することで、学習者にどのような変化が見られるかを考察する。

3.1 分析対象

分析対象とした講義は、大学 1 年次の学生を対象に入学直後から行われる専門科目への導入として位置づけられる講義で、「情報工学序説」という講義である。内容は、情報工学全般をテーマに、各専門科目への入門を扱っている。この講義では、LMS を利用し、毎回の説明を提示し、事前課題や事後課題の回収を行っている。講義は、全 15 回で行われ、大きく次の 3 つの学習活動で構成される。

- 学習活動の解説と練習
- 実際の学習活動
- テストと振り返り

「学習活動の解説と練習」は、初年次の受講者が授業で行う学習活動を円滑に行うために、はじめの 2 回の授業で、学習活動に必要な資料作成方法および資料作成のポイントについて実際の学習内容を扱いながら取り組み、事前課題の作成練習や学習活動の練習、事後課題の作成練習を行う。

「実際の学習活動」は、この授業の主たる部分であり、授業前、授業中、授業後の 3 段階にわけて行われる活動で

ある(表1)。授業では教科書を指定し、各学習活動のために教科書の範囲を教員が指定し、教員が指定した範囲の学習内容についてグループ学習を行い、内容理解を行う形態で行っている。

表1 学習活動

時間	活動内容
授業前	担当する範囲の解説スライドと確認用クイズの作成
授業中	グループ活動(相互に説明, 相互にクイズ出題)
授業中から授業後	まとめスライドを作成(すべての範囲)

受講者は授業の前までに担当する範囲を解説するためのスライドを作成し、同時に解説内容を理解できたか確かめるためのクイズを作成して授業時間に臨む。受講者は授業の前までに資料を作成し、LMSへ事前課題として提出する。事前授業は講義日の前日を締切として提出するようLMSの機能で制御している。授業中には、教員が時間を決めて合図を出し、当該授業回の内容説明およびクイズ出題を各担当者が順番に行う。授業時間の最後に各自でグループ活動の結果を1つのスライドにまとめる活動を行う。学習内容をひとつにまとめたスライド(まとめスライド)は未完成であっても成果物として授業時間中に1度提出を求めている。そして、受講者は授業時間外にまとめスライドを完成させ、学習成果物としてあらためて提出することができる。1回の授業で提出すべき成果物は、事前課題、授業中にまとめた資料、授業時間外にさらにまとめた資料の計3つであり、すべての提出物に対してLMS上に提出先が設けられている。

1回の学習活動に関わるLMS上のコンテンツは、以下の5つである。「授業中の成果物提出制御のためのスイッチ」とは、授業に出席していない学習者の提出を制限するために設置しているものである。

- 事前課題の提出
- 学習活動の流れを示したテキスト
- 授業中の成果物提出制御のためのスイッチ
- 授業中の成果物提出
- 授業時間外の成果物提出

「テストと振り返り」では、それまでの学習内容を確認するテストを行い、各自が理解できていなかったポイントをレポートに記述する活動とそれまでの学習内容をもとにしたPCを利用した演習形式での学習活動を行った。「テストと振り返り」では、テストの予想点を入力した資料を事前課題として提出するよう課しており、授業時間には、演習内容を記述したテキストをLMSに掲載し、時間内に

成果物の提出を課している。LMSの学習コースには、「学習活動」と同じコンテンツが必要であり、「学習活動」のコンテンツに加えて、テストがコンテンツとして追加される。

授業全体を大きく4つのブロックに分け、ブロック1からブロック4の各ブロックの学習内容の確認として「テストと振り返り」を行った。第1回から第15回までの構成を表2に示す。

表2 第1回から第15回までの構成

授業回	活動内容
1	解説(学習目標, 評価方法, 学習方法, 15回の流れ, スケジュール, 学習マップ作成)
2	解説と活動(担当分の解説スライド作成とクイズ作成)
3	ブロック1学習活動1(第3回事前課題スライド提出練習, グループ活動, まとめスライド作成)
4	ブロック1学習活動2
5	ブロック1テストと振り返り
6	ブロック2学習活動1
7	ブロック2学習活動2
8	ブロック2テストと振り返り
9	ブロック3学習活動1
10	ブロック3学習活動2
11	ブロック3学習活動3
12	ブロック3テストと振り返り
13	ブロック3学習活動1
13	ブロック3学習活動2
15	ブロック4テストと振り返り

