Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 159 (2019) 707–716

Procedia Computer Science

www.elsevier.com/locate/procedia

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Discovery of closed spatio-temporal sequential patterns from event data

Piotr S. Maciąg*, Marzena Kryszkiewicz, Robert Bembenik

Institute of Computer Science, Warsaw University of Technology
Nowowiejska 15/19,
00-665, Warsaw, Poland

Abstract

In the paper, we first thoroughly examine and prove properties of the participation index of spatio-temporal sequential patterns. Then, we introduce notions of a closure of a spatio-temporal sequential pattern and a closed spatio-temporal sequential pattern, as well as investigate and prove their properties. In particular, we prove that the set of all participation index strong closed spatio-temporal sequential patterns constitues a lossless representation of all participation index strong spatio-temporal sequential patterns. We also propose an algorithm, called CST-SPMiner, for discovering all participation index strong closed spatio-temporal sequential patterns. CST-SPMiner is an adaptation of the STBFM algorithm, which was proposed recently for the discovery of spatio-temporal sequential patterns with high participation index. While STBFM uses the CSP-tree structure for time-efficient candidate patterns generation and evaluation, CST-SPMiner uses it also for fast identification of closed patterns. Efficiency and effectiveness of our algorithm were verified on real crime data for Boston.

Keywords: closed spatio-temporal sequential patterns, event spatio-temporal data, participation index, crime data

1. Introduction

Event spatio-temporal data is collected by many modern sensing devices. This type of data is characterized by both geographical (spatial) location and temporal aspects of data objects as well as a set of related event types [9, 10]. Crime related data, where each crime incident is characterized by a geographical location, occurrence time and crime type, is an example of event spatio-temporal data. Another example is a set of disease related data: occurrences of several types of diseases can be observed on some area over a certain time span. Discovering relations in the form of

* Corresponding author
E-mail address: {P.Maciag, M.Kryszkiewicz, R.Bembenik}@ii.pw.edu.pl

10.1016/j.procs.2019.09.226

sequential patterns showing time dependencies between spatially close occurrences of different types of diseases or crimes provides a useful insight, potentially preventing future undesirable consequences.

In [11], we introduced the participation index in order to evaluate significance of spatio-temporal sequential patterns. We also proposed the STBFM algorithm and the SPTree structure to efficiently discover such patterns with high participation index. As follows from the performed experiments [11], the number of discovered patterns strongly depends on the participation index threshold value. For lower threshold values, their number is huge and practically impossible to analyse. In order to alleviate this problem, in this paper, we focus on efficient discovery of a concise representation of spatio-temporal sequential patterns in the form of *closed spatio-temporal sequential patterns*.

Problem statement. The problem considered in this paper is: (1) to derive properties of the participation index of spatio-temporal sequential patterns and participation index closed spatio-temporal sequential patterns; (2) to apply the derived properties for efficient discovery of closed spatio-temporal sequential patterns with the participation index greater than a given threshold θ value.

Related work. The problem of frequent sequential patterns discovery from temporal data without spatial dimension was formulated in [1] and [12]. An overview of methods of discovering this kind of patterns is given e.g. in [14, 5]. In [16], frequent closed sequential patterns were proposed as a lossless and concise representation of all frequent sequential patterns. The CloSpan algorithm was offered there for discovering such patterns. Since then several efficient algorithms such as ClaSP [7] and CloFAST [6] for mining frequent closed sequential patterns were proposed.

Recently, there is an increased interest in discovery of sequential patterns from event spatio-temporal data. The problem of discovering sequential patterns with high sequence index from event spatio-temporal data was introduced in [8]. In [13], discovery of cascade spatio-temporal patterns was proposed. [3] considered spatio-temporal sequential patterns discovery from evolving region-based instances. A survey of types of sequential patterns and methods of their discovery from event spatio-temporal data can be found in [9, 15]. Recently, discovery of spatio-temporal sequential patterns with high participation index was proposed by us in [11]. Our study in [11] is similar to those in [8] and [13], however significance of sequential patterns is defined differently in all these works.

While a number of approaches to mining sequential patterns from event spatio-temporal data is available in the literature, to the best of our knowledge no methods of discovering concise representations of spatio-temporal sequential patterns from such data were offered. This motivated us to investigate this topic and propose an efficient method for discovering closed spatio-temporal sequential patterns.

Contributions. In this paper, we provide the following new contributions:

(i) In Section 3, we thoroughly examine and prove properties of the participation index (as defined in [11]) of spatio-temporal sequential patterns. In particular, we prove that the participation index preserves anti-monotonicity property not only when extending a sequence by inserting elements at its end, but also by inserting elements at its beginning (see Theorem 1). We also provide an example showing that anti-monotonicity of the participation index is not guaranteed when sequences are extended in another way.

(ii) In Section 4, we introduce notions of a closure of a spatio-temporal sequential pattern and a closed spatio-temporal sequential pattern, as well as thoroughly examine and prove their properties. In particular, we prove that the set of all closed spatio-temporal sequential patterns with the participation index greater than θ is a lossless representation of all spatio-temporal sequential patterns with participation index greater than θ (see Theorem 3). We also prove that a spatio-temporal sequential pattern $\overrightarrow{s}$ is closed if and only if its participation index is different both from the participation index of its subsequence obtainable from $\overrightarrow{s}$ by removing its last element and from the participation index of its subsequence obtainable from $\overrightarrow{s}$ by removing its first element (see Theorem 4). This property provides an efficient method for distinguishing between closed and non-closed (in terms of the participation index) spatio-temporal sequential patterns.

(iii) In Section 5, we propose a new algorithm called CST-SPMiner for discovering all closed spatio-temporal sequential patterns with the participation index greater than θ. CST-Miner is an extension of the STBFM algorithm in that CST-SPMiner uses the STBFM method for finding spatio-temporal sequential patterns with participation index greater than θ, but additionally uses Theorem 4 to identify the closed ones among them. The CSP-tree structure is used in CST-SPMiner not only for efficient generation and evaluation of candidate patterns as in STBFM, but also for fast identification of closed patterns. (iv) In Section 6, we present the results of the experiments we performed to verify efficiency and effectiveness of our algorithm on real crime data for Boston.

2. Basic Notions

Let **F** denote a set of n event types and **D** denote a dataset of event instances such that for each event instance $e \in \mathbf{D}$, its spatial location, occurrence time (timestamp) and event type $F \in \mathbf{F}$ are provided. **D** will be called an *event spatio-temporal dataset*. The set of all event instances of type F in dataset **D** will be denoted by $\mathbf{D}(F)$.

