Spatio-temporal Co-occurrence Pattern Mining in Data Sets with Evolving Regions

Karthik Ganesan Pillai, Rafal A. Angryk, Juan M. Banda, Michael A. Schuh, Tim Wylie
Department of Computer Science
Montana State University
Bozeman, USA
{*k.ganesanpillai, angryk, juan.banda, michael.schuh, timothy.wylie*}*@cs.montana.edu*

***Abstract*—Spatio-temporal co-occurring patterns represent subsets of event types that occur together in both space and time. In comparison to previous work in this field, we present a general framework to identify spatio-temporal co-occurring patterns for continuously evolving spatio-temporal events that have polygon-like representations. We also propose a set of measures to identify spatio-temporal co-occurring patterns and propose an Apriori-based spatio-temporal co-occurrence mining algorithm to find prevalent spatio-temporal co-occurring patterns for extended spatial representations that evolve over time. We evaluate our framework on real-life data to demonstrate the effectiveness of our measures and the algorithm. We present results highlighting the importance of our measures in identifying spatio-temporal co-occurrence patterns.**

***Keywords*-evolving spatio-temporal events, extended spatial representations, spatio-temporal co-occurring patterns**

I. Introduction

Spatio-temporal pattern mining in data sets with evolving extended spatial representations is an important problem for many application domains such as weather monitoring, astronomy, and solar physics - which is our application focus. Spatio-temporal co-occurring patterns frequently occur among various solar events. Fig. 1 shows two types of solar phenomena, Filaments (green) and Sigmoids (red) in spatial context with their corresponding shapes and bounding boxes occurring at different locations on the Sun. As seen in Fig. 1, the shapes of the solar events are represented in extended spatial representations as polygons (each event is represented by a contour and a box enclosing the contour), and the shape, size, and location of the solar events continuously evolve over time. All of these factors influence relationships between various solar events, which lead to complex spatial and temporal interactions. Mining spatio-temporal co-occurring patterns on the Sun could help us better understand relationships between solar events and lead to better modeling and forecasting of important events such as solar flares and coronal mass ejections, which impact our safety on Earth.

In this paper we use the concepts of spatio-temporal predicates introduced in [7] to define spatio-temporal co-occurring patterns, and propose a novel look at the interest measures to find prevalent (frequent) spatio-temporal co-occurring patterns in data sets with region-based spatial representations that evolve over time.

The major contributions of this paper are: 1) We present a novel framework for mining region-based prevalent spatio-temporal co-occurrence patterns. 2) Event types are modeled as $3D$ spatial objects to capture different spatial and temporal characteristics of evolving extended spatial representations. 3) We define multiple measures to capture the strength of spatio-temporal relation Overlap occurrence between instances of different event types based on volumes resulting from Overlap. 4) We present an Apriori-based [2] algorithm to discover prevalent spatio-temporal co-occurrence patterns in data sets with evolving extended spatial representations.

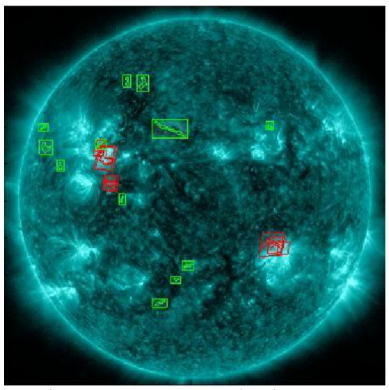

Figure 1: Image showing two types of solar events - Filaments are marked in green, and Sigmoids are marked in red.

The rest of the paper is organized as follows: Section II gives a background on related work. We explain important concepts of temporal relations, and spatio-temporal predicates that are crucial for understanding our work in Section III. We define our framework for finding spatio-temporal co-occurring patterns in data sets with evolving extended spatial representations in Section IV. Finally, we present a variety of experiments demonstrating the effectiveness of our approach, concluding with a summary of results and future work.

II. Related Work

Since spatio-temporal data mining is an important area, many algorithms have been proposed in literature for co-location mining in spatio-temporal databases: Topological Pattern Mining [12], Co-location Episodes [4], Mixed Drove Co-occurrence Mining [5], Spatial Co-location Pattern Mining from extended spatial representations [13], Spatio-temporal Pattern Mining in scientific data [14], and Interval Orientation Patterns [10]. In this section we review the work for co-location pattern mining in spatial and spatio-temporal

databases.

Mining topological patterns, also called co-location patterns, from spatio-temporal databases was introduced by Wang *et al.* in [12]. In this paper, the authors used a summary-structure to approximate the instance counts of a co-location pattern. The authors also introduced the TopologyMiner algorithm to discover the co-location patterns. The algorithm discovers frequent co-location patterns in a depth-first manner. The TopologyMiner algorithm divides the search space into a set of partitions, and then in each partition it uses a set of locally frequent features to grow patterns.

Cao *et al.* introduced the problem of mining co-location episodes in spatio-temporal data [4]. In this paper, the authors define a *co-location episode* as a sequence of co-location patterns with some common feature type across consecutive time slots. The authors also introduced a two-step framework for mining co-location episodes. In the first step of the framework, object pairs of different feature types (f_i, f_j) that have close concurrent subsequences are identified. In the next step of the framework, the authors use an Apriori-based [2] technique to discover the frequent episodes.