2015年4月15日から2015年8月5日まで実施された受講者60名の講義において講義を実践した。

4. 学習ログ

LMSにはMoodle(Ver.2.8)を用いており、学習ログはMoodleの標準ログを用いる。本来であれば学習者のすべての行動のログをもとに考察するべきであるが、今回は講義で設置した学習コースに限定して考察を行うことにした。

Moodleの標準ログを取得すると、以下の9つの項目の記述がある。

- (1) イベントの時間
- (2) ユーザフルネーム
- (3) 影響を受けたユーザ
- (4) イベントコンテキスト
- (5) コンポーネント
- (6) イベント名
- (7) 説明
- (8) オリジン
- (9) IPアドレス

「イベントの時間」は、受講者あるいは教員がコンテンツに対して何か行動を行った時刻を示す。「ユーザフルネーム」は、どの学習者あるいは教員が行った行動であることを示す。「イベントコンテキスト」は、どのコンテンツへアクセスしたかを示す。「コンポーネント」は、Moodleのどのモジュールを使っているかを示す。「イベント名」は、受講者あるいは教員が何を行ったかを示す。「IPアドレス」は、受講者あるいは教員のアクセスもとのIPアドレスが示される。

分析の対象とした学習コースの各コンテンツについて Moodle の標準ログを取得したところ、89630 レコードのログがあった。授業 15 回で、60 名の受講者であることから、1 人当たり 1 週間に平均 100 回程度アクセスしていることになる。

アクセスのあったユーザ数は 63 であり、受講者 60 名と教員 3 名のものである。イベントコンテキストの数は 85 であり、学習コースで利用されたコンテンツが 85 個あったことになる。授業が 15 回の構成と考えると 1 回の授業に平均して 5.7 個のコンテンツが配置されたことになり、前述の「学習活動」「テストと振り返り」とのコンテンツ数とほぼ一致している。学習コースを構成するために利用した Moodle のコンポーネントの数は、11 種類であった。イベントの種類は 37 であった。

Moodle 標準ログの各項目を数値化して分析を行う。11 種類であった Moodle のコンポーネントを例にあげると、項目「コンポーネント」の各要素に対して番号を付した。数値の意味は今回考慮していない。他の項目について要素すべてに番号を付して、1 つの CSV ファイルとした。

Moodle 標準ログのまま分析できないため、以下のデータクレンジング処理を行った。

- (1) 各項目のとり独立な要素に番号をつけ、標準ログを数値要素の CSV に編集する
- (2) 日時はすべて UNIX タイムに置き換える
- (3) 各 IP アドレスの第 2 オクテットまでが LMS サーバの IP アドレスと異なる場合に外部ネットワークとし、外部ならば第 1 と第 2 オクテットが同一であるものを同じネットワークとみなして番号付けを行う。また、第 2 オクテットまでが LMS サーバの IP アドレスと同じ場合、学内ネットワークと判断し、第 3 オクテットが同一であるものを同じネットワークとみなして番号付けを行う。

今回のデータクレンジング処理によって、ユニークなアクセス元 IP アドレスが 1341 個あったが、それらの IP アドレスが所属するネットワークにまとめることで 129 個に整理できた。

5. 学習ログの主成分分析

授業第 1 回から第 3 回にかけて授業内で行う学習活動の練習を含むため、通常の学習活動とは明らかに異なる操作やアクセスとなると予想されたため、本研究では第 4 回授業から 1 週間を基本データと考え、それ以降のデータについて考察することにした。第 4 回授業から 1 週間のデータが取得できれば、それ以降の学習者の変化が確認できるかどうかを本研究では確かめることになる。本来であれば、早い段階で学習者の変化を検知することが望まれることは明らかで、今後の課題である。