Fig. 1 illustrates the case, where the set of event types $\mathbf{F} = \{A, B, C, D, E\}$, the dataset of event instances $\mathbf{D} = \{a_1, a_2, b_1, \ldots, b_8, c_1, \ldots, c_8, d_1, \ldots, d_3, e_1, \ldots, e_5\}$, event instances a_1 and a_2 are of type A (i.e., $\mathbf{D}(A) = \{a_1, a_2\}$), event instances $b_1, \ldots, b_8$ are of type B (i.e., $\mathbf{D}(B) = \{b_1, \ldots, b_8\}$), etc. A spatial location and occurrence time of each instance $e \in \mathbf{D}$ is provided in Fig. 1 as values of the horizontal and vertical coordinates, respectively. We will refer to the dataset from Fig. 1 in all illustrating examples in the paper.

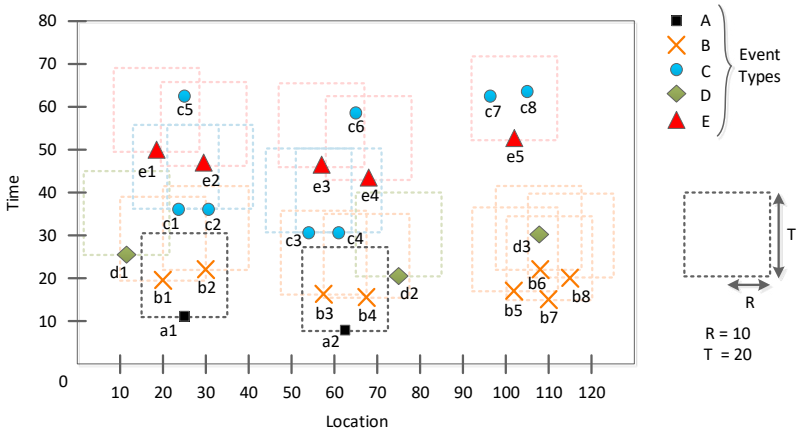

Fig. 1. An example event spatio-temporal dataset **D** showing event instances of corresponding event types. For an event instance $e \in \mathbf{D}$, dashed lines represent spatio-temporal space constisting of event instances that are distant from e by at most R and occurred later than e by at most T.

Definition 1 (Spatio-temporal sequential pattern, ST-sequential pattern). A *spatio-temporal sequential pattern* (in brief, *ST-sequential pattern*) is a sequence of elements each of which is an event type from **F**. The number of elements of the sequence is called its *length*. i-th element of sequence $\overrightarrow{s}$ is denoted by $\overrightarrow{s}[i]$. Sequence $\overrightarrow{s}$ consisting of m elements is denoted equivalently as $\overrightarrow{s}[1] \rightarrow \overrightarrow{s}[2] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$ or $\overrightarrow{s}[1] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$.

Definition 2 ((Immediate) subsequence / supersequence of an ST-sequential pattern). Let $\overrightarrow{s_1}$ and $\overrightarrow{s_2}$ be ST-sequential patterns. $\overrightarrow{s_1}$ is a *subsequence* of $\overrightarrow{s_2}$ and $\overrightarrow{s_2}$ is a *supersequence* of $\overrightarrow{s_1}$ if $\overrightarrow{s_1}$ can be obtained from $\overrightarrow{s_2}$ by removing any number of elements from $\overrightarrow{s_1}$. $\overrightarrow{s_1}$ is an *immediate subsequence* of $\overrightarrow{s_2}$ and $\overrightarrow{s_2}$ is an *immediate supersequence* of $\overrightarrow{s_1}$ if $\overrightarrow{s_1}$ is a subsequence of $\overrightarrow{s_2}$ and $\overrightarrow{s_1}$ contains exactly 1 element less than $\overrightarrow{s_2}$.

The notion of the participation index was introduced in [11] as a measure of significance of an ST-sequential pattern. In order to provide the definition of this measure, we will first introduce a number of useful auxiliary notions.

Definition 3 (Neighborhood of event instance with respect to event type [8]). For an event instance e, the *neighborhood of e with respect to an event type $F \in \mathbf{F}$* is denoted by $N_F(e)$ and is defined as follows: $N_F(e) = \{p | p \in \mathbf{D}(F) \wedge distance(p.location, e.location) \leq R \wedge (p.time - e.time) \in (0, T]\}$, where R and T denote user-given spatial distance threshold and time window threshold, respectively.

Thus, the neighborhood $N_F(e)$ of event instance e is the set of event instances of type F that are distant from e by at most R and occurred later than e by at most T. Clearly, $N_F(e)$ is a subset of $\mathbf{D}(F)$ for any event instance e. For example, $N_B(a_1) = \{b_1, b_2\}$, while $\mathbf{D}(B) = \{b_1, \ldots, b_8\}$. Hence, $N_B(a_1) \subseteq \mathbf{D}(B)$.

Definition 4 (Set of event instances supporting an element of an ST-sequential pattern [11]). *Set of event instances supporting i-th element of ST-sequential pattern* $\overrightarrow{s}$ is denoted by $I(\overrightarrow{s}, i)$ and is defined as follows:

$$I(\overrightarrow{s}, i) = \begin{cases} \mathbf{D}(\overrightarrow{s}[1]) & \text{when } i = 1, \\ \bigcup\limits_{e \in I(\overrightarrow{s}, i-1)} N_{\overrightarrow{s}[i]}(e) & \text{when } i > 1. \end{cases} \quad (1)$$

Set $I(\overrightarrow{s}, 1)$ of instances supporting the first event type of sequence $\overrightarrow{s}$ is defined as the set of all instances of this event type in $\mathbf{D}$; that is, to $\mathbf{D}(\overrightarrow{s}[1])$. Set $I(\overrightarrow{s}, i)$ of instances supporting each next $i-$th event type of $\overrightarrow{s}$ is defined as the set-theoretical union of the neighborhoods of the instances supporting $(i-1)-$th event type of $\overrightarrow{s}$. Let $\overrightarrow{s} = A \rightarrow B \rightarrow D$. Then, based on the dataset in Fig. 1, $I(\overrightarrow{s}[1]) = \{a1, a2\}, I(\overrightarrow{s}[2]) = \{b1, b2, b3, b4\}$, $I(\overrightarrow{s[3]}) = \{d1, d2\}$).

Definition 5 (Participation ratio [11]). The *participation ratio of an i-th element of ST-sequential pattern* $\overrightarrow{s}$, where $i \geq 1$, is denoted by $PR(\overrightarrow{s}, i)$ and is defined as the ratio of the cardinality of the set of event instances supporting i-th element of $\overrightarrow{s}$ to the number of all instances of type $\overrightarrow{s}[i]$ in the dataset $\mathbf{D}$; that is: $PR(\overrightarrow{s}, i) = \dfrac{\left|I(\overrightarrow{s}, i)\right|}{\left|\mathbf{D}(\overrightarrow{s}[i])\right|}$.

Clearly, the participation ratio takes values from the interval $[0, 1]$. Based on the dataset from Fig. 1, we conclude for ST-sequential pattern $\overrightarrow{s} = A \rightarrow B \rightarrow D$ that $PR(\overrightarrow{s}, 1) = \frac{2}{2} = 1$, $PR(\overrightarrow{s}, 2) = \frac{4}{8} = 0.5$ and $PR(\overrightarrow{s}, 3) = \frac{2}{3} = 0.66$.