Celik *et al.* introduced the problem of mining mixed-drove spatio-temporal co-occurrence patterns (MDCOPs) in spatio-temporal data [5]. In this paper, the authors define MDCOP as a subset of spatio-temporal mixed feature types whose instances are neighbors in space and time. They introduced the MDCOP-Miner algorithm, which extends standard spatial co-location mining algorithm [8] to include time information. The algorithm first discovers all *size-(k)* spatial prevalent MDCOPs, and then applies a time-prevalence based filtering to discover MDCOPs. Finally, the MDCOP-Miner algorithm generates *size-$(k+1)$* candidate MDCOPs using *size-(k)* MDCOPs.

Xiong *et al.* introduced the problem of mining spatial co-location patterns from extended spatial representations in [13]. In their buffer-based model the neighborhood of an extended spatial representation is defined by the spatial buffer operation. The authors introduced the EXCOM algorithm to find spatial co-location patterns in data sets with extended spatial representations. The EXCOM algorithm first uses a geometric filter step to eliminate a lot of feature sets that can not form co-location patterns. In the next step, an Apriori-based approach is used to generate spatial co-location patterns.

Mining spatio-temporal patterns in scientific data was first introduced by Yang *et al.* in [14]. In this paper, the authors introduce a general framework to discover spatial associations and spatio-temporal episodes for scientific data sets. The authors model features as geometric objects rather than points. They also extend their approach to accommodate temporal information and propose an algorithm to derive spatio-temporal episodes.

The problem of mining interval orientation patterns in spatio-temporal databases was introduced by Patel in [10]. In this work, the author modeled features by taking feature duration into account. Thus, the approach introduced was able to capture the temporal influence of a feature on other features within a spatial neighborhood. An Interval Orientation (IO) pattern is a frequent sequence of features with annotations of temporal and directional relationships between every pair of features. The author introduced an algorithm called IOMiner to mine frequent IO patterns. The algorithm uses a two-stage procedure to find IO patterns. In the first stage, a disjoint cubes hashing [12] is used to find IO patterns of size two. In the second stage, a hash-based join is used to find IO patterns of size three and more.

Methods available in the current literature consider features as spatial point representations with temporal information, or consider features as extended spatial representation with temporal information but do not take features' duration into account. Thus, these methods are not adequate for mining spatio-temporal co-occurrence patterns on data sets with extended spatial representations that evolve over time.

III. Important Concepts and Problem Statement

In this section, we formulate the problem of mining spatio-temporal co-occurring patterns on regions that evolve over time using spatio-temporal predicates to define the evolving regions' neighborhoods.

Allen introduced interval based temporal logic in [3]. The paper also introduces six asymmetric temporal relations and one symmetric temporal relation. These temporal relations (all 13 of them, i.e., 6 asymmetric and 1 symmetric) can be used to capture the relations between two time intervals.

However, for this work we are interested in finding spatio-temporal co-occurring patterns satisfying only a specific subset of Allen's temporal relations: *equal, meets, overlaps, during, starts,* and *finishes*. We only use one general spatial predicate: *spatial intersects* (see Fig. 2).

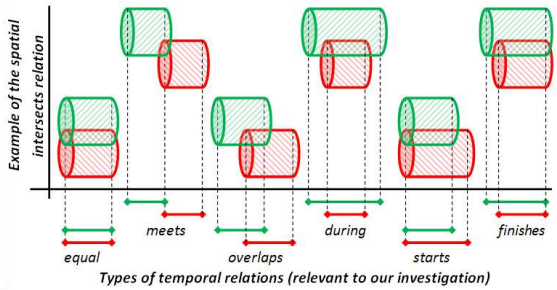

Figure 2: Two evolving polygons satisfying spatial intersects and temporal relations that are important for our investigation

Erwig and Schneider [7] presented a convenient way of thinking about spatio-temporal predicates by applying the idea of temporal lifting to spatial predicates. To distinguish spatio-temporal predicates from spatial predicates, following Erwig and Schneider's notation, we refer to spatio-temporal predicates by using a capital letter (to begin the word) and spatial predicates by using small letters. Often an evolving

spatial region can be represented as a three-dimensional object in three-dimensional space, where two dimensions represent spatial characteristics of the object, and the third dimension represents time. Fig. 2 represents some examples fulfilling the spatio-temporal relation "*Overlap*". In three-dimensional space, a moving point can be represented by a curve [7] and two co-occurring polygon-like objects can be represented as types with Overlapping trajectories.

A. Spatio-temporal Co-occurring Patterns

Let $E = \{e_1, \ldots, e_M\}$ be a set of spatio-temporal event types, and a set of instances of these event types, which evolve over time, $I = \{i_1, \ldots, i_N\}$ such that $M \ll N$. A spatio-temporal co-occurring pattern is a subset of spatio-temporal event types that co-occur in both space and time.

