Mining User Trajectories in Electronic Text Books

Ahcène Boubekki
Leuphana University of Lüneburg
boubekki@leuphana.de

Shailee Jain*
The University of Texas at Austin
shailee@cs.utexas.edu

Ulf Brefeld
Leuphana University of Lüneburg
brefeld@leuphana.de

ABSTRACT

Analyzing user behavior in electronic textbooks offers appealing insights into how pupils interact with the book and internalize the content. Using these insights may help to personalize the book, e.g., to support users with special educational needs. Conventional approaches often focus on atomic, user-triggered events like clicks or scrolls. In this paper, we propose to view all ongoing sessions in a classroom simultaneously and cast the problem as a multi-user problem over space and time. We devise two distance measures to compare the navigation behavior of pupils in different dimensions. Empirically, we observe that our metrics lead to interpretable clusters and serve as performance indicators.

Keywords

Sequential clustering, behavioral analyses, spatio-temporal trajectories

1. INTRODUCTION

The advent of information and communication technologies (ICT) in education has given teachers and educators a magic box full of possibilities [21]. Learning can now be made interactive and engaging for students. The digitization movement has further expanded with MOOCs [18, 10] that provide easy access to extensive and high quality courses online. Situated in-between traditional classrooms and online MOOCs, are electronic textbooks.

E-books incorporate the benefits of both traditionally printed copies and online media. Their structure closely resembles real books, thus rendering a look and feel familiar to students and teachers alike. Additionally, they often include interactive objects (hyperlinks, text boxes for comments) and interlinked media types to enhance the learning experience and delineate content better. Teachers can easily integrate the new technology in their classroom as they offer the full bandwidth, from traditional reading to creative exploring tasks. In addition, electronic books are usually designed to be self-contained and prevent the risk of students being lost in large amounts of content.

This work is part of a project that aims to evaluate the effectiveness of electronic textbooks as learning tools. Our study is based on a collaboration with psychologists and educators. The premise is an electronic text book called the 'mBook' [27, 28] that has been written and developed by a team of history teachers and didacticians. It is being deployed in the German-speaking community of Belgium since 2013.

The mBook records all user-triggered events like clicks and scroll operations such that every session can be replayed entirely. Quantities like the visible content at each timestamp can be derived straight forwardly from this data. We aim to use this information to identify usage patterns in the behavior of the pupils and analyze how they reflect on their performances.

Extracting patterns from log files has been a widely researched topic. Usual techniques range from Behavioral Sequential Analysis [2, 31, 9] to mixtures of Markov chains [6, 7, 15, 8]. However, all these methods are based on event transitions and do not consider historical events or past data. Higher-order Markov chains could possibly handle longer sequences that condition these transitions. Nevertheless, the computation becomes rapidly intractable.

The approach we choose here is to literally extend the navigation metaphor and build a structure to handle sessions as is they were spatio-temporal trajectories. For this purpose, we first extend the shortest path distance in a graph to handle extra events like the loss of focus. Secondly, we build a distance metric to compare trajectories independent of their length and duration. This measure is especially built for our use-case since it not only measures extent of difference between topics studied by two users, but also quantifies the differences in their navigation behavior. Such diverse aspects cannot be fully captured by traditional approaches that rely on simple statistics like the number of pages viewed. Additionally, by comparing navigation patterns between classmates, we characterize teaching style and detect outliers or specific learning patterns.

The rest of the paper is structured as follows. In Section

*Work done while at National Institute of Technology Karnataka, India.

2, we briefly introduce the mBook project. Notations and concepts necessary for the construction of the distance are presented in Section 3. We also review existing distance metrics based on three properties that a trajectory distance should satisfy, to successfully capture pupils' navigation patterns. Our page and trajectory distances are built in Section 4. In Section 5.1, the clustering qualities of our contribution are highlighted. Finally, in Section 5.2, we study how behavior patterns influence pupils performances and depend on the teaching style.

2. MBOOK

The mBook [27] is an electronic textbook for history, developed for students from grades 6 to 9. It is a part of a project regrouping didacticians, psychologist and computer scientists to study the influence of ICT on pupils and teaching staff. The ebook itself is a website based on a Typo3 environment so that it can be used independently of the device. However, tablets are the predominant device in most classrooms. The primary organization of the book is in the form of web-pages, grouped to represent different chapters/content. The book has 5 chapters that cover Antiquity, Middle Age, Renaissance, 19th Century, and the 20th and 21st Centuries. It also has an additional chapter on methods.

Figure 1: Screenshot of the mBook.

Content types cover five main components: text, galleries, audios or videos, information areas and a navigation bar. The primary content is in the form of text. A student can add notes to the text or highlight parts of it. Galleries comprise of pictures related to the text. Some audio or video files are directly integrated to the web-page and can be visualized from there. Information areas below the text provide additional information, beyond what is assigned for the chapter. These are usually organized in boxes that can be opened and accessed with a click/keypress event. Finally, the navigation bar at the bottom of the page allows the student to traverse sections and create highlights or notes. The section traversals include moving to either the previous, current or next section pages. In total, there are 738 pages, including 478 galleries and 537 exercises. Every page is assigned a unique identifier.

Since its deployment, the mBook was used by about 3,000 students in seven schools of the German-speaking community of Belgium. Since 2013, approximately 40,000 sessions were initiated and more than 7 million events (clicks, scrolls, key press, etc.) were tracked.

The project overseeing the deployment of the ebook also organized standardized tests at the end of each academic year. Based on these tests, the competency and knowledge of the pupils in history was regularly assessed using a Rasch model [23]. Additional variables like motivation, IT access and IT skill were obtained by questionnaires and MCQ tests.

3. PRELIMINARIES

In this section, we introduce notation and concepts that will become handy in sections to follow.

3.1 Notations

We begin with formally introducing trajectories.