第 4 回授業から 1 週間の学習ログデータを利用して主成分分析を行った結果、各主成分軸に対する係数は表 3 のようになる。第 1 主成分 (PC1) はイベントコンテキストおよびイベントが正の係数では大きく、コンポーネントが負の係数で影響が大きいことがわかる。一方で、ユーザと IP アドレスの影響が小さいことから、PC1 はユーザに影響されず学習コンテンツに依存する成分であると考えられる。第 2 主成分 (PC2) はユーザと IP アドレスの影響が大きく各ユーザに依存する成分であると考えられる。

表 3 各主成分軸に対する係数

	PC1	PC2	PC3
ユーザ	0.014	0.694	-0.625
イベントコンテキスト	0.590	0.011	0.320
コンポーネント	-0.662	-0.013	-0.029
イベント	0.462	-0.052	-0.430
IP アドレス	0.001	-0.718	-0.567

6. 異常値の検知

本研究では、主成分の値が正規分布すると仮定し、1 次元ホテリング理論に基づいて異常度を求め、1 % 水準で正規分布から外れるような異常度となる場合に異常値とする。正規分布という仮定および閾値について今後十分な検討が必要であると考えている。

その変化を考察するために第 4 回授業からの 1 週間以降の第 5 回から第 15 回までの学習ログデータを 1 週間毎に区切り、それぞれの期間中の主成分の値を求め、異常値をとった区間のデータを次に示す。

図 1, 図 2 図 3, 図 4, 図 5

7. まとめと今後に向けて

本研究では、LMS を利用し、初年次学生を対象に行った教育実践において、学習スタイルの調査を行い、学習スタイルの特徴ごとに分けたグループごとに学習ログについて考察した。