Definition 6 (Participation index [11]). The *participation index of ST-sequential pattern* $\overrightarrow{s} = \overrightarrow{s}[1] \rightarrow \overrightarrow{s}[2] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$ is denoted by $PI(\overrightarrow{s})$ and is defined as the minimum from the participation ratios of all elements of $\overrightarrow{s}$; that is, $PI(\overrightarrow{s}) = \min(\{PR(\overrightarrow{s}, i) |\ i = 1, 2, \ldots, m\})$.

Thus, the participation ratio of $\overrightarrow{s} = A \rightarrow B \rightarrow D$ equals the minimum from $PR(\overrightarrow{s}, 1)$, $PR(\overrightarrow{s}, 2)$ and $PR(\overrightarrow{s}, 3)$; that is, $PI(\overrightarrow{s}) = \min(\{1, 0.5, 0.66\}) = 0.5$.

Definition 7 (PI-strong ST-sequential pattern). An ST-sequential pattern is called *PI-strong* if its participation index PI is greater than a given participation index threshold θ.

3. Theoretical Foundations of Spatio-Temporal Sequential Patterns

Knowledge of properties of patterns can be used to devise efficient algorithms for their discovery. In particular, the STBFM algorithm [11] uses the following properties to reduce the number of candidates for PI-strong ST-sequential patterns: If $\overrightarrow{s}$ is an ST-sequential pattern, $\overrightarrow{s_1}$ is its subsequence obtainable from $\overrightarrow{s}$ by removing a number of its last elements and $\overrightarrow{s_2}$ is its subsequence obtainable from $\overrightarrow{s}$ by removing a number of its first elements, then $PI(\overrightarrow{s}) \leq PI(\overrightarrow{s_1})$ and $PI(\overrightarrow{s}) \leq PI(\overrightarrow{s_2})$. The former property follows trivially from the definition of the participation index, however, the second property does not. In fact, no proofs of these properties were provided in [11]. In this section, we thoroughly examine and prove these two and a number of other properties of ST-sequential patterns.

Let us start with an example showing that neither monotonicity nor anti-monotonicity (w.r.t. sequences containment) of the participation index is guaranteed. Let us consider ST-sequential patterns $\overrightarrow{s_1} = A \rightarrow C$, $\overrightarrow{s_2} = A \rightarrow B \rightarrow C$, $\overrightarrow{s_3} = A \rightarrow B \rightarrow C \rightarrow E$ and $\overrightarrow{s_4} = A \rightarrow B \rightarrow C \rightarrow E \rightarrow C$. Clearly, $\overrightarrow{s_1}$ is a subsequence of $\overrightarrow{s_2}$, which is a subsequence of $\overrightarrow{s_3}$, which, in turn, is a subsequence of $\overrightarrow{s_4}$. Their participation indices calculated with respect to the dataset and threshold values presented in Fig. 1 are as follows: $PI(\overrightarrow{s_1}) = 0$, $PI(\overrightarrow{s_2}) = 0.5$, $PI(\overrightarrow{s_3}) = 0.5$, $PI(\overrightarrow{s_4}) = 0.25$.

Even though the participation index is not guaranteed to be non-increasing for supersequences, we will show that it is non-increasing for a specific type of supersequences called *contiguous*.

Definition 8 ((Proper) contiguous subsequence / supersequence of an ST-sequential pattern). Let $\overrightarrow{s_1} = \overrightarrow{s_1}[1] \rightarrow \overrightarrow{s_1}[2] \rightarrow \cdots \rightarrow \overrightarrow{s_1}[m_1]$ and $\overrightarrow{s_2} = \overrightarrow{s_2}[1] \rightarrow \overrightarrow{s_2}[2] \rightarrow \cdots \rightarrow \overrightarrow{s_2}[m_2]$ be ST-sequential patterns. $\overrightarrow{s_1}$ is a *contiguous*

subsequence of $\overrightarrow{s_2}$ and $\overrightarrow{s_2}$ is a *contiguous supersequence* of $\overrightarrow{s_1}$ if $m_1 \leq m_2$ and there exits an integer k, where $0 \leq k \leq m_2 - m_1$, such that $\overrightarrow{s_1}[1] = \overrightarrow{s_2}[1+k] \wedge \overrightarrow{s_1}[2] = \overrightarrow{s_2}[2+k] \wedge \cdots \wedge \overrightarrow{s_1}[m_1] = \overrightarrow{s_2}[m_1+k]$. If $m_1 < m_2$, then $\overrightarrow{s}_1$ is a *proper contiguous subsequence* of $\overrightarrow{s}_2$ and $\overrightarrow{s}_2$ is a *proper contiguous supersequence* of $\overrightarrow{s}_1$.

Let $\overrightarrow{s_1}$ and $\overrightarrow{s_2}$ be ST-sequential patterns of lengths m_1 and m_2, respectively, and $\overrightarrow{s_1}$ be a contiguous subsequence of $\overrightarrow{s_2}$. This means that $\overrightarrow{s_1}$ is obtainable from $\overrightarrow{s_2}$ by removing its first k elements and last $(m_2 - m_1 - k)$ elements, where $0 \leq k \leq m_2 - m_1$.

Lemma 1 (Event instances supporting an element of a contiguous subsequence of an ST-sequential pattern). *Let $\overrightarrow{s_1}$ and $\overrightarrow{s_2}$ be ST-sequential patterns of length m_1 and m_2, respectively, such that $\overrightarrow{s_1}$ is a contiguous subsequence of $\overrightarrow{s_2}$. Let k, where $0 \leq k \leq m_2 - m_1$, be an integer such that $\overrightarrow{s_1}[1] = \overrightarrow{s_2}[1+k]$. Then, $I(\overrightarrow{s_1}, i) \supseteq I(\overrightarrow{s_2}, i+k)$ for $i = 1..m_1$.*

Proof. Since $\overrightarrow{s}_1$ is a contiguous subsequence of $\overrightarrow{s}_2$ and $\overrightarrow{s_1}[1] = \overrightarrow{s_2}[1+k]$, then $\overrightarrow{s_1}[i] = \overrightarrow{s_2}[i+k]$ for $i = 1..m_1$ (by Definition 8). Let $F_i = \overrightarrow{s_1}[i]$ for $i = 1..m_1$, Then, $\overrightarrow{s_2}[i+k] = F_i$ for $i = 1..m_1$.
Case $k = 0$**:** ST-sequential pattern $\overrightarrow{s_1} = \overrightarrow{s_1}[1] \rightarrow \overrightarrow{s_1}[2] \rightarrow \cdots \rightarrow \overrightarrow{s_1}[m_1]$ is equal to subsequence $\overrightarrow{s_2}[1+k] \rightarrow \overrightarrow{s_2}[2+k] \rightarrow \cdots \rightarrow \overrightarrow{s_2}[m_1+k]$ of ST-sequential pattern $\overrightarrow{s}_2$; that is, to subsequence $\overrightarrow{s_2}[1] \rightarrow \overrightarrow{s_2}[2] \rightarrow \cdots \rightarrow \overrightarrow{s_2}[m_1]$ of ST-sequential pattern $\overrightarrow{s_2}$. In this case, $I(\overrightarrow{s_1}, i) = I(\overrightarrow{s_2}, i) = I(\overrightarrow{s_2}, i+k)$ for $i = 1..m_1$.
Case $k > 0$**:** Here, we will apply Definition 4 to determine sets of event instances supporting elements of ST-sequential patterns $\overrightarrow{s_1}$ and $\overrightarrow{s_2}$.