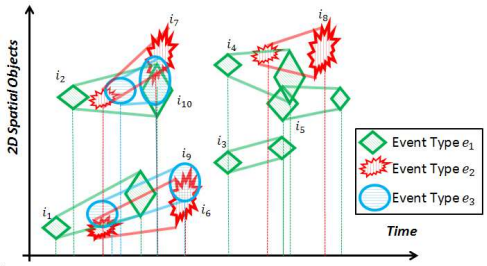

Figure 3: An example of spatio-temporal data

Table I: Temporal information about event instances of data shown in Fig. 3

Instance ID	Event Type	Start Time (HH:MM)	End Time (HH:MM)
i_1	e_1	10:00	10:30
i_2	e_1	10:10	10:40
i_3	e_1	11:00	11:20
i_4	e_1	11:00	11:30
i_5	e_1	11:20	11:50
i_6	e_2	10:20	10:50
i_7	e_2	10:20	10:40
i_8	e_2	11:20	11:40
i_9	e_3	10:20	10:50
i_{10}	e_3	10:30	10:40

Fig. 3 shows an example data set that we will use to explain our definitions in detail. In Tab. I, we show the instance ID, start time, and end time of instances of different event types from our example data set shown in Fig. 3. This example data set contains three event types. The event type e_1 has a total of five spatio-temporal instances occurring at various locations that evolve over time. Event type e_2 has a total of three spatio-temporal instances and event type e_3 has a total of two spatio-temporal instances occurring at various locations that evolve over time. For simplicity, in this example data set we did not show the values of sequences of $2D$ shapes of the evolving instances of event types e_1, e_2, and e_3. In our example, $E = \{e_1, e_2, e_3\}$ and $M = 3$, whereas $N = 10$, all instance IDs listed in the first column of Tab. I.

Definition 1. A *size-(k)* spatio-temporal co-occurrence is denoted as $SE = \{e_1, \ldots, e_k\}$, where $SE \subseteq E$, $SE \neq \emptyset$ and $1 \leq k \leq M$.

We can have multiple *size-(k)* spatio-temporal co-occurrences derived from the set E, so to separate them we will subscript future definitions (e.g., SE_i) with an arbitrarily chosen subscript to denote uniqueness, i.e., $SE_i \neq SE_j$. Note that indices (i or j) do not indicate the size of the co-occurrence. For the size we reserve the symbol k.

Definition 2. $pat\_instance$ is a *pattern instance* of a spatio-temporal co-occurrence SE_i if $pat\_instance$ contains an instance of all events in SE_i and no proper subset of $pat\_instance$ is also a *pattern instance*.

For example, $\{i_1, i_6, i_9\}$ is a *size*-3 ($k = 3$) pattern instance of co-occurrence $SE_i = \{e_1, e_2, e_3\}$ in the example spatio-temporal data set presented in Fig. 3 and Tab. I.

Definition 3. A collection of pattern instances of SE_i is a *table instance* of SE_i, and is denoted as $tab\_instance(SE_i)$.

For example, $\{\{i_1, i_6, i_9\}, \{i_2, i_7, i_{10}\}\}$ is a *size*-3 ($k = 3$) $tab\_instance(SE_i = \{e_1, e_2, e_3\})$ in the example spatio-temporal data set presented in Fig. 3 and Tab. I.

Definition 4. A spatio-temporal co-occurring prevalent pattern is of the form $SE_i(cce, p)$, where SE_i is a spatio-temporal co-occurrence, and parameters cce, p characterize the prevalent pattern in the following manner.

1) cce is an indicator of the strength of spatio-temporal relation's occurrence that is investigated. For our application we used spatio-temporal Overlap. Some examples of spatio-temporal Overlap are $\{i_1, i_6\}$, $\{i_2, i_7\}$, and $\{i_7, i_{10}\}$ as shown in Fig. 3. We will discuss this more in detail in Sec. *III.B*,
2) p is the *prevalence measure*. The *prevalence measure* emphasizes how interesting the spatio-temporal co-occurrences are. In our investigation we used the participation index (pi) [8] as the *prevalence measure*. The participation index monotonically decreases when the size of the spatio-temporal co-occurrence pattern increases, which can be exploited for computational efficiency [8].

B. Our Measures

To calculate cce (in our case the strength of spatio-temporal Overlap) of a *size-(k)* spatio-temporal co-occurrence SE_i, we introduce a spatio-temporal co-occurrence co-efficient. Our spatio-temporal co-occurrence co-efficient is closely related to the co-efficient of areal correspondence (CAC) proposed in [11]. CAC is computed for any two (or more, for longer patterns) overlapping polygons as the area of intersection, divided by the area of union. We extend CAC to three dimensions (two dimensions correspond to space and the third dimension corresponds to time), and calculate the spatio-temporal co-occurrence co-efficient based on volumes.

Definition 5. Spatio-temporal Intersection volume (I_v) of a $pat\_instance$: The I_v for a pattern instance is the volume of $3D$ object resulting from Intersection of trajectories of all instances of spatio-temporal event types in a pattern instance.

Definition 6. Spatio-temporal Union volume (U_v) of a $pat\_instance$: The U_v for a pattern instance is the volume of $3D$ object resulting from Union of trajectories of all instances of spatio-temporal event types in a pattern instance, where all the trajectories are limited in time to the interval of their common existence. See Fig. 4 for an example.

Spatio-temporal co-occurrence coefficient (cce) could be calculated for a $pat\_instance$ as the ratio $\frac{I_v}{U_v}$. The ratio $\frac{I_v}{U_v}$ represents the *Jaccard Coefficient* [9]. However, we show that multiple measures can be utilized as cce, which will be shown in Sec. *V.B*.

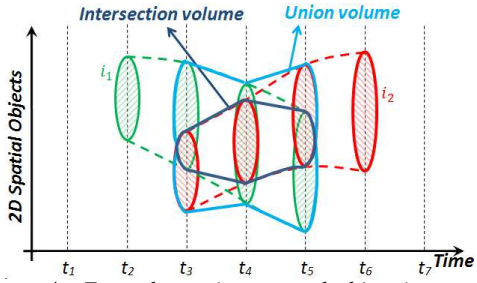

Figure 4: Example spatio-temporal object interaction

Computing cce for extended spatio-temporal representations such as moving polygons is not a trivial task. In Fig. 4 we show the movement of a pair of instances of two event types that changes shape, size, and direction across different time slots. We also show the region of Intersection and the region of Union at different time slots. Moreover, the volumes resulting from the Intersection (our I_v) and Union (our U_v) trajectories are shown in the Fig. 4. We give a detailed description of our approach to calculate cce in $3D$ spatial objects in Sec. *IV*.