Definition 1 (Trajectory). *Let Ω be a set. A trajectory $X = (x_i, t_i)_{0 \le i \le N}$ on Ω is a sequence of points x_i of Ω and of time-stamps t_i counted relative to t_0 such that $t_i \le t_{i+1}$. The length of the trajectory X is $N+1$ and its duration is t_N.*

When the time component is not relevant, the t_i will be omitted. To ease legibility, a sequence $(x_i)_{0 \le i \le N}$ will be abbreviated $(x_i)_N$ whenever the context allows.
Trajectories are essentially time-series of spatial points. In order to later have a notion of similarity between two trajectories, one needs to have a notion of distance between two points. A sequence of elements of Ω is an element of the power set of Ω. Thus, we give an abstract definition of a *distance* that could then be used for points or sequences of points.

Definition 2 (Distance). *Let Ω be a set. The function $d : \Omega \times \Omega \to \mathbb{R}$ is called a* distance *if it satisfies these properties for any elements $x, y, z \in \Omega$:*

- $\Delta(x, x) = 0$,
- *Non-negativity:* $\Delta(x, y) \ge 0$,
- *Symmetry:* $\Delta(x, y) = \Delta(y, x)$.

It is a metric *if it also satisfies:*

- *Identity of indiscernibles:* $\Delta(x, y) = 0 \Leftrightarrow x = y$,
- *Triangle inequality:* $\Delta(x, z) \leq \Delta(x, y) + \Delta(y, z)$.

In the following, we will prefer the notion of *distance* which is less restrictive than a *metric*. However, the distinction can be crucial to some clustering algorithms such as DBSCAN [12, 19] or k-medoids [17, 3] that assume the triangle inequality holds and thus require a metric between points. Other approaches like k-means and many hierarchical clustering methods [24] work well with non-metric distances. One exception is Ward's method [30] that is even more restrictive and relies on Euclidean distance.
Since every metric is also a distance, in the remainder, we denote generic distances between points and trajectories using d and Δ respectively.

3.2 Requirements

The aim of the work is to regroup pupils trajectories of various durations, within the mBook. This grouping should depend on the visited pages and be independent of session start. Additionally, we would like similar behaviors to be regrouped together. This can be controlled by enforcing the distance to satisfy certain properties.

P1: If Y last longer than X, for any truncation Y' of Y lasting longer than X, $\Delta(X, Y') = \Delta(X, Y)$.

P2: If X' and Y' go through the same sequence of points as X and Y but slower (or faster), $\Delta(X, Y) = \Delta(X', Y')$.

P3: If X and Y are loops, i.e. they start and end at the same point, their n-iterations are denoted as X^n and Y^n. If X and Y have the same duration, then $\Delta(X^n, Y^n) = \Delta(X, Y)$.

To motivate these three properties, we will make use of an analogy using a track and field race. Let X and Y be competing athletes and Δ an observer measuring the distance between the runners. Once one of the athletes finishes the race or gives up, the competition ends and Δ cannot make any further measurements. This is what property P1 encloses.
Now suppose that two other competitors X' and Y' perform exactly like the previous ones, but they run at half the speed of X and Y. Δ would make the same observations as above, relative to the total duration of the race. Hence, as stated in P2, we require that $\Delta(X, Y) = \Delta(X', Y')$.
To illustrate P3, X and Y finish the first lap in the same time. They continue similarly for the remaining laps.Thus, the information Δ extracts is the same for every lap. In other words, as stated in P3, $\Delta(X^n, Y^n) = \Delta(X, Y)$.

The first property P1 implies that a trajectory and its sub-trajectories are considered as equal. Sequences of different lengths or durations can then have a distance of 0. Consequently, the identity of indiscernibles is prohibited. Note that property P2 requires that $\Delta(X, Y) = \Delta(X', Y')$, however in the general case, $\Delta(X, Y) \neq \Delta(X, Y')$.

3.3 Distances

Distances on trajectories can be split into two groups [5]: shape-based and warping-based approaches. Warping-based approaches [4, 29] aim at handling sequences of various length by finding an alignment that minimizes a cost function. Dynamic Time Warping (DTW) [4] is often used in speech recognition tasks, but can be leveraged for any type of time series. The main limitation of this measure is that the evaluation algorithm is computationally demanding and has a time complexity of $O(N^2)$ in the length of the longest trajectory. Approximations have been developed to bring the complexity to an almost linear asymptote [26] but at the cost of a lower precision.

DEFINITION 3 (DTW). *Given two trajectories $X = (x_i)_N$ and $Y = (y_j)_M$, dynamic time warping (DTW) computes an alignment $W = (w_k)_K$ with the following properties:*

- $w_k = (x_i, y_j)$, $1 \leq i \leq N$, $1 \leq j \leq M$,
- $w_1 = (x_1, y_1)$,
- $w_K = (x_N, y_M)$,
- $d(w_k) = d(x_i, y_j)$,
- $w_k = (x_i, y_j) \Rightarrow w_{k+1} \in \left\{ \begin{array}{c} (x_i, y_{j+1}) \\ (x_{i+1}, y_j) \\ (x_{i+1}, y_{j+1}) \end{array} \right\}$.

Finally the distance between X and Y is then given by:

$$\mathrm{DTW}(X, Y) = \min_W \sum_{k=1}^{|W|} d(w_k).$$

The final result is the sum of the distances of the aligned points. Hence, the value grows with the length of the trajectories. This prevents DTW from satisfying P1 and P3. Note that the time-stamps are not considered here. As a consequence, P2 is naturally satisfied given that the duration between two points is irrelevant.

Shape-based distances aim at capturing geometric properties of the trajectories. A representatives of this family are for example Hausdorff [16], as well as more recent ones like the One-Way-Distance [20] and the Symmetrized Segment-Path Distance [5].