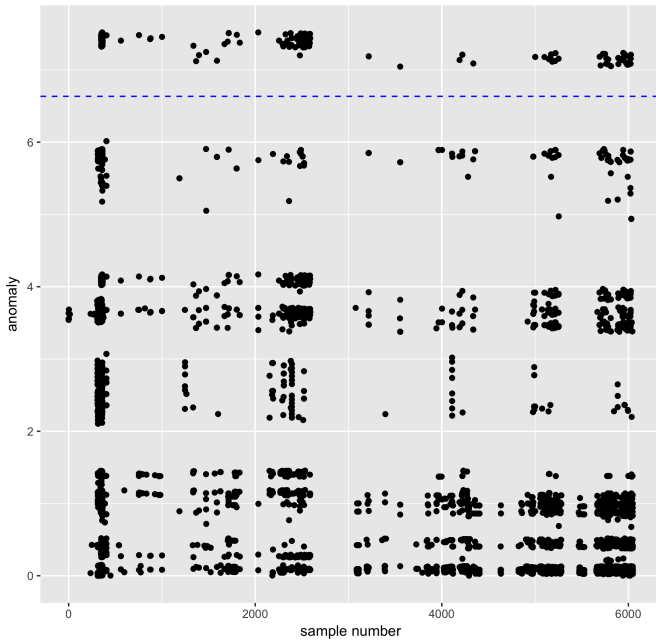


図 1 第 12 回授業から 1 週間の PC1 の時系列データ

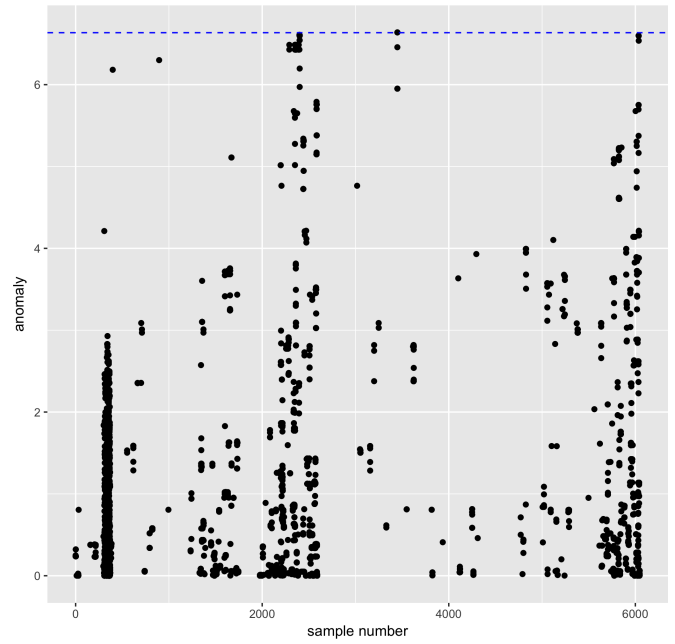


図 3 第 7 回授業から 1 週間の PC2 の時系列データ

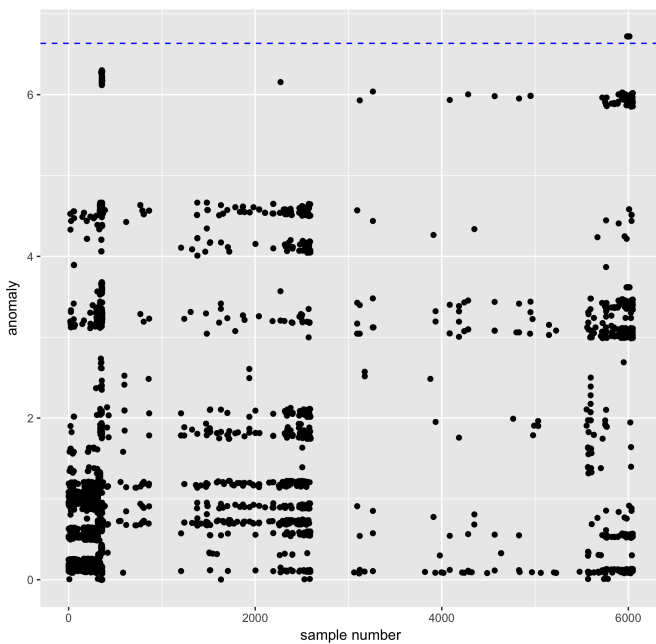


図 2 第 13 回授業から 1 週間の PC1 の時系列データ

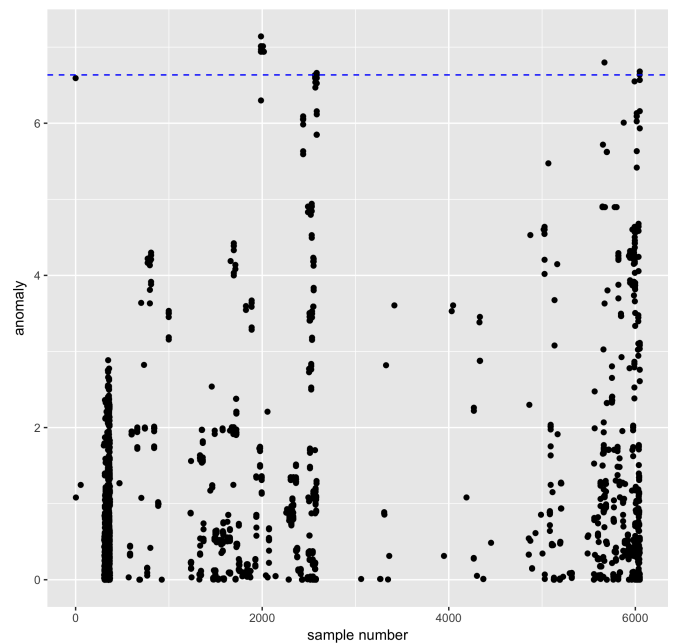


図 4 第 10 回授業から 1 週間の PC2 の時系列データ

実践した講義の中で行った Felder らの学習スタイル質問紙の回答をもとに、因子分析を行い、3つの因子について議論を進めた。今回、因子を3つとしたが、得られた結果の寄与率をみると十分な寄与率とは言えないものであると考えている。それぞれの因子は、Felder の学習スタイルモデルの指標を単独で含む形で現れており、Felder らの指標の有効性を確認することができた。

本研究では、学習ログと学習スタイルとの関係について調べるアプローチとして、学習スタイルごとに分類される

人の学習ログはどんな特徴をもっているか確認することとした。そこで、因子分析の結果である因子得点を用いて、クラスタ分析を行い、複数因子を示す学習者群が3つまたは2つに分類されることを確認し、各グループごとに学習ログを考察し、その特徴を確認した。今回のアプローチとは別に、学習ログをもとに分類を行い、その分類されたグループがどのような学習スタイルを示す傾向にあるか確認することも必要だと考えている。

