Face-to-Face Teaching Analytics: Extracting Teaching Activities from E-book Logs via Time-Series Analysis

Suehiro, Daiki
Faculty of Arts and Science, Kyushu University

Taniguchi, Yuta
Faculty of Arts and Science, Kyushu University

Shimada, Atsushi
Faculty of Arts and Science, Kyushu University

Ogata, Hiroaki
Faculty of Arts and Science, Kyushu University

https://hdl.handle.net/2324/4068613

出版情報：2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT). 2017, pp.267-268, 2017-08-08. IEEE
バージョン：

Face-to-Face Teaching Analytics: Extracting Teaching Activities from E-book Logs via Time-Series Analysis

Daiki Suehiro*, Yuta Taniguchi*, Atsushi Shimada*, Hiroaki Ogata*
* Faculty of Arts and Science, Kyushu University, Motooka, Nishi-ku, Fukuoka, 819-0395, Japan,
Email:{suehiro, taniguchi, atsushi, ogata}@artsci.kyushu-u.ac.jp

***Abstract*—Teaching analytics have been studied for many years with the aim of analyzing and improving the teaching process. To discover and amass teaching knowledge efficiently, we must extract the various teaching activities from educational data. However, data cannot be collected in face-to-face classes as efficiently as through online learning. In this paper, through the use of e-book logs, we describe a method of efficiently collecting and analyzing educational data. Moreover, we describe a novel approach to practicing teaching analytics in face-to-face classes; one which enable us to extract the teaching activity efficiently and accurately.**

Keywords*—*Teaching Analytics; E-book; Time-Series Analysis; Data Mining;

I. Introduction

Various learning logs can be collected using digital-learning tools, and these can be analyzed as a means of improving the learning and teaching process. The development of learning analytics has been largely linked to the development of online learning, which allows the efficient collection of data on learners or teachers. However, in a face-to-face class, such data cannot be collected in the same manner, and in a blended class (e.g., a face-to-face class using e-books), data on the online learning of learners but not teachers can be analyzed.

A. Related Work

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 one attempt to address this, Aitkin et al. proposed the statistical modeling of data as a means of representing various discriminative teaching activities [1]. The various knowledges of the teaching activities that these researchers discerned in their study may allow teachers to complement each other. However, the cost of collecting the data required to extract the teaching activities is very high in terms of time and effort. Recent developments in information technology and digital tools, have made it possible to collect various data for teaching analytics. Although some basic studies exist that involved the use of data collection and analytics to extract teaching activities [2], [3], the variables based on the teaching activities are usually manually determined by domain experts. Although the produced teaching activities or hypotheses are easily understandable, the results depend on the knowledge of domain experts. Furthermore, the preparations for considering such variables of teaching activities are very time-consuming. Prieto et al. attempted to determine teaching activity by using automatically extract orchestration graphs [4], and the results showed that the extracted orchestration graphs represent teaching activities in practice, and are useful for planning the next class. The researchers also took an impressive first step towards data-driven teaching analytics by employing various wearable sensors. 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. Moreover, unfortunately, the wearable sensors or and other direct monitoring tools are only useful for analysis, but not for learning or teaching.

B. Our Contribution

In this paper, as written below, we take a novel approach for practicing teaching analytics in face-to-face classes; one which enable us to extract the teaching activity efficiently and accurately.

First, to collect the data efficiently (without optional devices or tools) in a face-to-face class, we use e-books with logs. An e-book system is not only useful for providing course materials but is also useful for collecting learning logs. While e-book logs have mainly been used for analyzing learning styles (e.g., [5]–[7]), in this study, we focus on logs that concern the slides that a teacher shows in class and logs that show what students reading. Teachers usually do not retain the same logs as students (e.g., memos or highlights are usually not used during class), however, for example, the page transitions may indicate the teaching activity orchestrating the class.