- $I(\overrightarrow{s_1}, 1) = \mathbf{D}(F_1)$, while $I(\overrightarrow{s_2}, 1+k) = \bigcup_{e \in I(\overrightarrow{s_2}, k)} N_{F_1}(e)$. Thus, $I(\overrightarrow{s_1}, 1) \supseteq I(\overrightarrow{s_2}, 1+k)$.
- $I(\overrightarrow{s_1}, 2) = \bigcup_{e \in I(\overrightarrow{s_1}, 1)} N_{F_2}(e)$, while $I(\overrightarrow{s_2}, 2+k) = \bigcup_{e \in I(\overrightarrow{s_2}, 1+k)} N_{F_2}(e)$. Thus, $I(\overrightarrow{s_1}, 2) \supseteq I(\overrightarrow{s_2}, 2+k)$.
- ...
- $I(\overrightarrow{s_1}, m_1) = \bigcup_{e \in I(\overrightarrow{s_1}, m_1-1)} N_{F_{m_1}}(e)$, while $I(\overrightarrow{s_2}, m_1+k) = \bigcup_{e \in I(\overrightarrow{s_2}, m_1-1+k)} N_{F_{m_1}}(e)$. Thus, $I(\overrightarrow{s_1}, m_1) \supseteq I(\overrightarrow{s_2}, m_1+k)$. □

Lemma 2 (Property of the participation ratio). *Let $\overrightarrow{s_1}$ and $\overrightarrow{s_2}$ be ST-sequential patterns of length m_1 and m_2, respectively, such that $\overrightarrow{s}_1$ is a contiguous subsequence of $\overrightarrow{s}_2$. Let k, where $0 \leq k \leq m_2 - m_1$, be an integer such that $\overrightarrow{s_1}[1] = \overrightarrow{s_2}[1+k]$. Then, $PR(\overrightarrow{s_1}, i) \geq PR(\overrightarrow{s_2}, i+k)$ for $i = 1..m_1$.*

Proof. Since $\overrightarrow{s}_1$ is a contiguous subsequence of $\overrightarrow{s}_2$ and $\overrightarrow{s_1}[1] = \overrightarrow{s_2}[1+k]$, then each i-th element of sequence $\overrightarrow{s_1}$ is the same as $(i+k)$-th element of sequence $\overrightarrow{s_2}$ (by Definition 8). Thus, $|\mathbf{D}(\overrightarrow{s_1}[i])| = |\mathbf{D}(\overrightarrow{s_2}[i+k])|$. On the other hand, Lemma 1 implies that $|I(\overrightarrow{s_1}, i)| \geq |I(\overrightarrow{s_2}, i+k)|$. Therefore, $PR(\overrightarrow{s_1}, i) \geq PR(\overrightarrow{s_2}, i+k)$ for $i = 1..m_1$. □

Theorem 1 (Anti-monotonicity property of the participation index for contiguous supersequences). *Let $\overrightarrow{s_1}$ and $\overrightarrow{s_2}$ be ST-sequential patterns. If $\overrightarrow{s_1}$ is a contiguous subsequence of $\overrightarrow{s_2}$, then $PI(\overrightarrow{s_1}) \geq PI(\overrightarrow{s_2})$.*

Proof. Theorem 1 follows from the definition of the participation index (Definition 6) and Lemma 2. □

Let us consider again ST-sequential patterns $\overrightarrow{s_2} = A \rightarrow B \rightarrow C$, $\overrightarrow{s_3} = A \rightarrow B \rightarrow C \rightarrow E$ and $\overrightarrow{s_4} = A \rightarrow B \rightarrow C \rightarrow E \rightarrow C$. ST-sequential pattern $\overrightarrow{s_2}$ is a contiguous subsequence of $\overrightarrow{s_3}$, which is a contiguous subsequence of $\overrightarrow{s_4}$. $PI(\overrightarrow{s_2}) = 0.5 \geq PI(\overrightarrow{s_3}) = 0.5$ and $PI(\overrightarrow{s_3}) = 0.5 \geq PI(\overrightarrow{s_4}) = 0.25$.

Lemma 3. *Let $\overrightarrow{s_1}$ and $\overrightarrow{s_2}$ be ST-sequential patterns such that $PI(\overrightarrow{s_1}) = PI(\overrightarrow{s_2})$. For each ST-sequential pattern $\overrightarrow{s_3}$ such that $\overrightarrow{s_1}$ is a contiguous subsequence of $\overrightarrow{s_3}$ and $\overrightarrow{s_3}$ is a contiguous subsequence of $\overrightarrow{s_2}$, $PI(\overrightarrow{s_1}) = PI(\overrightarrow{s_3}) = PI(\overrightarrow{s_2})$.*

Proof. By Theorem 1: $PI(\overrightarrow{s_1}) \geq PI(\overrightarrow{s_3})$ and $PI(\overrightarrow{s_3}) \geq PI(\overrightarrow{s_2})$. Since $PI(\overrightarrow{s_1}) = PI(\overrightarrow{s_2})$, then $PI(\overrightarrow{s_1}) = PI(\overrightarrow{s_3}) = PI(\overrightarrow{s_2})$. □

Theorem 2 (Contiguous (supersequences) subsequences of (non-) PI-strong ST-sequential patterns). *If $\overrightarrow{s}$ is a PI-strong ST-sequential pattern, then all contiguous subsequences of $\overrightarrow{s}$ are PI-strong ST-sequential patterns. If $\overrightarrow{s}$ is not a PI-strong ST-sequential pattern, then all contiguous supersequences of $\overrightarrow{s}$ are not PI-strong ST-sequential patterns.*

Proof. Theorem 2 follows from Theorem 1. □

We will consider now properties of immediate contiguous subsequences of an ST-sequential pattern.

Lemma 4 (Immediate contiguous subsequences of an ST-sequential pattern). *Let* $\overrightarrow{s} = \overrightarrow{s}[1] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$, *where* $m \geq 2$. *ST-sequential patterns* $\overrightarrow{s_1} = \overrightarrow{s}[1] \rightarrow \cdots \rightarrow \overrightarrow{s}[m-1]$ *and* $\overrightarrow{s_2} = \overrightarrow{s}[2] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$ *are all and only immediate contiguous subsequences of* $\overrightarrow{s}$.