Definition 7. The participation index $pi(SE_i)$ of a spatio-temporal co-occurrence SE_i is,

$$min_{j=1}^{k} pr(SE_i, e_j) \quad (1)$$

where k is the length of the pattern (i.e., cardinality of SE_i, $|SE_i|$), and the participation ratio $pr(SE_i, e_j)$ for a spatio-temporal event type e_j is the fraction of the total number of instances of e_j forming spatio-temporal co-occurring instances in SE_i.

The value of cce_{SE_i} of all pattern instances in a table instance of a spatio-temporal co-occurring pattern SE_i, monotonically decreases with the *size-k* of the spatio-temporal co-occurrence pattern increasing. However, we do not use cce_{SE_i} as a measure of how interesting a spatio-temporal co-occurrence pattern is. cce_{SE_i} does not consider the total number of instances of event types forming spatio-temporal co-occurring instances in SE_i and it does not provide any information about the statistics of the relationship. Thus, we use participation index to capture how interesting a spatio-temporal co-occurrence pattern is.

C. Problem Statement

Given:

1) A set of spatio-temporal event types $E = \{e_1, e_2, \ldots, e_M\}$ over a common spatio-temporal framework.
2) A set of N event instances $I = \{i_1, i_2, \ldots, i_N\}$, which evolves over space and time such that $M \ll N$, and each $i_j \in I$ is a tuple $<$ *instance-id, spatio-temporal event type, sequence of 2D shapes, start time, end time* $>$ over a common spatio-temporal framework, where start and end time reflect lifetime of each event, and the sequence of shapes reflect the time-evolving spatial event.
3) A user-specified spatio-temporal co-occurrence coefficient threshold (cce_{th}).
4) A user-specified participation index threshold (pi_{th}), that we use as our prevalence measure.
5) A time interval of data sampling denoted t_s. All events are sampled with the same interval making the shapes of individual events exactly aligned in time. We can change granularity through the modification of our time sampling interval.

Objective: *Find the complete and correct result set of spatio-temporal co-occurring patterns with* $cce > cce_{th}$ *and* $pi \geq pi_{th}$.

The selection of threshold values will influence the computational cost associated with mining spatio-temporal co-occurring patterns. Small threshold values can usually generate many patterns, but will increase the computational cost. Large threshold values can prune a lot of significantly interesting patterns, reducing the computational cost, but a small set of highly frequent patterns may be not interesting for the user, due to the reporting of patterns that are so frequent that they are already well-known to the domain experts. Thus, selection of threshold values largely depends on the purpose of the analysis.

IV. Apriori-based Spatio-temporal Co-occurrence Miner For Evolving Shapes

In this section we first propose a geometric solution to calculate cce for a pattern instance. Then, we introduce an algorithm to mine spatio-temporal co-occurring patterns inspired from the spatial co-location algorithm introduced in [8].

Table II: Calculation of cce from $tab\_instance(e_1, e_2)$ *derived from data shown in Fig. 4*

Instance of e_1	Instance of e_2	Time Instant	Union area (U_a)	Intersection area (I_a)	$\frac{I_a}{U_a}$
i_1	i_2	t_3	100	20	0.2
i_1	i_2	t_4	80	60	0.75
i_1	i_2	t_5	120	30	0.25

A. 3D Geometric Challenges

In Fig. 4 we show an example movement of a pair of instances of two event types that change shape, size, and direction during their lifetime. We will use this example to explain our steps to calculate cce for pattern instances of a spatio-temporal co-occurrence pattern. In Fig. 4 we also show the regions of Intersection and Union at different time

slots, as well as the volumes resulting from the Intersection and Union of instances' trajectories (I_v,U_v, respectively).

Input :

(1) E= A set of spatio-temporal event types, which can be represented as $2D$ shapes at each time step.

(2) I= *<instance-id, spatio-temporal event type, sequence of 2D shapes, start time, end time>*.

(3) User-specified thresholds: minimum spatio-temporal co-occurrence coefficient cce_{th} and minimum participation index pi_{th}.

(4) A user-specified sampling interval (t_s), measured as duration between snapshots of evolving objects.

Output :

A set of spatio-temporal co-occurring patterns with cce and pi greater than the user-specified minimum threshold values given on input.

Variables :

(1) k the co-occurrence size (see Def. 1).

(2) C_k: a set of candidates for size-(k) spatio-temporal co-occurring patterns derived from *size-($k-1$)* prevalent spatio-temporal co-occurring patterns.

(3) T_k: set of instances of size-(k) spatio-temporal co-occurrences (see Def. 3).

(4) P_k: a set of *size-(k)* prevalent spatio-temporal co-occurring patterns derived from *size-(k)* candidate spatio-temporal co-occurring patterns (see Def. 4).

(6) P_{final}: union of all prevalent spatio-temporal co-occurring patterns (patterns of all sizes).