Second, to extract the teaching activity accurately and efficiently, we consider the page transition as time-series data, and provide a novel analysis scheme using *time-series shapelets* [8]–[11] (we call *shapelets* for short). Shapelets-based time-series classification methods can provide us with data-driven and interpretive variables (to be described later) from time-series data. In contrast to several methods based on manual coding variables (e.g. statistical values of time windows), shapelets-based methods do not require explicit variables and automatically extract the discriminative variables from data. Hence, observing the extracted shapelets from teachers help us to efficiently find important teaching activities.

Finally, the results show that our analysis succeeded in extracting the teaching activities accurately. Further, we observe that time-series analysis based on shapelets gives us concrete and interpretive teaching activity.

II. Data Collection

A. Using E-book Logs

In some classes at our university, an e-book system is employed. The teachers show slides and students read the slides on their terminals, as shown in Figure 1.

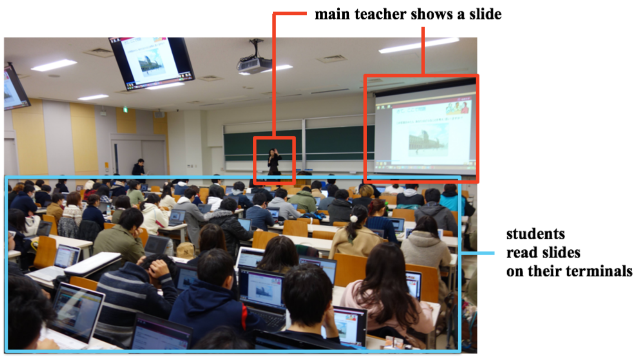

Figure 1. A class using e-books.

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 I).

TABLE I. Example of e-book logs

User	Slides Name	Page	Time	···
Teacher 1A	Statistics 1	3	16/04/30 08:50:09	···
Student 1B	Statistics 1	4	16/04/30 08:50:15	···
Student 1C	Statistics 1	3	16/04/30 08:50:20	···
Student 1B	Statistics 1	5	16/04/30 08:50:21	···
⋮	⋮	⋮	⋮	⋮
Student 2C	English Read. 2	22	16/05/23 13:34:30	···
Teacher 2A	English Read. 2	25	16/05/23 13:34:32	···

In this study, we collected e-book logs from seven "Information science" courses that were taught by four teachers during first semester of 2016[1]. All of the materials for teaching are shared in the courses. We provide the details of the seven courses in Table II. The materials contain 32 sets of slides based on the course contents. Each teacher may choose several sets of slides for classroom use. In each lecture, the teacher teaches for 90 minutes; the five courses taught by teachers A, B, D consists of a total of 15 lectures in the semester, and the remaining two courses, taught by teacher C, consists of a total of eight lectures in the semester. The number of students in each course are also described in Table II. Although the disciplines of the students are biased by the courses, we performed our analysis without referring to any influence the bias may create.

B. Time-Series Data on Page Transitions

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).

[1]In fact, there are nine courses taught by five teachers in total, but one of the teachers did not use the e-book system in two courses.

TABLE II. 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	

First, using time-stamp data as in Figure I, we created a time-series data of page-view numbers 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. For example, in Figure I, "student 1B" did not stamp from 08:16 to 08:20; hence, we fill the last value, "4" at 08:15 and, thus, we can create a contiguous time-series $(4, 4, 4, 4, 4, 4, 5)$ from 08:15 to 08:21.

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 (e.g., image recognition, information theory, robotics), a DPTS is created for each 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 the page number of the teacher at a given time in seconds. Hence, the value of DPTS at time p indicates following properties:

- If the value is positive, then student i is ahead of the teacher.
- If the value is negative, then a student i is behind the teacher or reviewing.
- If the value is "0", then student i is following the teacher[2].

The motivation for using DPTS is as follows; first, 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; second, in our prior analysis, we observed that all of the distributions of DPTS by the four teachers were roughly Gaussian with zero mean as shown in Figure 2. These facts encouraged us to conduct a smooth analysis across the various-condition classes.