DEFINITION 4 (HAUSDORFF). *Given two trajectories $X = (x_i)_N$ and $Y = (y_j)_M$. The Hausdorff distance is defined as*

$$\mathrm{HAUS}(X, Y) = \max \left(\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right).$$

The Hausdorff distance is independent of the timestamps of the points, hence property P2 is satisfied; the computation relies only on their distribution. The number of times each point is visited does however influence the distance. In

particular the situation described by P3 is holds.
A limitation of this measure is that it can be easily deceived by odd point distributions. Consider the three trajectories X, Y and Z represented in Figure 2. Although the shapes are very different, $Haus(X, Z) = Haus(Y, Z) = 3$.
If the last point of X were removed, $Haus(X, Z)$ would decrease. This is in contradiction with P1.

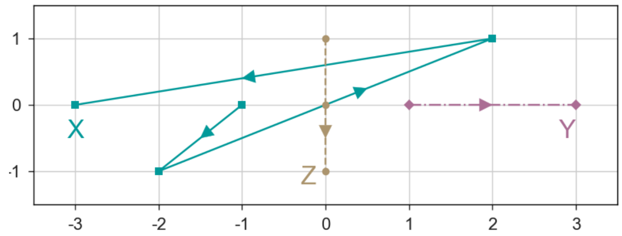

Figure 2: Three trajectories on the plane such that $Haus(X, Z) = Haus(Y, Z) = 3$. The arrows indicate the points orders.

The definitions of the One-Way-Distance (OWD) and Symmetrized Segment-Path Distance (SSPD) require to define the distance from a point to a trajectory:

Definition 5 (Distance Point-Trajectory). *Let x be a point of Ω and $Y = (y_j)_M$ be a trajectory. A segment of Y is a pair of successive points of Y, $[y_j, y_{j+1}]$. The distance between x and a segment of Y is the shortest distance between x and any point of the segment:*

$$d(x, [y_j, y_j + 1] = \min_{\tau \in [0,1]} (d(x, y_j \tau + (1 - \tau) y_{j+1}))$$

The distance between x and Y is the shortest distance between x and the segments of Y:

$$d(x, Y) = \min_j d(x, [y_j, y_j + 1]).$$

Definition 6 (OWD). *The one-way-distance (or OWD) between two trajectories $X = (x_i, t_i)_N$ and $Y = (y_j, t'_j)_M$ is defined as the integral of the distance from points of X to trajectory Y divided by the duration of X :*

$$OWD(X; Y) = \frac{1}{t_N} \int_{x \in X} d(x, Y) dx.$$

The symmetric OWD is the average of the OWD between X and Y:

$$sOWD(X, Y) = \frac{OWD(X; Y) + OWD(Y; X)}{2}.$$

The sOWD is close to the distance we want to build. Thanks to the normalization with duration, the measure satisfies P2 and P3. However it is not invariant per truncation as required by P1. If Y is truncated into Y', the duration of the later is shorter than the former, hence $OWD(Y'; X) \neq OWD(Y; X)$ in general.
Given that Y' is said in P1 to last longer than X, $OWD(X; Y') = OWD(X; Y)$. Yet, $\frac{1}{2}(OWD(X; Y') + OWD(Y'; X))$ is different from $\frac{1}{2}(OWD(X; Y) + OWD(Y; X))$ in general.

Definition 7 (SSPD). *The Segment-Path Distance, SPD, between two trajectories $X = (x_i)_N$ and $Y = (y_j)_M$ is :*

$$SPD(X; Y) = \frac{1}{N + 1} \sum_{i=0}^{N} d(x_i, Y).$$

The Symmetric Segment-Path Distance is the average of the SPD between X and Y:

$$SSPD(X, Y) = \frac{SPD(X; Y) + SPD(Y; X)}{2}.$$

The distance SSPD is independent of the time indexing, hence P2 is automatically satisfied. Besides the normalization by the number of points assure that the distance between loop trajectories is invariant with the number of iterations. Thus SSPD complies with P3.
However similarly than for OWD, the Symmetric Segment-Path Distance does not satisfy P1. Indeed if Y last longer than X and Y' is a truncation Y lasting as well longer than X, $SPD(Y'; X) \neq SPD(Y; X)$ while $SPD(X; Y') = SPD(X; Y)$. The averages are hence also different.

4. WEB TRAJECTORIES

Consider a website $\mathcal{W}$ whose structure is given by the page graph $G = (\mathcal{P}, \mathcal{E})$. We refer to the corresponding web-page of a node $p \in \mathcal{P}$ by $\mathcal{W}(p)$. That is, a node $p \in \mathcal{P}$ has a child $p' \in \mathcal{P}$ if users can transfer from page $\mathcal{W}(p)$ to $\mathcal{W}(p')$ by clicking a link or using the navigation bar. In that case $(p, p') \in \mathcal{E}$ holds. A loss of focus happens when the user turns off the screen of the tablet, or visit another tab. In order to handle this event, we add a dummy page F to $\mathcal{P}$. As it can happen anytime, F is connected to all the other pages.

A session on $\mathcal{W}$ can be represented as a sequence of pairs $P = (p_i, t_i)_{0 \leq i < l}$, where a user views page $\mathcal{W}(p_i)$ at timestamp t_i. For simplicity, we represent timestamps relatively to t_0, to retain the elapsed time on page and site. To call P a trajectory, we need to define a metric between its points.

4.1 Distances between pages

A natural distance measure for pages is the shortest path between the corresponding nodes in the underlying graph G. However, the auxiliary state F needs to be appropriately incorporated to allow for a meaningful application of a shortest path algorithm. Despite being connected to all the pages, we thus set the distance between F and any other page p to $dF \in \mathbb{R}_+$ such that

$$\max_{p,q \in \mathcal{P}} \text{ShortestPath}(p, q) < dF.$$

We motivate this choice by the fact that we want the clustering algorithm to consider a loss of focus as a special state. By making it very costly with respect to the other costs, we favor clusters of sessions that frequently visit F.

Definition 8 (Page Distance).
The distance d between two pages $p, q \in \mathcal{P}$ is defined

as follows.

$$d(p,q) = \begin{cases} \textsc{ShortestPath}(p,q) & , \text{ if } p \neq F \text{ and } q \neq F \\ dF & , \text{ if } p \neq F \text{ and } q = F \\ 0 & , \text{ if } p = F \text{ and } q = F \end{cases}$$

This *page distance* now allows the comparison of points inside a page graph and can be used by existing measures comparing trajectories. In order to assure that its usage does not remove the distance properties out of these measures, d needs to be a distance as well.