Algorithm :

1 k=1, C_1=E, $P_1 = E$, $P_{final} = \emptyset$;
2 $T_1 = gen\_loc(C_1, I, t_s)$;
3 **while** ($P_k \neq \emptyset$) {
4 $C_{(k+1)} = gen\_candidate\_coocc(P_k)$;
5 $T_{(k+1)} = gen\_tab\_ins\_coocc(C_{(k+1)}, cce_{th})$;
6 $P_{(k+1)} = pre\_prune\_coocc(C_{(k+1)}, pi_{th})$;
7 $P_{final} = P_{final} \cup P_{(k+1)}$;
8 $k = k + 1$;
9 }
10 return P_{final};

Figure 5: Spatio-temporal Co-occurrence Mining Algorithm

We calculate cce for instances of a spatio-temporal co-occurrence pattern during the time interval of their Overlap as follows. First, we calculate the ratio of Intersection and Union area between instances of spatio-temporal events in a *pattern instance* at each time stamp. In Tab. II, we show Union area, Intersection area, and ratio of Intersection to Union area for time stamps t_3, t_4, and t_5, during which instances i_1 and i_2, shown in Fig. 4, satisfy the spatio-temporal relation Overlap. Second, cce for each pattern instance (i_1, i_2 in our example) is calculated by computing the average ratio of Intersection and Union area across all time slots. For the example instances shown in Fig. 4, cce can be calculated from averaging the values of the last column of Tab. II, so cce is equal to 0.4. Note, all our evolving shapes are sampled at the same interval; thus, no weights are necessary to calculate average.

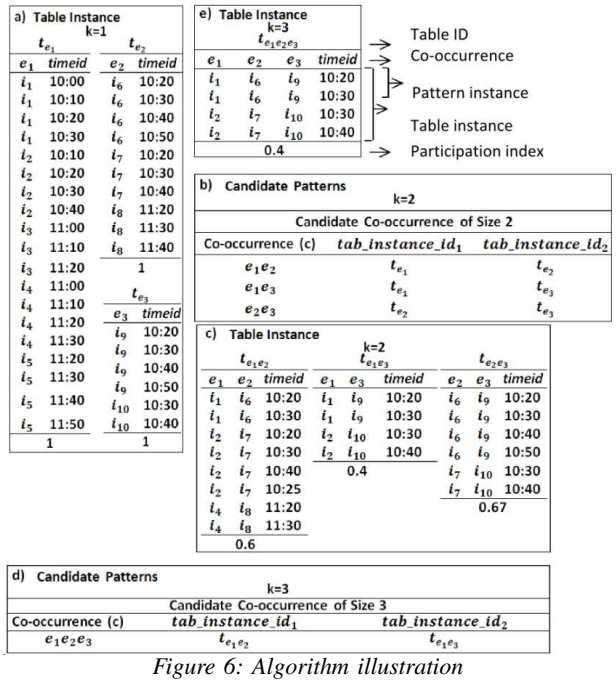

Figure 6: Algorithm illustration

We also save the geometric shapes resulting from the Intersection and Union operations. These geometric shapes will be used for finding the cce of spatio-temporal co-occurring patterns of size three or more. The computation cost involved in finding cce will directly depend on the interval of sampling (t_s) used for the calculation. However, sampling interval t_s can be tuned to a specific user analysis.

B. Implementation Details

Fig. 5 gives the pseudocode of our spatio-temporal co-occurrence pattern mining algorithm. The inputs are as defined at the beginning of Sec. *III.C*. In the algorithm, steps 1 and 2 initialize the parameters and data structures, steps 3 through 9 give an iterative process to mine spatio-temporal co-occurring patterns, and step 10 returns a union of the results of the spatio-temporal co-occurring patterns (rules of all size). Steps 3 through 9 continue until there are no candidate spatio-temporal co-occurring patterns to be mined. Next we explain the functions in the algorithm.

Step 2, i.e., $T_1 = gen\_loc(C_1, I, t_s)$: In this step, we generate table instances of size one. This function is similar to the table instance initialization step of the spatial co-location miner algorithm [8]. However, this function takes an additional argument t_s, which represents the increment in number of time steps. In this method we project the movement of instances of events from its start to end time slot, using t_s to increment the number of time steps between time slots. The combination of the event instance ID and the time stamp will allow us to uniquely identify an event

instance at a particular moment.

For example, Fig. 6 (a) shows the key columns of table instances of size one for our sample spatio-temporal data set from Fig. 3 and Tab. I. Here, the t_s value was set to 10 minutes. The column denoted t_{e_1} represents the table instance of size one for event type e_1. Similarly, the columns denoted t_{e_2} and t_{e_3} represent the tables instances of size one for event types e_2 and e_3. The columns: *union geometry*, *intersection geometry*, and cce are not shown in Fig. 6 (a) for simplicity.

Step 4, i.e., $C_{(k+1)} = gen\_candidate\_coocc(P_k)$: In this step, we generate candidate spatio-temporal co-occurring patterns. An Apriori-based [2] approach is used to generate the candidates of size-$(k+1)$ using spatio-temporal co-occurring prevalent patterns of *size-(k)* [8]. Thus, prevalent patterns of *size-(k)*, which satisfy the user-specified threshold value of a minimum participation index pi_{th}, are used to generate candidate patterns of size-$(k+1)$.

For example, Fig. 6 (b) and (d) show the candidate co-occurrence patterns of size $k = 2$ and $k = 3$, respectively, for our example spatio-temporal data set (Fig. 3 and Tab. I).

