Extracting Teaching Activities from E-book Logs Using Time-Series Shapelets

Daiki Suehiro[1,a)] Yuta Taniguchi[1,b)] Atsushi Shimada[1,c)] Hiroaki Ogata[2,d)]

1. Introduction

Teaching analytics for face-to-face classes have been researched for many years. In particular, determining the teaching activity from data is highly important. In many of the related works, the teaching activities or hypotheses are manually designed to be easily understandable. However, the results depend on the knowledge of domain experts and the preparations for considering useful variables of teaching activities are very time-consuming. The data-driven approaches using digital devices such as wearable sensors achieve moderate success [4]. However, teaching analytics using IT and digital tools are still at an early stage, and the cost of using such digital tools for large-scale verification is high.

In this paper, we take a novel approach for practicing teaching analytics in face-to-face classes as following: First, to collect the data efficiently (without optional devices or tools) in a face-to-face class, we use e-books with logs. Second, to extract the teaching activity, we consider the page transition as time-series data, and provide a novel analysis scheme using *time-series shapelets* [7] (we call *shapelets* for short) and machine learning techniques [6]. Shapelets-based time-series analysis can provide us with data-driven and interpretive variables from time-series data. Finally, the results show that our analysis succeeded in extracting the teaching activities accurately.

2. Data Collection

In some classes at Kyushu university in Japan, an e-book system is employed. The teachers show slides and students read the slides on their terminals. The e-book system allows us to efficiently collect various logs on learners or teachers, such as the page numbers users are viewing, memos, and other learning actions and along with the time (e.g., Table 1). In this study, we collected e-book logs from seven "Information science" courses that were taught by four teachers during first semester of 2016. All of the materials for teaching are shared in the courses. We provide the details of the seven courses in Table 2. The materials contain 32 sets of slides based on the course contents. Each teacher may choose several sets of slides for classroom use. The number of students in each course are also described in Table 2.

Table. 1. Example of e-book logs

User	Slides Name	Page	Time	···
Teacher 1	Statistics 1	3	16/04/30 08:50:09	···
Student 1	Statistics 1	4	16/04/30 08:50:15	···
⋮	⋮	⋮	⋮	⋮
Student 9	English 2	22	16/05/23 13:34:30	···
Teacher 7	English 2	25	16/05/23 13:34:32	···

Table. 2. Details of seven "Information science" courses

Course ID	Teacher	# students	# lectures
c1	A	26	15
c2		46	
c3	B	109	15
c4	C	165	8
c5		239	
c6	D	111	15
c7		43	

This paper focuses on the teachers' and the students' page-view logs. We believe that we can use these logs to observe some teaching activities that relate to controlling the class (e.g., teaching speed that considers students' learning capabilities). First, using time-stamp data as in Figure 1, we created a time-series data of page-view num-

[1] Kyushu University
[2] Kyoto University
[a)] suehiro@ait.kyushu-u.ac.jp
[b)] taniguchi@artsci.kyushu-u.ac.jp
[c)] atsushi@ait.kyushu-u.ac.jp
[d)] hiroaki.ogata@gmail.com

bers for each student and teacher. To create contiguous time-series data, we simply fill the last-viewed page number for each time of non-stamped time. Second, in one of the main points of this paper, we create secondary data that is defined by the differences between the pages that the teacher is showing and those that each student is viewing, which we call DPTS. Moreover, to observe teaching activities that can provide information on the expertness of each teacher, a DPTS is created for 32 content, yielding 32 DPTSs in total. The value of each data point in the time series in a DPTS is defined as the page number of a student minus that of the teacher at a given time in seconds. Hence, the value of DPTS at each time indicates following properties:

- If the value is positive, then the student is ahead of the teacher.
- If the value is negative, then the student is behind the teacher or reviewing.
- If the value is "0", then student is following the teacher.

We think that the page number itself does not have a meaning, but the value of the difference might indicate whether students are able to follow the teacher.

2.1 Problem Setting

For the rDPTS of each content, we set a (multi-class) time-series classification problem. More precisely, given time-series data rDPTS = $(t_1, y_1), \ldots, (t_N, y_N)$, we wish to find a rule that detects a teacher of unlabeled data, where t_i is a time-series example and y_i is the label in corresponding to the teacher, N is the number of the students for each teacher. It is important to note that detecting teachers is not our true objective but rather is a procedure towards extracting the teaching activities of the experts. Our true goal is to find teaching activities that allow us to accurately discriminate between the teachers.