Definition 9 (First (second) parent of an ST-sequential pattern). Let $\overrightarrow{s} = \overrightarrow{s}[1] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$, where $m \geq 2$. The *first parent of* $\overrightarrow{s}$ is denoted by $firstParent(\overrightarrow{s})$ and is defined as the immediate contiguous subsequence $\overrightarrow{s_1} = \overrightarrow{s}[1] \rightarrow \cdots \rightarrow \overrightarrow{s}[m-1]$ of $\overrightarrow{s}$. The *second parent of* $\overrightarrow{s}$ is denoted by $secondParent(\overrightarrow{s})$ and is defined as the immediate contiguous subsequence $\overrightarrow{s_2} = \overrightarrow{s}[2] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$ of $\overrightarrow{s}$.

Thus, the only immediate contiguous subsequences of an ST-sequential pattern are its first parent and its second parent.

Lemma 5 (Construction of an ST-sequential patterns from its first parent and second parent). *Let* $\overrightarrow{s}$ *be an ST-sequential pattern of length* $m \geq 3$ *and* $\overrightarrow{s_1}$ *be the first parent of* $\overrightarrow{s}$. *Then,* $\overrightarrow{s} = \overrightarrow{s_1} \rightarrow F_m$, *where* F_m *is the last element of* $secondParent(\overrightarrow{s})$, *and* $secondParent(\overrightarrow{s}) = secondParent(\overrightarrow{s_1}) \rightarrow F_m$.

Proof. Let $\overrightarrow{s} = \overrightarrow{s}[1] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$. Then, $firstParent(\overrightarrow{s}) = \overrightarrow{s_1} = \overrightarrow{s}[1] \rightarrow \cdots \rightarrow \overrightarrow{s}[m-1]$ and $secondParent(\overrightarrow{s}) = \overrightarrow{s}[2] \rightarrow \cdots \rightarrow \overrightarrow{s}[m]$. Hence, $secondParent(\overrightarrow{s_1}) = \overrightarrow{s}[2] \rightarrow \cdots \rightarrow \overrightarrow{s}[m-1]$ and the last element F_m of the $secondParent(\overrightarrow{s})$ equals $\overrightarrow{s}[m]$. Therefore, $secondParent(\overrightarrow{s}) = secondParent(\overrightarrow{s_1}) \rightarrow F_m$, F_m is the last element of $secondParent(\overrightarrow{s})$ and $\overrightarrow{s} = \overrightarrow{s_1} \rightarrow F_m$. □

4. PI-Strong Closed Spatio-Temporal Sequential Patterns

In the data mining literature, closed patterns are defined for different types of patterns and data, e.g. closed frequent itemsets, closed sequential patterns in transaction data etc. They are typically defined as patterns having value of a respective anti-monotonous (or partly anti-monotonous) interestingness measure different from values of their all (some) proper super-patterns (supersets, supersequences, etc.). Following this way of defining closed patterns, we introduce a notion of a *closed ST-sequential pattern* with respect to the participation index, which, as we proved in the previous section, preserves anti-monotonicity for contiguous supersequences.

Definition 10 ((PI-)closed ST-sequential pattern and (PI-)closure of an ST-sequential pattern). *ST-sequential pattern* $\overrightarrow{s_1}$ is *PI-closed* (in brief, *closed*) if there exists no proper contiguous supersequence $\overrightarrow{s_2}$ of $\overrightarrow{s_1}$ such that the participation index $PI(\overrightarrow{s_2}) = PI(\overrightarrow{s_1})$. A *PI-closure* (in brief, *closure*) of ST-sequential pattern $\overrightarrow{s_1}$ is a contiguous supersequence $\overrightarrow{s_2}$ of $\overrightarrow{s_1}$ such that $\overrightarrow{s_2}$ is a closed ST-sequential pattern and $PI(\overrightarrow{s_2}) = PI(\overrightarrow{s_1})$.

Please note that an ST-sequential pattern may have more than one closure. For example, ST-sequential pattern $B \rightarrow C$ has the following two closures: $A \rightarrow B \rightarrow C \rightarrow E$ and $B \rightarrow B \rightarrow C \rightarrow E$. All the three ST-sequential patterns have the same participation index, which is equal to 0.5 (please see Fig. 1 and Fig. 2).

Definition 11 (The set of all PI-strong closed ST-sequential patterns, PI-SC). The set of all *PI-strong closed ST-sequential patterns* is denoted by PI-SC and is defined as the set of all closed ST-sequential patterns that are PI-strong.

Theorem 3 (PI-SC as a lossless representation of all PI-strong ST-sequential patterns).

a) The set PI-SC of all PI-strong closed ST-sequential patterns enables determining for each ST-sequential pattern whether it is PI-strong or not; namely, an ST-sequential pattern is PI-strong if and only if there exists its contiguous supersequence in PI-SC.

b) If $\overrightarrow{s}$ *is a PI-strong ST-sequential pattern, then* $PI(\overrightarrow{s}) = max\{PI(\overrightarrow{s_i}) \mid \overrightarrow{s_i}$ *is a contiguous supersequence of* $\overrightarrow{s}$ *and* $\overrightarrow{s_i} \in$ *PI-SC*$\}$. □

Proof. Ad Theorem 3a) Each ST-sequential pattern $\overrightarrow{s}$ has a closure being its proper or improper contiguous supersequence among closed ST-sequential patterns and the participation index of $\overrightarrow{s}$ is the same as that of its closure. If there is no contiguous supersequence of $\overrightarrow{s}$ in PI-SC, then a closure of $\overrightarrow{s}$ is among closed ST-sequential patterns that are not PI-strong. In such a case, $\overrightarrow{s}$ is not PI-strong as its closure is not. Otherwise, a closure of $\overrightarrow{s}$ is among its contiguous supersequences in PI-SC, which means that $\overrightarrow{s}$ is PI-strong as its closure is.

Ad Theorem 3b) Let $\overrightarrow{s}$ be a PI-strong ST-sequential pattern. Hence, its closures are also PI-strong. The question is which contiguous supereresequences of $\overrightarrow{s}$ in PI-SC are closures of $\overrightarrow{s}$. By Theorem 1, $PI(\overrightarrow{s})$ is not greater than the participation index of any of its contiguous supersequences. This condition implies that the closures of $PI(\overrightarrow{s})$ are those contiguous supersequences of $\overrightarrow{s}$ among all contiguous supersequences of $\overrightarrow{s}$ in PI-SC that have the greatest value of the participation index. □

Lemma 6 (Immediate contiguous supersequence of an ST-sequential pattern). *Let* $\overrightarrow{s_1} = \overrightarrow{s_1}[1] \rightarrow \overrightarrow{s_1}[2] \rightarrow \cdots \rightarrow \overrightarrow{s_1}[m_1]$ *and* $\overrightarrow{s_2} = \overrightarrow{s_2}[1] \rightarrow \overrightarrow{s_2}[2] \rightarrow \cdots \rightarrow \overrightarrow{s_2}[m_2]$ *be ST-sequential patterns such that* $\overrightarrow{s_1}$ *is a proper non-immediate contiguous subsequence of* $\overrightarrow{s_2}$ *and* $\overrightarrow{s_1}[1] = \overrightarrow{s_2}[1+k]$ *for some integer* $0 \leq k \leq m_2 - m_1$. *Then, there is at least one immediate contiguous supersequence of* $\overrightarrow{s_1}$*; namely,* $\overrightarrow{s_2}[k] \rightarrow \overrightarrow{s_1}[1] \rightarrow \cdots \rightarrow \overrightarrow{s_1}[m_1]$, *provided* $k > 0$, *and/or* $\overrightarrow{s_1}[1] \rightarrow \cdots \rightarrow \overrightarrow{s_1}[m_1] \rightarrow \overrightarrow{s_2}[m_1 + k + 1]$, *provided* $k < m_2 - m_1$.