C. Data Conversion

Our time-series data set is very large, since by simple arithmetic tells us that each student has a time series of length $60(\text{sec}) \times 90(\text{min}) \times 15(\text{lectures}) = 81,000$ data points. The time complexities of almost all of the methods for analyzing

[2]It is important to note that we actually define the value to be "0" not only when the teacher's page number is equal to the student's page number, but also in certain cases when the student is not viewing the same content as the teacher. This deficiency in our method is the result of a desperate resort to define the values of DPTS contiguously. Although it is an important problem for future work, it does not greatly impact this paper because such examples were very rare in our data.

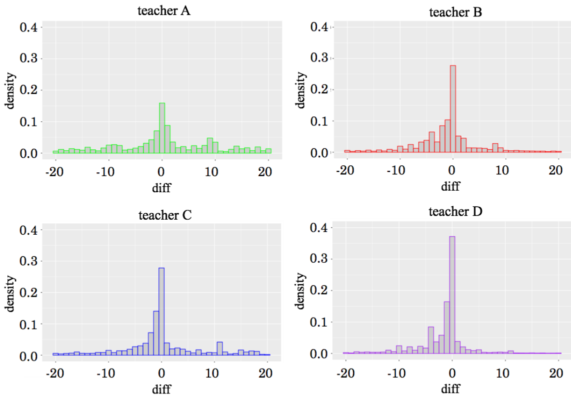

Figure 2. The distributions of page-view differences between teacher and the students in "information science" courses.

time series depend on the length of the time series [8], [12]. Although we want to decrease the frequency of time-series data, doing so puts us at risk of losing important information in the time series because the values are discrete. To avoid this problem, we use a smoothing technique for each time series[3]. Smoothed values are continuous and retain more of a time series' information after a reduction in frequency than the original does in its discrete counterpart. Moreover, smoothing also acts as a low-pass filter in denoising the time series. The DPTS after smoothing and frequency reduction by 10 seconds have been applied is referred to as rDPTS.

III. Data Mining Strategy

A. 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.

B. Our Data Mining Method

There are many methods of resolving the time-series classification problem [8], [12]. 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 3). Moreover, as many algorithms

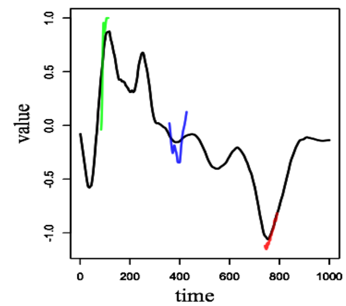

Figure 3. The black line is an original time series, the blue, green, and red line are the shapelets. The feature values of the time series are given by minimum distances between any sub-sequence of the time series and each shapelet.

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 [14];

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

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. [11], 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[4]. However, in this paper, we use sparse *AUC-maximizing* SVM [15], which we abbreviate to *SASVM*. SASVM finds the linear hypothesis f that maximizes AUC (Area Under the ROC Curve) [16]. 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;

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

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 II 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

[3] We use Friedman's Super Smoother [13] in our analysis.

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

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:

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

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.

IV. Results

A. Teacher-prediction Accuracy

For the seven "information science" courses of the first semester of the 2016 school year, we prepared a set of rDPTS by content. The logs stated that 18 of 32 contents were taught by four teachers (each teacher can design their own courses rather freely). We searched for shapelets from short time series of lengths of 300(sec.) $\leq |s_j| \leq$ 900(sec.), which meant that the "activated" or "orchestrated" time ranges we expected were returned. We excluded extremely short examples (less than 900(sec.)). Table III shows the AUCs for each content evaluated using 5-fold cross-validation. For four of 18 contents, we could not analyze because the sets of slides were used by only one teacher or because SASVM was unable to find any solution[5].

We can see that our teacher-predicting method achieves high AUC scores[6]. 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.