Step 5 i.e., $T_{(k+1)} = gen\_tab\_ins\_coocc(C_{(k+1)}, cce_{th})$: In this step, we generate table instances for candidates of *size-$(k+1)$*. This method is similar to table instances of candidate locations in [8], but include two additional calculations. The first is a check of the time slot ID between events instances, and the second saves the resulting geometries of Intersection and Union from the instances of *size-(k)* spatio-temporal co-occurring patterns. Thus computation for *size-$(k+1)$* candidates can be expressed as spatio-temporal join query shown in Fig. 7.

forall co-occurrence $c \in C_{k+1}$
  insert into T_c /* T_c is the table instance of co-occurrence c */
  select $p.e_1, p.e_2, \ldots, p.e_k,$
    $q.e_k,$
    $p.timeid, uge(p.ug, q.ug), ige(p.ig, q.ig),$
    $area(ige(p.ig, q.ig))/area(uge(p.ug, q.ug))$
  from $c.tab\_instance\_id_1\ p, c.tab\_instance\_id_2\ q$
  where $p.e_1 = q.e_1, \ldots,$
    $p.e_{k-1} = q.e_{k-1},$
    $p.timeid = q.timeid;$
end;

Figure 7: Spatio-temporal JOIN Query

In Fig. 7, $uge(.,.)$ is a function that takes two geometries as arguments and returns the Union of those two geometries, and $ige(.,.)$ is a function that takes two geometries as arguments and returns the Intersection of those two geometries, ug and ig represent Union and Intersection geometry. $area(.)$ is a function that takes a geometry as an argument and returns the area of that geometry.

The spatio-temporal co-occurrence coefficient for each pattern instance is calculated by summing the ratios of the area of intersection to union across all the time slots, and then averaging the sum by the total number of time slots. Pattern pairs that have cce below the user-specified cce_{th} value are deleted from the table instance.

For example, Fig. 6 (c) and (e) show the table instances of sizes two and three for our example spatio-temporal data set (Fig. 3 and Tab. I).

Step 6, i.e., $P_{(k+1)} = pre\_prune\_coocc(C_{(k+1)}, pi_{th})$: In this step we find prevalent spatio-temporal co-occurring patterns. Spatio-temporal co-occurring patterns P of *size-$(k+1)$* are found by pruning $C_{(k+1)}$ that have spatio-temporal co-occurring $pi < pi_{th}$ [8].

As defined and shown with the example in Sec. *III.B* (Def. 7), participation index for spatio-temporal co-occurrence pattern SE_i is calculated using Eq. 1.

For example, Fig. 6 (c), shows participation index values with candidate co-occurrence of size two. As shown in the figure, the candidate pattern $SE_i = \{e_1, e_3\}$ will be pruned if a value of 0.5 is set to pi_{th} and the algorithm will stop.

Step 7 i.e., $P_{final} = P_{final} \cup P_{(k+1)}$ of the algorithm calculates the union of prevalent patterns P_{final} and P_{k+1}. The algorithm runs iteratively until no more spatio-temporal co-occurring patterns can be generated, and returns the union of all the found prevalent spatio-temporal co-occurring patterns in Step 10.

V. Experimental Evaluation

In this section, we evaluate our framework on a real data set from solar physics domain. Specifically, we evaluate our framework using six types of evolving solar phenomena and four different measures (used to evaluate our cce). We report results based on performance and impact of our measures on capturing the different spatial and temporal characteristics of the real spatio-temporal data set.

A. Real-life Data Set

Our data set contains evolving instances of six different solar event types. We obtained our data set from a real data repository called Heliophysics Event Knowledgebase (HEK) [1]. Our data set consists of 1228 instances of *Active region*, 708 instances of *Emerging Flux*, 681 instances of *Filament*, 1297 instances of *Flare*, 531 instances of *Sigmoid*, and 308 instances of *Sunspot* for the period 01/01/2012 through 01/31/2012. The real data set represents significantly different spatial and temporal characteristics of the six different solar event types (some events (e.g., *Active Region*) are large and long lasting, while others (e.g., *Emerging Flux*) are relatively small with short duration).

For all the experiments, sampling interval t_s was set to 10 minutes. All experiments were performed using PostgreSQL 9.1.4 and PostGIS 1.5.4.

B. Experimental Design

To accurately capture the different spatial and temporal characteristics of co-occurrences of the six different solar events, represented as evolving polygons, where each instance of these events has different spatial size and duration of life-

time, we investigated four different measures applicable for our algorithm's cce parameter (Tab. III).

Table III: Measures evaluating spatio-temporal relation Overlap (our cce)

Name	Formula	Output Range
Overlap coefficient	$\frac{Volume(i_1 \cap i_2)}{\min(Volume(i_1), Volume(i_2))}$	$[0, 1]$
Cosine coefficient	$\frac{Volume(i_1 \cap i_2)}{\sqrt{Volume(i_1) \times Volume(i_2)}}$	$[0, 1]$
Dice coefficient	$\frac{2 \times Volume(i_1 \cap i_2)}{Volume(i_1) + Volume(i_2)}$	$[0, 1]$
Jaccard (or Coherence or Tanimoto) coefficient	$\frac{Volume(i_1 \cap i_2)}{Volume(i_1 \cup i_2)}$	$[0, 1]$

The first column of Tab. III shows the name of the measure, and the second column shows the formula to calculate the measure in order to characterize the strength of spatio-temporal Overlap for any two spatio-temporal co-occurrence patterns. The third column of Tab. III shows the range of the output of the coefficients. Since the output ranges are identical, we could run our experiments without modification of cce_{th}.