LEMMA 1. *The functions $d : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$ is a metric.*

PROOF. Non-negativity, symmetry and the identity of indiscernibles directly apply from the SHORTESTPATH which is a metric on $\mathcal{P} \setminus F$.
Let us prove the triangle inequality, i.e for p, q, s in $\mathcal{P}$: $d(p,r) \leq d(p,q) + d(q,r)$

- If $r = F$ and $q = F$, $d(F,F) = 0$.
- If $r = F$ and $q \neq F$, per non-negativity of d: $d(p,F) \leq dF \leq d(p,q) + dF = d(p,q) + d(q,r)$
- If none of the pages is F, then d is simply the SHORTESTPATH, which satisfies the triangle inequality.

□

4.2 Distances between trajectories

Following Definition 1, sessions can now be viewed as trajectories, more precisely web trajectories. In opposition to spatial trajectories, the position of a web trajectory between two timestamps does not evolve. Hence the position at any timestamp is precisely the one of the most recent point. We define the cross-product C of two trajectories X and Y to keep track the positions changes of X and Y.

DEFINITION 9 (CROSS-PRODUCT). *Let $X = (x_i, t_i)_N$ and $Y = (y_j, t'_j)_M)$ be two trajectories such that $t_N \leq t_M$. The cross-product of X and Y is the sequence $C = C(X,Y) = (c_k)_K = (\bar{t}_k, \bar{x}_k, \bar{y}_k)_{0 \leq k \leq K}$ defined as follows:*

- $\bar{t}_k \in \{t_i, 0 \leq i < N\} \cup \{t'_j, 0 \leq j < M \text{ and } t'_j \leq t_N\}$
- $c_0 = (0, x_0, y_0)$,
- *For $0 \leq k < K+1$, $c_k = (\bar{t}_k, \bar{x}_k, \bar{y}_k)$, with $\bar{x}_k = x_i$ such that $t_i \leq \bar{t}_k < t_{i+1}$, and $\bar{y}_k = y_j$ such that $t'_j \leq \bar{t}_k < t'_{j+1}$,*
- $c_K = (t_N, x_N, y_j)$ *such that* $t'_j \leq \bar{t}_N < t'_{j+1}$.

Now we devise a distance Δ for web-trajectories. Δ is defined as the normalized area spanned between them until the shortest one ends.

DEFINITION 10 (TRAJECTORY DISTANCE). *Let $X = (x_i, t_i)_N$, $Y = (y_j, t'_j)_M)$ be two trajectories and $C = (\bar{t}_k, \bar{x}_k, \bar{y}_k)_K$ their cross product:*

$$\Delta(X,Y) = \frac{1}{t_N} \sum_{k=1}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})(\bar{t}_{k+1} - \bar{t}_k)$$

In Section 3, we formulated three requirements for trajectory distances to assure certain properties in the clustering. The fact that none of the reviewed distances fulfills all of them, motivated the construction of Δ. We will now prove that our distance complies with the three conditions.

LEMMA 2. *The function Δ defined on pairs of web-trajectories satisfies the three properties P1, P2 and P3.*

PROOF. Let $X = (x_i, t_i)_N$ and $Y = (y_j, t'_j)_M$ be two trajectories and $C = (\bar{t}_k, \bar{x}_k, \bar{y}_k)_{0 \leq k \leq K}$ their cross product. We suppose that Y last longer: $t_N \leq t'_M$. Let us prove that each property is satisfied.

P1: The distance Δ depends only on the cross product of the two trajectories. Per construction, the cross-product contains only the points happening before that the shortest one ends, here X.
Hence for any truncation $Y' = (y_j, t'_j)_{0 \leq j < M'+1}$ of Y such that $M' < M$ and $t_N \leq t'_{M'}$, $C(X', Y) = C(X,Y)$. This implies $\Delta(X, Y') = \Delta(X,Y)$.

P2: For $\lambda > 1$, X' and Y' travel the same path than X and Y but λ times slower means that $X' = (x_i, \lambda t_i)_N$ and $Y' = (y_j, \lambda t'_j)_M)$. Their cross product is $C' = (\lambda \bar{t}_k, \bar{x}_k, \bar{y}_k)_{0 \leq k < K+1}$.

$$\begin{aligned} \Delta(X',Y') &= \frac{1}{\lambda t_N} \textstyle\sum_{k=1}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})(\lambda \bar{t}_{k+1} - \lambda \bar{t}_k) \\ &= \frac{\lambda}{\lambda t_N} \textstyle\sum_{k=1}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})(\bar{t}_{k+1} - \bar{t}_k) \\ \Delta(X',Y') &= \Delta(X,Y) \end{aligned}$$

P3: We will prove this property for $n = 2$, but it can be extended for any value. In this case X is a loop, i.e. $x_0 = x_N$, and $t_N = t'_M$. A trajectory X^2 traveling two times through X is of duration $2t_N$ and does not visit twice the initial position, i.e.

$$X^2 = (x_i, t_i)_{0 \leq i \leq N} \cup (x_i, t_i + t_N)_{1 \leq i \leq N}.$$

In turn, $C(X^2, Y^2) = (\bar{t}_k, \bar{x}_k, \bar{y}_k)_K \cup (\bar{t}_k + \bar{t}_K, \bar{x}_k, \bar{y}_k)_{1 \leq k \leq K}$. Hence:

$$\begin{aligned} \Delta(X^2,Y^2) &= \frac{1}{2t_N} \Big(\textstyle\sum_{k=0}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})(\bar{t}_{k+1} - \bar{t}_k) \\ &\quad + d(\bar{x}_K, \bar{y}_K)(\bar{t}_K + (\bar{t}_1 + t_N)) \\ &\quad + \textstyle\sum_{k=1}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})((\bar{t}_{k+1} + t_N) - (\bar{t}_k + t_N)) \Big) \end{aligned}$$