2.2 Our Data Mining Method

There are many methods of resolving the time-series classification problem [5, 8]. In practice, some statistical values of a time window (e.g., mean, max) are used as the variables because they are easy for domain experts to understand their meaning. However, such statistical values sometimes omit several important features, such as small changes of value over a short-period. Therefore, in order to realize our objective, we employ shapelets-based time-series classification methods [8]. Informally speaking, shapelets are a set of "short" time series, which indicates characteristic patterns of the partial transitions, used to detect the label of a time series by whether a given shapelet partly matches or mismatches for a time series (illustrated in Figure 1). Moreover, as many algorithms

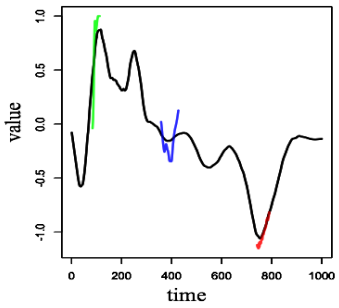

Figure 1. The black line is an original time series, the blue, green, and red line are the shapelets. The feature values of the time series are given by minimum distances between any sub-sequence of the time series and each shapelet.

to find shapelets retain the full information of the original time series, we may find more precise features (characteristic transition-patterns of each teacher) using this process than with other methods. With such a property, shapelets may find discriminative and interpretive patterns in time series.

Here we show more a formal definition. Let $(t_1, \ldots, t_N)$ be the sample data of a time series, where N is the number ofe xamples. Given a set of K shapelets $\mathcal{S} = \{s_1, s_2, \ldots, s_K\}$, the feature values $(x_{i,1}, \ldots, x_{i,K})$ of a time-series t_i are defined as the minimum Euclidean distance between shapelet s_j and any sub-sequences of t_i [2];

$$x_{i,j} = \min_l (t_i[l : l + |s_j| - 1] - s_j)^2, \qquad (1)$$

where $t[a : b]$ sub-sequence $t_a, t_{a+1}, \ldots, t_b$. In this paper, we find suitable shapelets using the method proposed by Suehiro et al. [7], which finds efficiently and accurately finds good shapelets. The method uses sparse SVM (Support Vector Machines), which is one of the most popular machine-learning algorithms for binary classification problems[*1]. However, in this paper, we use sparse *AUC-maximizing* SVM [6], which we abbreviate to *SASVM*. SASVM finds the linear hypothesis f that maximizes AUC (Area Under the ROC Curve) [1]. Informally speaking, AUC, as known as c-statistic, is a validation measure that is often used for heavily biased data. More formally, the AUC of any hypothesis h is defined as follows;

[*1] "Sparse" means that the method induces good feature selection. Hence, we can obtain well-selected teaching activities.

$$\mathrm{AUC}(h) = \frac{1}{pn}\sum_{i=1}^{p}\sum_{j=1}^{n} I\left(h(\boldsymbol{x}_i) - h(\boldsymbol{x}'_j) \geq 0\right), \qquad (2)$$

where $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)$ is an example of target label and $(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_n)$ are examples of the other label, $I(\cdot)$ is the indicator function that returns "1" if $(\cdot)$ is true and "0" otherwise. In all pairs of the target examples and the other examples, the more closely the h values match the target examples, the higher the AUC score (close to 1.0). Table 2 shows that the number of students in each course is biased. Furthermore, a multi-class classification problem is usually decomposed into one-vs-all binary class classification sub-problems (e.g., the examples of teacher A are labeled as positive and the examples of the other teachers are labeled as negative). However, for example, if we try to solve a one-vs-all sub-problem in which the examples of teacher A are labeled as positive, the number of positive examples is 72 and the number of negative examples is 667, we encounter highly biased data. For this reason, we want to maximize AUC but not simple classification accuracy. SASVM inputs training examples that are represented as feature vectors and returns a linear function that gives higher values to positive examples than negative examples:

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} = w_1 x_1 + \cdots + w_K x_K, \qquad (3)$$

where $\boldsymbol{w}$ is a weight vector that holds $\sum_{j=1}^{K} |w_j| = 1$ and $\cdot$ indicates the inner product. Each element of w represents the contributing rate for each shapelet because the feature value x_j is given by shapelet s_j.

After we solve the multi-class classification problem for each content, we obtain the useful shapelets $\mathcal{S}$ and the contribution $\boldsymbol{w}$ that discriminates between the teachers. Moreover, f has a property that ranks the time series of the target teacher higher than the other teachers. In other words, the time series with a higher value might be the typical time series of the target teachers.

3. Results

3.1 Teacher-prediction Accuracy

For the seven "information science" courses of the first semester of the 2016 school year, we prepared a set of DPTS by content. In this paper, we show the result of prediction accuracy measured by AUC, for three contents of all 32 contents. Table 3 shows the AUCs for each content evaluated using 5-fold cross-validation. We can see that our teacher-predicting method achieves high AUC scores[*2]. Therefore, it is clear that the shapelets extracted by our method provide some discriminative power for predicting each target teacher. In other words, the extracted shapelets represent some of the teaching activities or the styles of each teacher.