B. Observing Teaching Activities via Shapelet Analysis

To observe good teaching activities in extracted shapelets, in this paper, we focus on the specific content "12J1W" and teacher C. In fact, the content of "12J1W" is "image

[5]We set the parameter of SASVM so that the output function achieves AUC = 0.6 at minimum on training data.

[6]It is said that a hypothesis is considered excellent discrimination if the AUC is more than 0.8 [17].

TABLE III. AUCs for rDPTS by each content

Contents ID	AUCs			
	Teacher A	Teacher B	Teacher C	Teacher D
12IYZ	0.872	0.906	0.900	0.889
12IZB	0.901	-	0.906	-
12IZE	0.928	0.927	-	0.922
12IZH	0.883	0.882	-	0.890
12IZK	0.931	0.963	-	0.895
12IZT	0.807	-	-	0.792
12IZW	0.852	0.694	-	0.661
12J15	-	0.963	-	0.931
12J1B	0.875	0.959	-	0.931
12J1Q	0.852	0.924	-	-
12J1T	-	-	0.861	0.867
12J1W	0.894	0.916	0.834	0.856
12J1Z	-	0.923	0.857	0.869
16XSW	0.890	0.871	0.834	0.860

recognition" and teacher C is an expert in image recognition[7]. Then, we consider obtaining some good teaching activities by observing the shapelets and time series of teacher C for "12J1W." First, using the rDPTS of "12J1W," we create the prediction functions f of teacher C. Next, we calculate the values of rDPTS of "12J1W" using the prediction functions. Figure 4 exemplifies the three typical time-series of teacher C, which are evaluated as the top-3 of all time-series. The legend describes the weight of each shapelet, and the shapelets shown are the most and second most important for representing the activities. It is important to note that the shapelets with negative weight (e.g., the red shapelet as seen in Figure 4) 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. This is why the value of f for teacher C becomes larger if shapelets with negative weight match the time series. On the other hand, if shapelets with positive weight match the time series, the value of f for teacher C becomes smaller. 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

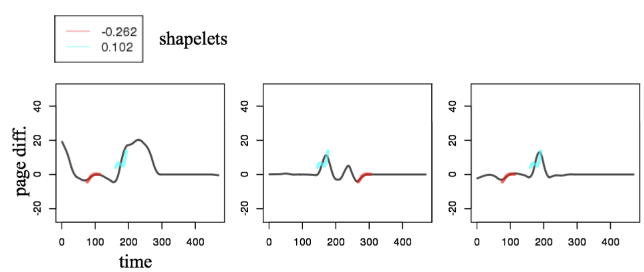

Figure 4. Top-3 typical time-series of rDPTS for teacher C (black line). The colored lines are highly important shapelets of teacher C. The legend means the weight of each shapelet. 1 time-unit is 10 seconds.

Figure 4, we discern the following things:

1) The three typical students seem to follow the teacher C because the page-differences basically remain at zero.

[7]Of course, teacher A, B, and D are experts in information science but their domains do not concern image recognition.

2) The shapelet described by the red line seems to constitute an orchestrated teaching activity by teacher C that induces 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 teaching 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. 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." It is just one example of our results, but such knowledge can be used as a useful feedback for the other teachers.

V. Discussion

The above results were obtained from e-books through the use of logs and machine-learning techniques and without the use of optional devices or manually employed variables. It is important to note that we can easily apply the other conditions of the courses (e.g., this method is applicable more than just information science). That is why the concept of DPTS allows us to determine the teaching activities of the experts, as the difference value is a common measure in several classes that use slides. In contrast to the variables based on statistical values, shapelets enable us to understand more precise teaching activities of each teacher through matching the original data. Although our methods are verified for the same courses, it is an effective way to begin teaching analytics using DPTS and shapelets. In this paper, we gave a teaching activity concerning orchestration as an example, but it is possible to employ this method to other various teaching activities.