Lemma 7. *ST-sequential pattern* $\overrightarrow{s}$ *is not closed if and only if there exists an immediate contiguous supersequence* $\overrightarrow{s_i}$ *of* $\overrightarrow{s}$ *such that the participation index* $PI(\overrightarrow{s_i}) = PI(\overrightarrow{s})$.

Proof. $\overrightarrow{s_1}$ is not a closed ST-sequential pattern $\iff$ there is a contiguous supersequence $\overrightarrow{s_j}$ of $\overrightarrow{s}$ such that $PI(\overrightarrow{s_j}) = PI(\overrightarrow{s})$ $\iff$ (by Lemma 3 and Lemma 6) there is an immediate contiguous supersequence $\overrightarrow{s_i}$ of $\overrightarrow{s}$ such that $PI(\overrightarrow{s_i}) = PI(\overrightarrow{s})$. □

Theorem 4. *ST-sequential pattern* $\overrightarrow{s}$ *is not closed if and only if* $\overrightarrow{s}$ *is the first parent or second parent of an ST-sequential pattern* $\overrightarrow{s_i}$ *such that the participation index* $PI(\overrightarrow{s_i}) = PI(\overrightarrow{s})$.

Proof. Let $\overrightarrow{s}$ be a non-closed ST-sequential pattern. By Lemma 7, there is an immediate contiguous supersequence of $\overrightarrow{s}$, say $\overrightarrow{s_i}$, such that $PI(\overrightarrow{s_i}) = PI(\overrightarrow{s})$. Hence, $\overrightarrow{s}$ is an immediate contiguous subsequence of $\overrightarrow{s_i}$. Thus, by Lemma 4, $\overrightarrow{s}$ is the first parent or second parent of $\overrightarrow{s_i}$ such that such $PI(\overrightarrow{s_i}) = PI(\overrightarrow{s})$. □

5. Discovering PI-Strong Closed Spatio-Temporal Sequential Patterns

In this section, we propose the CST-SPMiner algorithm (Algorithm 1) for discovering all PI-strong closed ST-sequential patterns (PI-SC). The algorithm bases on candidate generation and verification schema of the STBFM algorithm and the prefix SPTree, which were introduced in [11]. CST-SPMiner uses the methodology of the STBFM algorithm for mining PI-strong ST-sequential patterns iteratively. In each k-th iteration, the set of PI-strong ST-sequential patterns of length k (L_k) is mined. Initially, each pattern in L_k is marked as *closed*. However, in iteration $k+1$ its state may be changed from *closed* to *non-closed* according to Theorem 4. The resulting set of discovered PI-strong ST-sequential patterns that remained marked as *closed* is returned as the sought PI-SC patterns.

The execution of CST-SPMiner proceeds as follows. First, ST-sequential patterns of length 1 are generated from all event types in **F**. The participation indices of these patterns are set to 1 according to Definitions 4, 5, 6. Then, patterns of length 2 are generated and verified in two nested loops as presented in lines 3-12 of Algorithm 1. PI-SC patterns of length $k \geq 3$ are discovered by Algorithm 2 as follows: (1) for each pattern $\overrightarrow{s_i} \in L_{k-1}$, and each child $\overrightarrow{s_j}$ of the second parent of $\overrightarrow{s_i}$, a new pattern $\overrightarrow{s}$ is generated by concatenating ST-sequential pattern $\overrightarrow{s_i}$ with the last event type of $\overrightarrow{s_j}$ (see Lemma 5). Next, $\overrightarrow{s_i}$ and $\overrightarrow{s_j}$ are set as being first and second parent of $\overrightarrow{s}$, respectively. (2) The participation index (PI) of the newly created pattern is calculated. If PI is greater than the threshold θ, the pattern is added to the list of *children* of its first parent ($\overrightarrow{s_i}$) and stored as PI-strong ST-sequential pattern in L_k. (3) If the participation index of its first parent or second parent is equal to the participation index of $\overrightarrow{s}$, then such first or second parent is marked as non-closed in L_{k-1} according to Theorem 4 and is not returned in the PI-SC patterns set.

To efficiently calculate the participation index of candidate patterns, neighborhoods of event instances are determined with the plane sweep algorithm proposed in [2].

The result of the execution of Algorithm 1 with input parameters: $R = 10$, $T = 20$, and $\theta = 0.2$ for the example synthetic dataset presented in Fig. 1 is the CSP-tree shown in Fig. 2.

Algorithm 1 CST-SPMiner: Closed Spatio-Temporal Sequential Patterns Miner Algorithm

Input: **D** - event spatio-temporal dataset, **F** - set of event types, R - spatial threshold value, T - time window threshold value, θ - the participation index threshold value.
Assumption: L_k - a set of PI-strong ST-sequential patterns of length k, each of which initially is marked as closed.
Output: PI-SC = $\bigcup_k \{\overrightarrow{s} \in L_k \mid \overrightarrow{s}$ is marked as closed$\}$.

L_1 := generate PI-SC patterns of length 1 from all event types in **F**;
$L_2 := \emptyset$;
for each $\overrightarrow{s_i} \in L_1$ **do**
  for each $\overrightarrow{s_j} \in L_1$ **do**
    $\overrightarrow{s} := \overrightarrow{s_i}[1] \rightarrow \overrightarrow{s_j}[1]$; $\overrightarrow{s}[2].I := \bigcup_{e \in \overrightarrow{s_i}[1].I} N_{\overrightarrow{s}[2]}(e)$; $\overrightarrow{s}.PI := PR(\overrightarrow{s}, 2)$;
    if $\overrightarrow{s}.PI > \theta$ **then**
      $\overrightarrow{s}.firstParent := \overrightarrow{s_i}$, $\overrightarrow{s}.secondParent := \overrightarrow{s_j}$; Add $\overrightarrow{s}$ to $\overrightarrow{s_i}.children$; Add $\overrightarrow{s}$ to L_2;
      if $\overrightarrow{s}.PI = \overrightarrow{s}.firstParent.PI$ **then** mark $\overrightarrow{s}.firstParent$ as non-closed; **end if**
      if $\overrightarrow{s}.PI = \overrightarrow{s}.secondParent.PI$ **then** mark $\overrightarrow{s}.secondParent$ as non-closed; **end if**
    end if
  end for
end for
$k := 3$;
while $L_{k-1} \neq \emptyset$ **do**
  L_k := GenAndVerify(L_{k-1}); $k := k + 1$;
end while
return $\bigcup_k \{\overrightarrow{s} \in L_k \mid \overrightarrow{s}$ is marked as closed$\}$;

Algorithm 2 GenAndVerify(L_{k-1})

Input: L_{k-1} - a set of PI-strong ST-sequential patterns of length $k-1$.
Output: L_k - a set of PI-strong ST-sequential patterns of length k.