The formulas take $3D$ objects as inputs (denoted i_1 and i_2 in Tab. III) representing instances of $2D$ spatial objects that evolve over time (our $3D$) and assess the strength of their spatio-temporal Overlap by comparing volumes of the objects' Overlapping trajectories. Note, cce as originally explained in Sec. *III.B* is matching the *Jaccard coefficient* [9] in Tab. III. However, as we will show below, all of the measures listed in Tab. III can be used to evaluate the strength of spatio-temporal Overlap relation. Furthermore, each of the measures has different properties. For instance, the *Jaccard coefficient* acts similar to *Dice coefficient* [9]; however, it penalizes objects with smaller Intersection volumes (i.e., it gives much lower values than *Dice* to objects which have small Intersection (common) volume - giving a penalty to some of our events that are small in the area and short-lasting. *Overlap coefficient* [9] gives a value of one if one object is totally contained with the other. We could say that it reflects inclusion (i.e., *Overlap coefficient* value is a maximum value when $i_1 \subseteq i_2$ or $i_2 \subseteq i_1$), which benefits the objects that are almost equal in space and time. *Cosine coefficient* [9] is more resistant to size of the objects, making it more appropriate for our real-life data set that contains events with drastically different life spans and area (sizes). We report the impact of the measures on capturing the different spatial and temporal characteristics by the counts of spatio-temporal co-occurrence instances found and the number of unique prevalent spatio-temporal co-occurrence patterns generated (as shown in Fig. 8).

C. Impact of Measures

In this section, we present the results of impact of our measures for the characterization of spatio-temporal Overlap that were presented in Tab. III. We compare the count of spatio-temporal co-occurrence pattern instances found,

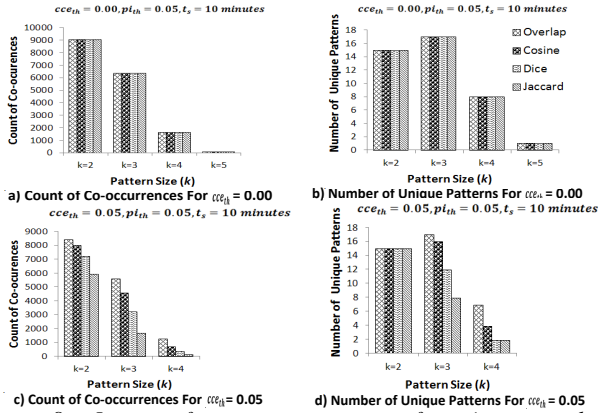

Figure 8: Impact of measures on count of spatio-temporal co-occurrence instances found and number of unique patterns found

and also compare the number of unique prevalent spatio-temporal co-occurrence patterns found for the four different measures. Fig. 8 (a) and (c) show the impact of our measures on the count of spatio-temporal co-occurrence pattern instances for different spatio-temporal pattern sizes (our k) with $cce_{th} = 0.00$ and $cce_{th} = 0.05$, respectively, while pi_{th} and t_s are unchanged. Fig. 8 (b) and (d) show the impact of our measures on the number of unique prevalent spatio-temporal co-occurrence patterns found for different spatio-temporal pattern sizes with $cce_{th} = 0.00$ and $cce_{th} = 0.05$, respectively, while pi_{th} and t_s are unchanged.

As expected for values of *Overlap, Cosine, Dice,* and *Jaccard coefficient* > 0.00, all the measures generated the same number of spatio-temporal co-occurrence pattern instances which led to the same number of unique prevalent spatio-temporal patterns as shown in Fig. 8 (a) and (b). However, in Figs. 8 (c) and (d), we show the number of spatio-temporal co-occurrence pattern instances found and the number of unique prevalent spatio-temporal patterns found for values of *Overlap, Cosine, Dice,* and *Jaccard coefficient* > 0.05. Fig. 8 shows the relation between different measures in capturing the spatio-temporal co-occurrence patterns. An interesting ordering relation on the selectivity of the boolean versions of *Overlap, Cosine, Dice, and Jaccard coefficients* was shown in [6]. We experimentally confirm it for the real positive numbers that reflect volumes (please see Fig. 8 (c) and (d)). Thus, the number of spatio-temporal co-occurrence patterns generated follows the order *Overlap coefficient* $\geq$ *Cosine coefficient* $\geq$ *Dice coefficient* $\geq$ *Jaccard coefficient*, and this is evident from the results of Fig. 8 (c) and (d). This is a useful property for estimating selectivity of generated patterns, since with the same values of parameters, we now know that any of the measures cannot generate more confidence for our values than the *Overlap coefficient*.

Next we looked into the influence of our different measures (i.e., cce) on solar events with drastically different spatial and temporal characteristics. First we picked a solar event type *Active Region* that has a large volume. *Active Region* is large (spatially) in comparison to other five solar

events of our data set, and it has a longer lifetime. Second we picked a solar event type *Emerging flux* that has a small volume. *Emerging Flux* is small (spatially and temporally) in comparison to the other five solar events of our data set.