However, considerable work remains to put this methodology into practice. In order to enhance the utility of our method, we must ensure that the feedback is useful for the other teachers. To advance to the next step, we plan to collect video data in classes and provide specific teaching activities or, alternatively, clips from the videos extracted by the shapelets that demonstrate good teaching.

VI. 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 succeeded to extract teaching activities by following two key ideas. First, 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 his or her 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 this suggests the possiblity of efficiently discovering teaching activities. Although this paper yielded fundamental results, future work should verify the data across various courses. Furthermore, for the further development of teaching analytics, we must analyze the procedure in combination with the other effective procedures, such as wearable sensors [4] or teaching movies.

References

[1] J. H. Murray Aitkin, Dorothy Anderson, "Statistical modelling of data on teaching styles," *Journal of the Royal Statistical Society. Series A (General)*, vol. 144, no. 4, pp. 419–461, 1981.

[2] B. Xu and M. Recker, "Teaching analytics: A clustering and triangulation study of digital library user data," *Educational Technology & Society*, vol. 15, no. 3, pp. 103–115, 2012.

[3] F. Domínguez, K. Chiluiza, V. Echeverria, and X. Ochoa, "Multimodal selfies: Designing a multimodal recording device for students in traditional classrooms," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI '15. New York, NY, USA: ACM, 2015, pp. 567–574.

[4] L. P. Prieto, K. Sharma, P. Dillenbourg, and M. Jesús, "Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors," in *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, 2016, pp. 148–157.

[5] A. Shimada, F. Okubo, and H. Ogata, "Browsing-pattern mining from e-book logs with non-negative matrix factorization," in *the 9th International Conference on Educational Data Mining*, 6 2016, pp. 636–637.

[6] K. Mouri, F. Okubo, A. Shimada, and H. Ogata, "Bayesian network for predicting students' final grade using e-book logs in university education," in *IEEE International Conference on Advanced Learning Technologies (ICALT2016)*, 7 2016, pp. 85–89.

[7] J. Wang, H. Ogata, and A. Shimada, "A knowledge comparison environment for supporting meaningful learning of e-book users," *Systems*, vol. 4, no. 2, p. 21, 2016.

[8] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 947–956.

[9] E. J. Keogh and T. Rakthanmanon, "Fast shapelets: A scalable algorithm for discovering time series shapelets," in *Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA.*, 2013, pp. 668–676.

[10] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, "Learning time-series shapelets," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: ACM, 2014, pp. 392–401.

[11] D. Suehiro, K. Kuwahara, K. Hatano, and E. Takimoto, "Time series classification based on random shapelets," Neural Information Processing Systems (NIPS) 2016 Time Series Workshop, https://sites.google.com/site/nipsts2016/NIPS_2016_TSW_paper_42.pdf?attredirects=0&d=1.

[12] P. Senin and S. Malinchik, "Sax-vsm: Interpretable time series classification using sax and vector space model," in *2013 IEEE 13th International Conference on Data Mining*, Dec 2013, pp. 1175–1180.

[13] H. F. Friedman, "A variable span scatterplot smoother," Laboratory for Computational Statistics, Tech. Rep., 1984.

[14] J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall, "Classification of time series by shapelet transformation," *Data Mining and Knowlege Discovery*, vol. 28, no. 4, pp. 851–881, Jul. 2014.

[15] D. Suehiro, K. Hatano, and E. Takimoto, "Approximate reduction from AUC maximization to 1-norm soft margin optimization," in *Proceedings of the 22nd International Conference on Algorithmic Learning Theory*, 2011, pp. 324–337.

[16] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, pp. 1145–1159, 1997.

[17] D. W. Hosmer and S. Lemeshow, *Applied logistic regression*, ser. Wiley series in probability and statistics. New York, Chichester, Weinheim: John Wiley & Sons, Inc. A Wiley-Interscience Publication, 2000.