Given that $t_N = t'_M$ and that X and Y are loops, $\bar{x}_K = x_N = x_0$, $\bar{y}_K = y_N = y_0$ and $\bar{t}_K = t_N$. Besides following Definition 9 $\bar{t}_0 = 0$. Consequently ,

$$d(\bar{x}_K, \bar{y}_K)(\bar{t}_K + (\bar{t}_1 + t_N)) = d(\bar{x}_0, \bar{y}_0)(\bar{t}_0 + \bar{t}_1)$$

. This term can hence be integrated inside the second sum, such that we have:

$$
\begin{aligned}
\Delta(X^2, Y^2) &= \frac{1}{2t_N}\left(\sum_{k=0}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})(\bar{t}_{k+1} - \bar{t}_k)\right. \\
&\quad \left. + \sum_{k=0}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})(\bar{t}_{k+1} - \bar{t}_k)\right) \\
&= \frac{1}{2t_N}\left(2\sum_{k=0}^{K} d(\bar{x}_{k-1}, \bar{y}_{k-1})(\bar{t}_{k+1} - \bar{t}_k)\right) \\
\Delta(X^2, Y^2) &= \Delta(X, Y)
\end{aligned}
$$

□

Algorithm 1: $\Delta(X, Y)$

$\Delta \leftarrow 0$;
$T \leftarrow \min(t_N, t'_M)$;
Initialize a list C with $(0, x_0, y_0)$
foreach (t_i, x_i) *in* X *with* $i > 0$ *and* $t_i \leq T$ **do**
    Append (t_i, x_i, NAN) to C;
end
foreach (t'_j, y_j) *in* Y *with* $j > 0$ *and* $t'_j \leq T$ **do**
    Append (t'_j, NAN, y_j) to C;
end
Sort C accordingly to the first column;
$K \leftarrow$ length of C;
for $1 \leq k \leq K$ **do**
    $C_{k-1} = (t_{k-1}, x_{k_1}, y_{k-1})$;
    $C_k = (t_k, x_k, y_k)$;
    $\Delta \leftarrow \Delta + d(x_{k-1}, y_{k-1})(t_k - t_{k-1})$
    if x_k *is NAN* **then**
        $x_k \leftarrow x_{k-1}$;
    end
    if y_k *is NAN* **then**
        $y_k \leftarrow y_{k-1}$;
    end
end
Return Δ/T;

Algorithm 1 describes an efficient way to compute Δ.
Firstly, the distance Δ initialized to 0 and the shortest duration T is retrieved. The cross product C is a list of triplets : (t_k, x_k, y_k). The first coordinate indicates the timestamps, the two others the positions of X and Y at this time. The first tuple gives the initial positions of the two trajectories. Then all the positions of X and Y with a timestamp smaller or equal than T are included in C where the position of Y or X is set respectively as unknown. After that C is sorted accordingly to the timestamps.
Finally C is browsed starting from the second element ; Δ is updated accordingly to Definition 10 ; the missing positions are assigned using the last known positions.
Note that if X and Y have points with the same timestamp, C will contains tuples with the same timestamp. It is not problematic as they will cancel out each other during the update of Δ.
The time complexity of Algorithm 1 is $\mathcal{O}(N + M)$. It derives its efficiency from the fact that the assignments of the missing positions in C and the updates of Δ are done in the same loop.

4.3 Example

This section gives an example for the computation of the distance measure Δ. Consider the graph that is displayed in Figure 3. On the left, two trajectories are represented on

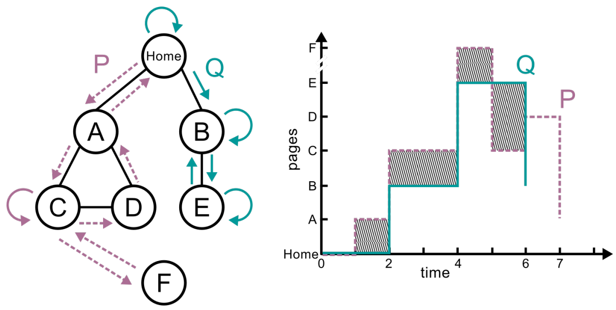

Figure 3: Trajectories on the page graph (left) and as timeseries (right). Edges between F and the other pages are not shown for legibility.

the page graph. Arrows represent a click that causes a page change. After vising page C, P loses the focus during one time unit. On the right, the progression of the trajectories over time is represented. The x-axis represents time and the y-axis the pages. The distance between P and Q is computed as follows.

$$
\begin{aligned}
\Delta(P, Q) &= \tfrac{1}{6}\left[d(H, H) + d(A, H) + d(C, B) * 2\right. \\
&\qquad \left. + d(F, E) + d(C, E)\right] \\
\Delta(P, Q) &= \tfrac{1}{6}\left[0 + 1 + 3 * 2 + dF + 4\right] \\
\Delta(P, Q) &= \frac{11 + dF}{6}
\end{aligned}
$$

5. EMPIRICAL RESULTS

5.1 Clustering

In this section, we report on clustering results that are obtained by using Hausdorff, DTW and the proposed Δ distances. We use K-means [24] as the underlying clustering algorithm. The distance of a trajectory to a cluster is the average distance between the trajectory and all the sessions in the cluster. We repeat every experiment 50 times and report on the best result for every measure.

The requirements stated in Section 3 aim to promote groupings of sessions that share long subsequences of viewed pages. To highlight the consequences of these choice, we restrict the data to only a single day. The subset contains 41 sessions from 37 users with an average duration of 32 minutes. The small scale allows for an interpretable analysis of the resulting clusterings. However, note that the computational complexity of DTW and Hausdorff quickly become infeasible with more data: The computation of the upper triangle of the DTW distance matrices using [4] requires more than 6 hours.