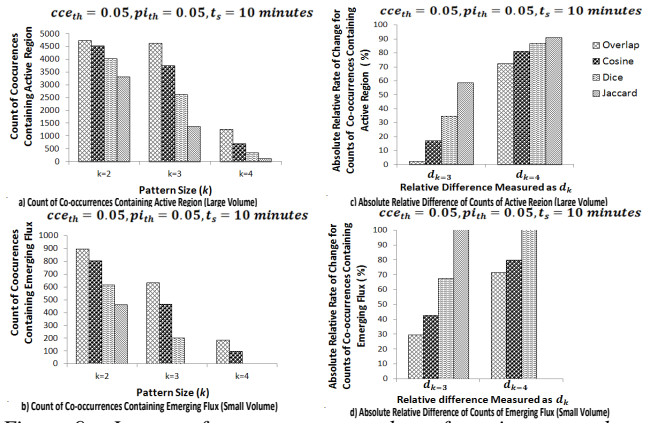

Figure 9: Impact of measures on number of spatio-temporal co-occurrence instances found and relative difference measured as d_k

In Fig. 9 (a) and (b) we show results on the number of co-occurrence instances in patterns that contain *Active Region* and *Emerging Flux*, with the same setup as earlier. Since the count of instances of both these events (we had 1228 *Active Regions* and 708 *Emerging Flux*) are different, we do not only present counts of co-occurrences that contain these events, but also ratio of these changes as shown in Fig. 9 (c) and (d). From the Fig. 9 (b) and (d), we observe that the *Jaccard* measure penalizes patterns with *Emerging Flux* (relatively small volumes) more than *Overlap* and *Cosine coefficient*. As *Cosine coefficient* is more resistant to the size of the objects, we were able to find more co-occurrences instances of *Emerging Flux* than *Dice* and *Jaccard coefficient*. Furthermore, we were able to find more co-occurrence instances with *Overlap coefficient* as it does not penalize objects with smaller intersection (common) volumes. A similar argument holds for co-occurrences containing *Active Region* as shown in Fig. 9 (a) and (c). Moreover, we can observe from Fig. 9 (c) and (d) the rate of change of different measures vary for co-occurrences containing *Active Region* and *Emerging Flux*. Again, this reflects the property of measures on co-occurrences containing the events of large and small volumes.

VI. Conclusion and Future Work

Empirical results on real life data set, drawn from the solar physics discipline, serve to validate the framework and the impact of our proposed measures on the number of spatio-temporal co-occurrence instances found and the number of unique prevalent spatio-temporal patterns found. From the results of Fig. 8, we showed the relation between *Overlap, Dice, Cosine,* and *Jaccard coefficients* in capturing the spatio-temporal co-occurrence patterns. Furthermore, we used Fig. 9 to discuss the behavior of our measures on two solar events that have drastically different spatial and temporal characteristics. Moreover, we demonstrated through our experiments the importance of choosing appropriate measures for identifying spatio-temporal co-occurrence patterns in data sets with evolving spatial representations.

For future work, we plan to investigate using a combination of our measures (i.e., multiple cce's) for different spatio-temporal region-based co-occurrence patterns. This would help us choose measures based on the different event types present in the spatio-temporal co-occurrence pattern. We also plan to investigate new computationally efficient algorithms for mining spatio-temporal co-occurrence patterns. One approach is to use geometric filter and refine paradigm to eliminate event sets that cannot form spatio-temporal co-occurrence patterns, thus reducing the number of geometric overlay operations [13].

VII. Acknowledgments

This work was supported by two National Aeronautics and Space Administration (NASA) grant awards, 1) No. NNX09AB03G and 2) No. NNX11AM13A.

References

[1] Heliophysics events registry http://www.lmsal.com/isolsearch, Feb. 2012.

[2] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2):207–216, June 1993.

[3] J. F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, Nov. 1983.

[4] H. Cao, N. Mamoulis, and D. W. Cheung. Discovery of collocation episodes in spatiotemporal data. In *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, pages 823–827, Washington, DC, USA, 2006. IEEE Computer Society.

[5] M. Celik, S. Shekhar, J. P. Rogers, J. A. Shine, and J. S. Yoo. Mixed-drove spatio-temporal co-occurence pattern mining: A summary of results. In *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, pages 119–128, Washington, DC, USA, 2006. IEEE Computer Society.

[6] L. Egghe and C. Michel. Strong similarity measures for ordered sets of documents in information retrieval. *Inf. Process. Manage.*, 38(6):823–848, Nov. 2002.

[7] M. Erwig and M. Schneider. Spatio-temporal predicates. *IEEE Trans. on Knowl. and Data Eng.*, 14(4):881–901, July 2002.

[8] Y. Huang, S. Shekhar, and H. Xiong. Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(12):1472–1485, 2004.

[9] C. D. Manning and H. Schütze. *Foundations of statistical natural language processing*. MIT Press, Cambridge, MA, USA, 1999.

[10] D. Patel. Interval-orientation patterns in spatio-temporal databases. In *DEXA (1)*, pages 416–431, 2010.

[11] P. Taylor. *Quantitative Methods in Geography: An Introduction to Spatial Analysis*. Houghton Mifflin, 1977.

[12] J. Wang, W. Hsu, and M. L. Lee. A framework for mining topological patterns in spatio-temporal databases. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, CIKM '05, pages 429–436, New York, NY, USA, 2005. ACM.

[13] H. Xiong, S. Shekhar, Y. Huang, V. Kumar, X. Ma, and J. S. Yoo. A framework for discovering co-location patterns in data sets with extended spatial objects. In *SDM*, 2004.

[14] H. Yang, S. Parthasarathy, and S. Mehta. A generalized framework for mining spatio-temporal patterns in scientific data. In *Patterns in Scientific Data, ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pages 716–721, 2005.