今回の分析に際して、Moodle の標準ログを取得して

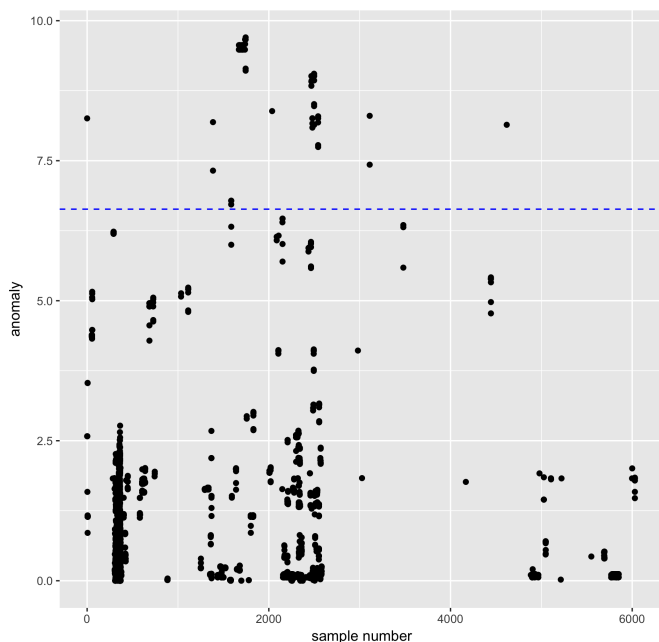


図5 第15回授業から1週間のPC2の時系列データ

行ったが、データクレンジングでは、コンポーネントなどの要素については、対応させた数値に意味を持たせてはなかった。将来的には、対応する数値に、学習コンテンツの順番や学習コンテンツとして利用される頻度の高さなどの情報を与える方法も検討しなければならない。

今後は、今回のデータで利用していない他の科目での学習ログの利用やLMSにはないログの利用など、より多面的に学習者の状況を把握する方法について考察する。

謝辞

本研究の一部は日本学術振興会の学術研究助成基金助成金（若手研究(B) 25750086）の支援により実施された。

参考文献

- [1] 竹生久美子, 辻靖彦: 学習履歴データから可視化する通信課程と通学課程の学習者特性, 日本教育工学会研究報告集, vol. 15, no. 1, pp. 119–126 (2015).
- [2] 中野裕司, 久保田真一郎, 松葉龍一, 戸田真志, 永井孝幸, 右田雅裕, 武藏泰雄, 喜多敏博, 杉谷賢一: 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, 電子情報通信学会技術研究報告. SC, サービスコンピューティング, vol. 114, no. 277, pp. 33–38 (2014).
- [3] 山川修: LMSと対面における学習者の行動特性の比較研究, 研究報告教育学習支援情報システム (CLE), vol. 2013, no. 7, pp. 1–4, (2013).
- [4] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, N. Taft: Structural Analysis of Network Traffic Flows., ACM SIGMETRICS Performance Evaluation Review, vol. 32, no. 1, pp. 61–72 (2004).
- [5] 井手剛: 入門 機械学習による異常検知 -Rによる実践ガイド-, コロナ社 (2015).