Although the sessions do not contain information about teachers, we will still evaluate the clusterings based on their similarity with the teachers' groupings. They should not be very different. Indeed, during one class, pupils tend to worked on the same subject. Thus, we expect them to be clustered together.
The teacher ID of the pupils behind session are represented by the y-axis of Figure 4.a. The connection times (x-axis) show six different classes. An analysis of the session logs shows that the closest classes in terms of topic and thus also

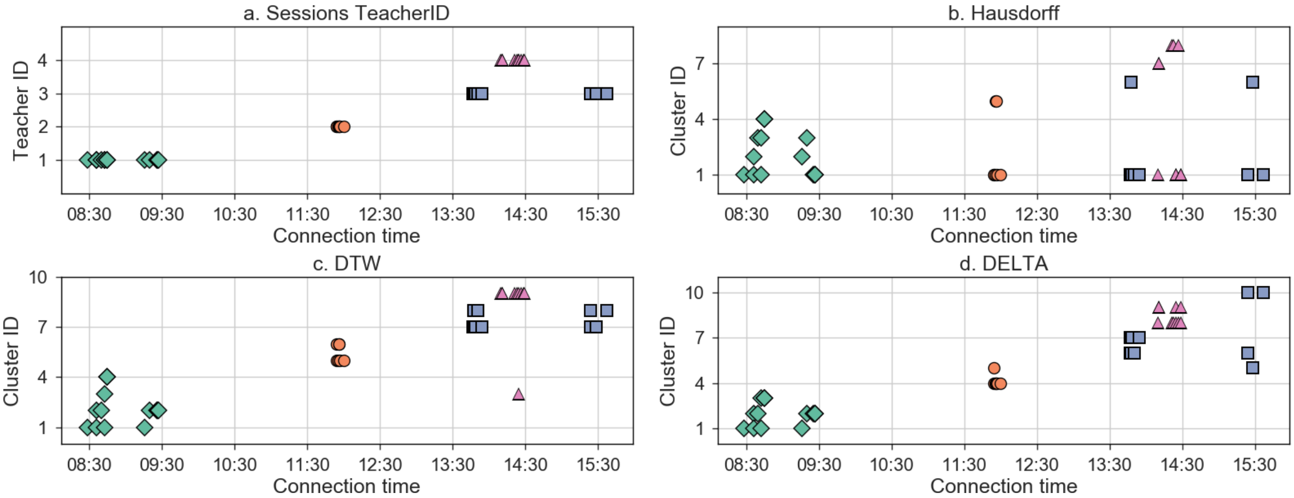

Figure 4: Teacher and cluster assignments of each sessions.

in terms of distance in the web-site graph are the ones of teacher 1 and 3, who dedicated all their lessons of this day to Alexander the Great and to the Roman Empire respectively. During a single class, teacher 2 focused on the situation of Belgium during WWII. The group of teacher 4 learned about the Reformation.

Two settings are evaluated. In the first one the number of clusters K is fixed to the number of teachers, that is $K = 4$. In the second experiment, K is chosen an order of magnitude higher to give the algorithm enough degrees of freedom to return the optimal amount of clusters for every measure. The returned clusters in this last setting are plotted in Figures 4.b to d. The final number of clusters found by each method and the homogeneity scores [25] of the clustering relatively to the teachers' distribution are given in Table 1. A homogeneity score of 1 indicates that no cluster contains sessions from multiple teachers.

Table 1: Number of clusters and homogeneities in the case of constrained or unconstrained clusterings.

	K=4		K=20	
Distance	# Cl.	Homog.	# Cl.	Homog.
Hausdorff	4	0.14	8	0.39
DTW	4	0.67	9	0.97
Δ	4	0.87	10	0.97

In both settings, the Hausdorff distance performs poorly. As shown in Figure 4.b, it fails at detecting class behaviors. The first cluster is spread all over the day, despite that each class studied different sections. By contrast, Δ's high homogeneities indicates that our proposed distance successfully detects the topics. Even when K is fixed to 4, Δ outperforms DTW and made few clustering errors. For $K = 20$, DTW and Δ create enough clusters such that all of them are pure with respect to the teacher, except for one session that is wrongly assigned in a cluster with sessions from another teacher. Interestingly for both distances, this mistake happens in a group of two sessions. DTW groups two sessions from teacher 1 and teacher 4 together, while Δ mistakenly associates a session from teacher 2 with a session from teacher 3, respectively.

For $K = 20$, the main difference between DTW and Δ is how they handle teacher 4. While DTW aims to group sessions associated with teacher 4 together, our distance measure splits them into two clusters. The trajectories of each cluster for each measure are shown in Figure 5. The pages are organized per chapter.

DTW detects the topic well as all the sessions dealing with Renaissance are grouped together. Cluster 3 in Figure 5.a is actually the DTW's cluster that is made only of two sessions from two different teachers. It is not clear why this artifact occurs. By contrast, our distance measure creates two groups out of all trajectories visiting the Renaissance's chapter. Cluster 8 shown in Figure 5.c contains those sessions that navigate more or less directly to the page about the Reformation and then stay on that page until the session is terminated. Sessions with more irregular trajectories are put into cluster 9. Thus, in addition to the topic, the shape of the trajectories is also a determining factor for Δ-based clusterings.

This section showed that pupils may exhibit very different types of behavior during the same class and that our distance measure performs well in detecting these behaviors. The next section investigates how the behaviors relate to the pupils performance in the class.

5.2 Assessments

In this section, we study the relation between the expressed behavior and the pupil's scores described in Section 2.