$L_k := \emptyset$;
for each $\overrightarrow{s_i} \in L_{k-1}$ **do**
  $\overrightarrow{s_l} := \overrightarrow{s_i}.secondParent$;
  for each $\overrightarrow{s_j} \in \overrightarrow{s_l}.children$ **do**
    $\overrightarrow{s} := \overrightarrow{s_i}[1] \rightarrow \overrightarrow{s_i}[2] \rightarrow \cdots \rightarrow \overrightarrow{s_i}[k-1] \rightarrow \overrightarrow{s_j}[k-1]$;
    $\overrightarrow{s}[k].I := \bigcup_{e \in \overrightarrow{s_i}[k-1].I} N_{\overrightarrow{s}[k]}(e)$;
    $\overrightarrow{s}.firstParent := \overrightarrow{s_i}$, $\overrightarrow{s}.secondParent := \overrightarrow{s_j}$;
    $\overrightarrow{s}.PI := \min(\overrightarrow{s}.firstParent.PI, PR(\overrightarrow{s}, k))$;
    if $\overrightarrow{s}.PI > \theta$ **then**
      Add $\overrightarrow{s}$ to $\overrightarrow{s_i}.children$; Add $\overrightarrow{s}$ to L_k;
      if $\overrightarrow{s}.PI = \overrightarrow{s}.firstParent.PI$ **then** mark $\overrightarrow{s}.firstParent$ as non-closed; **end if**
      if $\overrightarrow{s}.PI = \overrightarrow{s}.secondParent.PI$ **then** mark $\overrightarrow{s}.secondParent$ as non-closed; **end if**
    end if
  end for
end for
return L_k;

6. Experimental Evaluation

In the experiments, we compared execution times and the number of PI-SC patterns discovered by the offered CST-SPMiner algorithm with the execution times and the number of all PI-strong ST-sequential patterns discovered by the STBFM algorithm [11]. The experiments were carried out on crime related Boston city data for year 2014 [4]. First in our experiments, we used the *complete dataset* containing 40544 crime incidents of 26 crime types occurring

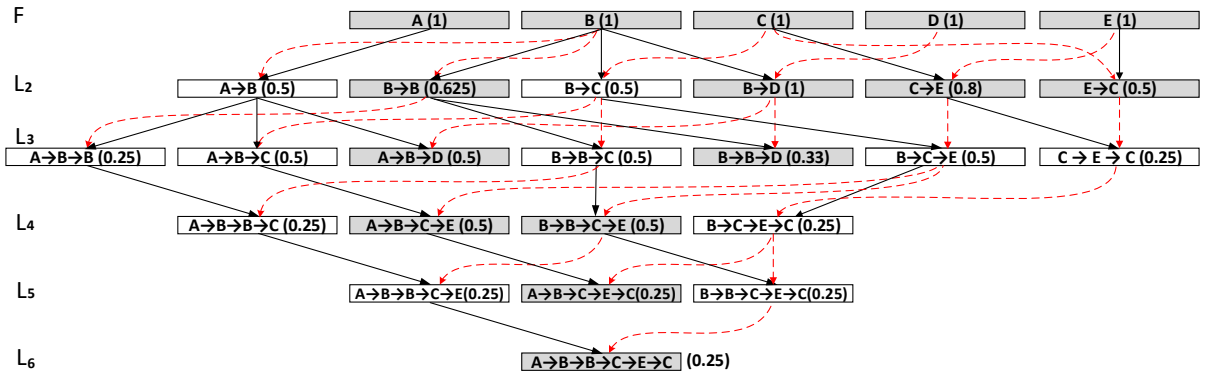

Fig. 2. The CSP-tree created for the dataset presented in Fig. 1. PI-strong ST-sequential patterns of length k (L_k) are stored in nodes on level k of the tree. Solid arrows indicate first parents of patterns, dashed arrows indicate second parents of patterns. Grey boxes represent PI-SC patterns, white boxes represent non-closed PI-strong patterns. Participation indices of patterns are provided in parentheses next to the patterns.

with various frequencies. The used dataset is publicly available at the Boston Police Department website. Then, we used the *reduced dataset* created by selecting only event instances of 10 event types occurring least frequently in the original dataset. These event types are: *violation of liquor laws, operating under influence, manslaughter, homicide, harassment, gambling offense, embezzlement, crimes against children, bomb, arson*. The reduced dataset contains 896 event instances overall.

Table 1 presents a comparison of the results for CST-SPMiner and STBFM for the complete dataset, while Table 2 shows a similar comparison for the reduced dataset. For the results obtained for the complete dataset, the greatest reduction of the discovered patterns was achieved for small values of θ and large values of time window T threshold. For example, for the neighborhood specification $R = 300$ meters, $T = 5760$ minutes (4 days) and the threshold $\theta = 0.015$, the number of all PI-strong ST-sequential patterns discovered by STBFM was 2 819 490, while the number of PI-SC patterns discovered by CST-SPMiner was 2 040 303, which means that PI-SC was 28% less numerous than the set of all PI-strong ST-sequential patterns. The difference between the number of discovered PI-SC patterns and all PI-strong ST-sequential patterns was even greater for the reduced dataset. For example, for $R = 500$ meters, $T = 57600$ minutes (40 days) and $\theta = 0.002$, the number of PI-SC patterns (8 023 267) was by 84.3% less than the number of all PI-strong ST-sequential patterns (49 213 713).

Table 1. Computation time (in seconds) and the number of discovered PI-SC patterns and PI-strong ST-sequential patterns for the *complete dataset*.