正誤表

原稿タイトル「LMSにおける学習ログを基にした学習者の振る舞い変化を検知する試み」

上記タイトルにて投稿させていただいた原稿について、著者の不注意により誤って完成前の原稿をアップロードしてしまいました。深くお詫び申し上げます。

下記に正誤表を記します。参考いただけますと幸いです。

ページ等	誤	正
4 ページ, 第 6 節「異常値の検知」, 10 行目	図 1, 図 2 図 3, 図 4, 図 5	<p>図 1 から図 5 のいずれの図も縦軸が異常度で横軸が授業日以降の 1 週間の時間である。図 1 は第 12 回授業から 1 週間の PC1 の異常度をプロットした図で、図 2 は第 13 回授業から 1 週間の異常度をプロットした図である。いずれも第 1 主成分において異常値を示した区間で、この他の区間で異常値は見られなかった。表 2 から第 12 回授業はテストと振り返りの回であり、イベントコンテキストやイベントに依存する第 1 主成分であるため、それまでの学習活動との相違が異常値として現れたと考えられる。</p> <p>図 3 は第 7 回授業から 1 週間の PC2 の異常度を、図 4 は第 10 回授業から 1 週間の PC2 の異常度を、図 5 は第 15 回授業から 1 週間の PC2 の異常度をプロットした図になる。いずれも第 2 主成分において異常値を含む区間で、この他の区間で異常値は見られなかった。第 7 回授業から 1 週間の PC2 が異常値を示すデータを確認したところ PC2 の値は -2.69 となっており、負の係数の影響が大きいと考えられる。表 3 の PC2 のうち負の値影響を及ぼす要素は IP アドレスであり、これにより異常値となったと考えられる。この異常値を示した際の IP アドレスを確認したところ、全レコード 86819 件のうち 396 件のレコード数で、アクセスの少ないネットワークに属する IP アドレスであったために、異常値を示したと考えられる。第 10 回授業から 1 週間の PC2 が異常値を示すデータを確認したところ PC2 の値の平均は -2.75 (標準偏差 0.03) であり、PC2 のうち負の値に影響を及ぼす要素である IP アドレスが大きく影響を与えていると考えられる。この異常値を示した際の IP アドレスを確認したところ、2 つの特定のネットワークに属する IP アドレスで、全レコード 86819 件のうち 2849 件、1955 件のレコード数であった。先の考察とは異なり、いずれもレコード数が多かったため、PC2 に影響が大きい要素であるユーザを確認したところ、特定のユーザが特定のネットワークからアクセスしていることがわかった。第 10 回授業から 1 週間の間にこれら特定のユーザに何らかの変化があった可能性があり、彼らの学習ログをさらに詳細に確認する必要があると思われる。第 15 回授業から 1 週間の PC2 が異常値を示すデータは 56 レコードもある。基本的には第 15 回授業は最終回であり、それ以降はテスト期間にも入るためアクセスの仕方がそれまでと大きく異なったためにこのようになったと考えている。</p>

ページ等	誤	正
4 ページ, 第 7 節「まとめと今後に向けて」, 全文	第 7 節「まとめと今後に向けて」の全文	<p>本研究では、LMS に記録される学習ログをもとに、ある期間の時系列データで構成される行列に対して主成分分析を適用し、その他の区間における主成分の値を考察した。考察には、1次元ホテリング理論に基づいた異常度を求め、正規分布を仮定するときの1判定を行い、異常値を検出した。この手法により学習ログデータのインプットから主成分の値を求め、異常度を算出することができ、異常度の判定を行う過程がシステム化できると考えられる。一方で、異常値データから学習者の変化を判断するまでには至っておらず、異常判定における閾値の検討、異常データから判断される学習者の環境変化についてより詳細な考察が必要であると考えられる。また、今回の検討では、15回授業のうち第4回授業から1週間の学習ログデータをもとに他の区間の異常度を考察した。このため第5週から異常度を監視する運用となり、早期に学習者の変化を検知することができない。これについては前年度学習ログデータを用いるなど早期に学習者の変化を考察できる工夫が必要である。</p> <p>今回の研究では簡単のために1次元ホテリング理論により異常値を検出したが、多変量のホテリング理論により異常な値を検知すべきである。また、ホテリング理論は正規分布を前提としており、学習ログデータが必ずしも正規分布に従うとは限らないため、正規分布によらない検知手法について検討が必要である。</p> <p>今回、Moodle の標準ログを取得して分析を行ったが、データクレンジングでは、コンポーネントなどの要素に対応する数値に意味を持たせてはいない。将来的には、学習コンテンツの順番や学習コンテンツとして利用される頻度の高さなどの情報を与える方法も検討が必要と考えている。また、今後は、他の科目で同様の手法により異常検知として学習者の変化を検知できないか検証が必要であると考えている。また、学習ログの利用に限らず LMS には記録されないログの利用など、より多面的に学習者の状況を把握する方法についても検討が必要である。</p>
6 ページ, 謝辞, 全文	謝辞の全文	本研究の一部は JSPS 科研費 16K01070 の助成を受けたものです。