The activity of a user during one session can be measured through statistics like the 'number of pages seen per minute' (PPM) or the 'number of events per minute' (EPM). The average distance between a pupil's session and the other class sessions indicates how much the pupil's usage diverges from the group's.
However, these values can not be used to compare the activity between classes. Indeed, in a class with an average

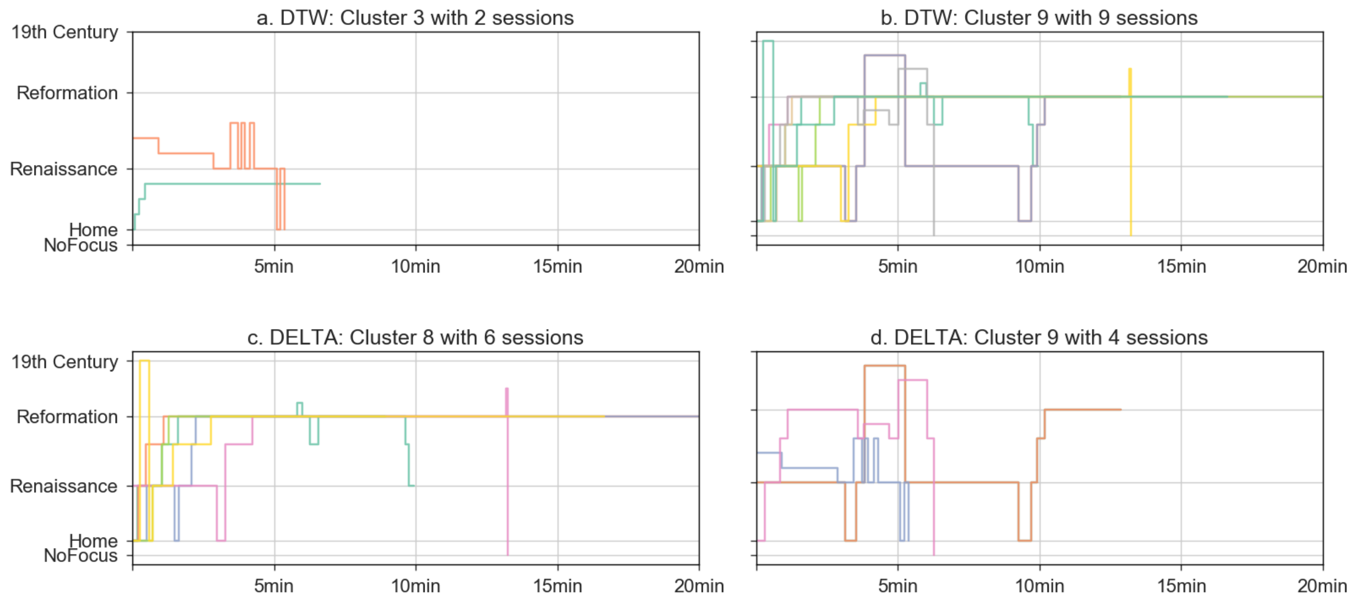

Figure 5: Trajectories of clusters obtained using DTW and Δ associated to the class of teacher 4.

of one page view per minute, a user viewing one page per minute will be considered as regular. However if the average of the class was 3, the same user would appear too inactive. Hence, these quantities need to be expressed relative to the average value of each class.
The average distance between trajectories of a class, also called the intra-class distance, is denoted as Ψ. The average distance of session P to the other class trajectories, also called divergence of the session, is denoted as $\psi(P)$.

We extract 400 class-sessions between February and July 2017, under the supervision of two teachers in two different schools. A class-session happens between 08:00 and 16:00 and contains at least five sessions from pupils with the same teacher that all start within 10 minutes. Table 3 contains the number of classes, sessions associated to the teacher, as well as the number of pupils. The average intra-class distance of the teachers' classes are given in the last column with standard deviations. Correlations between the measures and the pupils' scores are reported in Table 2. Pearson's correlations with a p-value smaller than 5% are marked in bold face. The displayed numbers indicate that the two groups show different behavior and that the teachers apply different teaching styles.

Table 2 suggests that while the three indicators correlate with the pupils competencies, they do so in different directions. For instance, pupils that possess a higher ψ, visit more pages per minute or interact more than the other pupils, during the same class. These pupils of teacher A perform better at the competency test. The opposite holds for the pupils of teacher B.
These differences can be interpreted only if put in the context of the average intra-class distances, given in Table 3. A Mann-Whitney U test [22, 13] between the Ψ of the two teachers' classes returns a U-value of 85 (<87 critical) and a one-sided p-value of 0.02. Thus, we can state that the pupils in teacher B's classes have more definite trajectories. And pupils who diverge from the predominant path tend to perform worst. To the contrary, the worst performing pupils of teacher A, whose classes present in average a bigger Ψ, are those that under-use the textbook.

The fact that all the indicators correlate with competency could mistakenly be interpreted as redundancy. However, we observe cases where only ψ is significant. For example, a small ψ correlates with high motivation in group A. This is remarkable, since it presents a correlation in the opposite direction of competency.
In the case of teacher B, pupils with low ψ perform better at the competency tests but also possess higher skills in information and communication technologies compared to their classmates. Indeed, among teacher B's pupils, the Pearson coefficient between these two scores indicate a correlation (0.399, p-value 0.0002); PPM and EPM fail to capture this effect.

In addition to the classical PPM and EMP, ψ appears to be a good indicator of the pupils' performances. Besides, it captures relations that are hidden to PPM and EMP and that are independent of connections between different scores.

6. DISCUSSION

In this paper, we focus on methods to extract diverse usage patterns of an e-book, through analysis of spatio-temporal, web-log trajectories. While conventional methods focus on individual events like page-clicks or scrolls, we extract and analyze trajectories within a web-page as a whole. To achieve this, we propose to embed the structure of electronic textbooks into graphs. Once pages of the ebook are associated with nodes in the graph, shortest path algorithms can be applied to compute distances between pages. Additionally, we also lift these distances to entire sessions, by making use of cross-products. The establishment of the distance metrics facilitates the use of spatial clustering methods to sessions of possibly unequal

Table 2: Pearson's correlations and associated p-values for each combination of pupil's activity indicators and score.