[*2] AUC is one of the standard measure for classification accuracy. It is said that a hypothesis is considered excellent discrimination if the AUC is more than 0.8 [3].

Table. 3. AUCs for DPTS by each content

Contents ID	AUCs			
	Teacher A	Teacher B	Teacher C	Teacher D
M1	0.872	0.906	0.900	0.889
M2	0.894	0.916	0.834	0.856
M3	0.890	0.871	0.834	0.860

3.2 Observing Teaching Activities via Shapelet Analysis

To observe good teaching activities in extracted shapelets, in this paper, we focus on the specific content "M2" and teacher C. In fact, the content of "M2" is "image recognition" and teacher C is an expert in image recognition. Then, we consider obtaining some good teaching activities by observing the shapelets and time series of teacher C for "M2." Figure 2 exemplifies the three representative results for teacher C. The legend describes the importance-score of each shapelet (scored by [6]), and the shapelets shown are the most and second most important characteristic pattern. Note that the shapelets with negative importance (e.g., the red shapelet as seen in Figure 2) represent the teaching activities of teacher C, but the shapelets with positive weight (e.g., the light-blue shapelet) represent the teaching activities of the other teachers. Thus, when we want to know the teaching activities of a targeted teacher, we only have to focus on the shapelets with negative weight. So, by observing Figure 2,

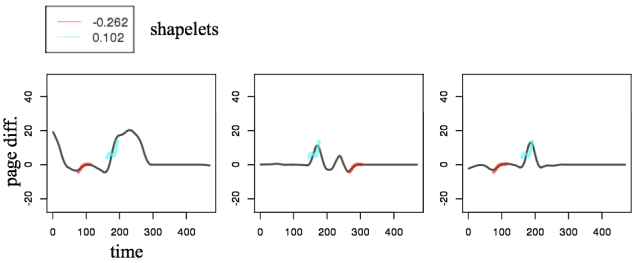

Figure 2. Three representative time-series of DPTS for teacher C (black line). The colored lines are highly important shapelets of teacher C. The legend means the weight of each shapelet. 1 time-unit is 10 seconds.

we discern the following things:

(1) The three typical students seem to follow the teacher

C because the page-differences basically remain at zero.

(2) The shapelet described by the red line seems to constitute a teaching activity by teacher C that controls the students to return to the current slide.

(3) The activity should be performed at approximately 90–110 or 280–300 (the times the red shapelet matches).

We can discern the details of the activities by checking the slides or videos used during the "activated" time. Indeed, we checked the slides used during the time, then we found that the slides explain a mathematical way of thinking and the students tend to be behind the teacher during the slides[*3].

4. Conclusion and Future Work

In this paper, we proposed a novel teaching analytics for use in face-to-face classes that involve the use of e-books. We focused on the difference between the page that the teacher was showing in the class at a given time and the pages that a given student was viewing at that time on their terminal. Second, we set a time-series classification problem, and we solve the problem using time-series shapelets and machine-learning techniques. The results show that some extracted shapelets are discriminative and interpretable, and suggests the possibility of efficiently discovering teaching activities.

References

[1] Bradley, A. P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition*, Vol. 30, pp. 1145–1159 (1997).

[2] Hills, J., Lines, J., Baranauskas, E., Mapp, J. and Bagnall, A.: Classification of Time Series by Shapelet Transformation, *Data Mining and Knowlege Discovery*, Vol. 28, No. 4, pp. 851–881 (2014).

[3] Hosmer, D. W. and Lemeshow, S.: *Applied logistic regression*, Wiley series in probability and statistics (2000).

[4] Prieto, L. P., Sharma, K., Dillenbourg, P. and Jesús, M.: Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors, *Proceedings of LAK '16*, pp. 148–157 (2016).

[5] Senin, P. and Malinchik, S.: SAX-VSM: Interpretable Time Series Classification Using SAX and Vector Space Model, *2013 IEEE 13th International Conference on Data Mining*, pp. 1175–1180 (2013).

[6] Suehiro, D., Hatano, K. and Takimoto, E.: Approximate Reduction from AUC Maximization to 1-norm Soft Margin Optimization, *Proceedings of ALT '11*, pp. 324–337 (2011).

[7] Suehiro, D., Kuwahara, K., Hatano, K. and Takimoto, E.: Time Series Classification Based on Random Shapelets, NIPS 2016 Time Series Workshop.

[8] Ye, L. and Keogh, E.: Time Series Shapelets: A New Primitive for Data Mining, *Proceedings of KDD '09*, ACM, pp. 947–956 (2009).

[*3] Unfortunately, we did not record the videos of the classes, however, teacher C told us to the teaching activity that "slowed down the teaching during the slides and instructed the students to lay weight on intuitive understanding."