R = 200 meters, T = 14400 (10 days)					**R = 300 meters, T = 11520 (8 days)**					**R = 300 meters, T = 5760 (4 days)**				
	CST-SPMiner		**STBFM**			**CST-SPMiner**		**STBFM**			**CST-SPMiner**		**STBFM**	
θ	time	# patterns	time	# patterns	θ	time	# patterns	time	# patterns	θ	time	# patterns	time	# patterns
0.09	33	3848	33	4598	0.55	10	27	10	28	0.055	13	3346	13	3982
0.085	40	5161	36	6233	0.5	10	34	10	35	0.05	14	4484	14	5386
0.08	46	6887	47	8419	0.45	12	55	12	56	0.045	16	6248	15	7550
0.075	53	10 233	48	12 810	0.4	12	86	12	91	0.04	19	9899	17	12 014
0.07	65	15 817	57	20 196	0.35	19	182	17	195	0.035	22	17 732	22	22 191
0.065	83	27 388	77	35 483	0.3	21	336	20	385	0.03	33	33 954	32	43 327
0.06	119	51 663	112	67 672	0.25	28	760	27	881	0.025	47	79 701	49	102 731
0.055	212	133 962	192	182 501	0.2	53	2799	50	3388	0.02	93	282 156	96	367 400
0.05	488	430 763	443	593 785	0.15	269	39 224	265	51 881	0.015	357	2 040 303	346	2 819 490

Below we present example PI-strong closed ST-sequential patterns discovered from the reduced dataset for the following threshold values: $\theta = 0.02$, $R = 600$ meters and $T = 28800$ minutes (20 days):

crime against child → *violation of liquor law* → *operating under influence* → *harassment* → *embezzlement* (PI = 0.03);
violation of liquor law → *operating under influence* → *bomb* → *homicide* (PI = 0.04);
violation of liquor law → *operating under influence* → *homicide* (PI = 0.11);

Table 2. Computation time (in seconds) and the number of discovered PI-SC patterns and PI-strong sequential patterns for the *reduced dataset*.

R = 500 meters, T = 43200 (30 days)					R = 500 meters, T = 57600 (40 days)					R = 600 meters, T = 28800 (20 days)				
	CST-SPMiner		STBFM			CST-SPMiner		STBFM			CST-SPMiner		STBFM	
θ	time	# patterns	time	# patterns	θ	time	# patterns	time	# patterns	θ	time	# patterns	time	# patterns
0.01	0.2	9987	0.2	34 102	0.01	3	156 310	2	631 282	0.01	0.04	1314	0.06	3996
0.009	0.2	9987	0.2	34 102	0.009	2	156 310	2	631 282	0.009	0.04	1314	0.05	3996
0.008	0.6	51 219	0.7	141 860	0.008	11	644 583	11	2 560 637	0.008	0.04	1867	0.07	6990
0.007	0.6	51 219	0.6	141 860	0.007	12	644 583	10	2 560 637	0.007	0.05	1867	0.05	6990
0.006	0.7	56 812	0.7	161 238	0.006	19	1 153 357	15	4 211 651	0.006	0.06	2416	0.07	9920
0.005	1	76 894	0.9	228 285	0.005	40	2 125 039	31	9 351 445	0.005	0.09	4338	0.09	20 449
0.004	9	714 557	9	2 399 217	0.004	213	8 023 267	173	49 213 713	0.004	0.41	24 183	0.44	120 576
0.003	9	714 557	9	2 399 217	0.003	212	8 023 267	180	49 213 713	0.003	0.42	24 183	0.44	120 576
0.002	9	714 557	10	2 399 217	0.002	238	8 023 267	196	49 213 713	0.002	0.43	24 183	0.41	120 576

7. Conclusions

In this paper, we formally defined the problem of discovering closed spatio-temporal sequential patterns with the participation index higher than θ (PI-SC). In particular, we thoroughly analyzed and proved properties of ST-sequential patterns, their closures and closed ST-sequential patterns. We also proved that PI-SC constitutes a lossless representation of all spatio-temporal sequential patterns with the participation index higher than θ (PI-strong ST-sequential patterns). Moreover, we proposed the CST-SPMiner algorithm for finding PI-SC. In the course of experiments carried out on real crime data for Boston, we showed that PI-SC is up to 84% less numerous than the set of all PI-strong ST-sequential patterns. In the experiments, the runtime of CST-SPMiner was almost the same as the runtime of the STBFM algorithm, which mines PI-strong ST-sequential patterns.

References

[1] Agrawal, R., Srikant, R., 1995. Mining sequential patterns, in: Proceedings of the Eleventh International Conference on Data Engineering, pp. 3–14.

[2] Arge, L., Procopiuc, O., Ramaswamy, S., Suel, T., Vitter, J.S., 1998. Scalable sweeping-based spatial join, in: Proceedings of the 24rd International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. pp. 570–581.

[3] Aydin, B., Angryk, R.A., 2016. Spatiotemporal event sequence mining from evolving regions, in: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 4172–4177.

[4] Boston-Police-Department, 2014. Boston police department: Crime incident reports.

[5] Fournier-Viger, P., Lin, J.C.W., Kiran, R.U., Koh, Y.S., Thomas, R., 2017. A survey of sequential pattern mining. Data Science and Pattern Recognition 1, 54–77.

[6] Fumarola, F., Lanotte, P.F., Ceci, M., Malerba, D., 2016. Clofast: closed sequential pattern mining using sparse and vertical id-lists. Knowledge and Information Systems 48, 429–463.

[7] Gomariz, A., Campos, M., Marin, R., Goethals, B., 2013. Clasp: An efficient algorithm for mining frequent closed sequences, in: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (Eds.), Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 50–61.

[8] Huang, Y., Zhang, L., Zhang, P., 2008. A framework for mining sequential patterns from spatio-temporal event data sets. IEEE Transactions on Knowledge and Data Engineering 20, 433–448.

[9] Li, Z., 2014. Spatiotemporal Pattern Mining: Algorithms and Applications. Springer International Publishing, Cham. pp. 283–306.

[10] Maciąg, P.S., 2017. A survey on data mining methods for clustering complex spatiotemporal data, in: Kozielski, S., Mrozek, D., Kasprowski, P., Małysiak-Mrozek, B., Kostrzewa, D. (Eds.), Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation, Springer International Publishing, Cham. pp. 115–126.

[11] Maciąg, P.S., Bembenik, R., 2019. A novel breadth-first strategy algorithm for discovering sequential patterns from spatio-temporal data, in: Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM,, INSTICC. SciTePress. pp. 459–466.

[12] Mannila, H., Toivonen, H., Verkamo, A.I., 1995. Discovering frequent episodes in sequences, in: Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, August 20-21, 1995, pp. 210–215.

[13] Mohan, P., Shekhar, S., Shine, J.A., Rogers, J.P., 2012. Cascading spatio-temporal pattern discovery. IEEE Transactions on Knowledge and Data Engineering 24, 1977–1992.

[14] Mooney, C.H., Roddick, J.F., 2013. Sequential pattern mining – approaches and algorithms. ACM Comput. Surv. 45, 19:1–19:39.

[15] Sunitha, G., Reddy, M., Rama, A., 2014. Mining frequent patterns from spatio-temporal data sets: A survey. Journal of Theoretical & Applied Information Technology 68.

[16] Yan, X., Han, J., Afshar, R., 2003. Clospan: Mining closed sequential patterns in large datasets, in: Proceedings of the 2003 SIAM International Conference on Data Mining, pp. 166–177.