Teacher A

	Competency		Knowledge		Motivation		IT Access		IT Skill	
	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value
ψ	**0.179**	0.012	0.096	0.182	**-0.17**	0.017	0.023	0.745	0.092	0.202
PPM	**0.145**	0.044	0.133	0.064	0.039	0.587	-0.002	0.979	0.019	0.789
EPM	**0.185**	0.009	**0.156**	0.03	-0.065	0.37	-0.022	0.761	0.063	0.381

Teacher B

	Competency		Knowledge		Motivation		IT Access		IT Skill	
	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value
ψ	**-0.224**	0.047	-0.165	0.146	0.096	0.402	-0.069	0.547	**-0.357**	0.001
PPM	**-0.232**	0.039	0.049	0.671	0.111	0.331	0.188	0.097	-0.156	0.171
EPM	**-0.232**	0.04	-0.141	0.216	-0.142	0.212	0.081	0.481	0.059	0.604

Table 3: Summary of the analyzed classes.

	#Class	#Sessions	#Pupils	Ψ
Teacher A	27	200	48	5.76 (1.41)
Teacher B	11	80	22	4.48 (1.61)

length.

Empirically, we show that pupils exhibit very different types of behavior during the same class; the proposed distance measure outperforms baseline measures in grouping and detecting these behaviors. Moreover, in another experiment, we show that our distance measure differentiates between teaching styles and facilitates comparison between user behavior and user competence. The average dissimilarity between sessions during a class can thus be turned into an effective indicator of pupil performance and teaching technique. This study thus facilitates a thorough understanding of the effectiveness of e-books, in a classroom setup.

The empirical success of the proposed distance metric establishes it as a useful tool to analyze learning and teaching behaviour in a classroom. We thus hope to further extend these experiments to detect more complex learning patterns, now that a suitable comparison metric has been developed. For instance, our technique could be extended to detect 'outliers' or pupils who completely contravene typical classroom behaviour. It will further be interesting to establish correlations between outliers and performance. This will throw more light on the effectiveness of the teaching style and the ebook medium.

Acknowledgments

This research has been funded in parts by the German Federal Ministry of Education and Science BMBF under grant QQM/01LSA1503C.

7. REFERENCES

[1] H. Alt and M. Godau. Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(01n02):75–91, 1995.

[2] R. Bakeman, B. F. Robinson, and V. Quera. Testing sequential association: Estimating exact p values using sampled permutations. *Psychological methods*, 1(1):4, 1996.

[3] S. Baraty, D. A. Simovici, and C. Zara. The impact of triangular inequality violations on medoid-based clustering. In *International Symposium on Methodologies for Intelligent Systems*, pages 280–289. Springer, 2011.

[4] D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

[5] P. Besse, B. Guillouet, J.-M. Loubes, and R. François. Review and perspective for distance based trajectory clustering. *arXiv preprint arXiv:1508.04904*, 2015.

[6] A. Boubekki, U. Kröhne, F. Goldhammer, W. Schreiber, and U. Brefeld. Toward data-driven analyses of electronic text books. In *EDM*, pages 592–593, 2015.

[7] A. Boubekki, U. Kröhne, F. Goldhammer, W. Schreiber, and U. Brefeld. Data-driven analyses of electronic text books. In *Solving Large Scale Learning Tasks. Challenges and Algorithms*, pages 362–376. Springer, 2016.

[8] I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White. Visualization of navigation patterns on a web site using model-based clustering. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 280–284. ACM, 2000.

[9] K.-E. Chang, C.-T. Chang, H.-T. Hou, Y.-T. Sung, H.-L. Chao, and C.-M. Lee. Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Computers & Education*, 71:185–197, 2014.

[10] G. Conole. Moocs as disruptive technologies: strategies for enhancing the learner experience and

quality of moocs. *RED. Revista de Educación a Distancia*, (50), 2016.

[11] T. Eiter and H. Mannila. Computing discrete fréchet distance. Technical report, Citeseer, 1994.

[12] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[13] M. P. Fay and M. A. Proschan. Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys*, 4:1, 2010.

[14] M. M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22(1):1–72, 1906.

[15] C. Geigle and C. Zhai. Modeling student behavior with two-layer hidden markov models. *JEDM| Journal of Educational Data Mining*, 9(1):1–24, 2017.

[16] F. Hausdorff. *Mengenlehre*. Walter de Gruyter Berlin, 1927.

[17] L. Kaufmann and P. Rousseeuw. Clustering by means of medoids. pages 405–416, 01 1987.

[18] R. Kop. The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *The International Review of Research in Open and Distributed Learning*, 12(3):19–38, 2011.

[19] M. Kryszkiewicz and P. Lasek. Ti-dbscan: Clustering with dbscan by means of the triangle inequality. In *International Conference on Rough Sets and Current Trends in Computing*, pages 60–69. Springer, 2010.

[20] B. Lin and J. Su. Shapes based trajectory queries for moving objects. In *Proceedings of the 13th annual ACM international workshop on Geographic information systems*, pages 21–30. ACM, 2005.

[21] S. Livingstone. Critical reflections on the benefits of ict in education. *Oxford review of education*, 38(1):9–24, 2012.

[22] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[23] G. Rasch. Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests. 1960.

[24] L. Rokach and O. Maimon. Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer, 2005.

[25] A. Rosenberg and J. Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 2007.

[26] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, 2007.

[27] W. Schreiber, A. Schöner, and F. Sochatzy. *Analyse von Schulbüchern als Grundlage empirischer Geschichtsdidaktik*. Kohlhammer Verlag, 2013.

[28] W. Schreiber, F. Sochatzy, and M. Ventzke. Auf dem weg zu digital-multimedialen lehr-und lernmitteln für kompetenzorientiertes inklusives unterrichten und lernen. *Online Publikation, Medienberatung NRW. Zugriff am*, 11:2017, 2014.

[29] M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. In *Data Engineering, 2002. Proceedings. 18th International Conference on*, pages 673–684. IEEE, 2002.

[30] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

[31] C. Yin, N. Uosaki, H.-C. Chu, G.-J. Hwang, J.-J. Hwang, I. Hatono, E. Kumamoto, and Y. Tabata. Learning behavioral pattern analysis based on students' logs in reading digital books. 12 2017.