

APPLICATION OF SPATIOTEMPORAL ASSOCIATION RULES ON SOLAR DATA TO SUPPORT SPACE WEATHER FORECASTING

Carlos Roberto Silveira Junior¹, José Roberto Cecatto², Marilde Terezinha Prado Santos¹ and Marcela Xavier Ribeiro¹

¹ Department of Computing, Federal University of São Carlos, São Carlos, Brazil

² National Institute of Space Research, São José dos Campos, Brazil

ABSTRACT

It is well known that solar energetic phenomena influence the Space Weather, in special those directed to the Earth environment. In this context, the analysis of Solar Data is a challenging task, particularly when are composed of Satellite Image Time Series (SITS). It is a multidisciplinary domain that generates a massive amount of data (several Gigabytes per year). It includes image processing, spatiotemporal characteristics, and the processing of semantic data. Aiming to enhance the SITS analysis, we propose an algorithm called "Miner of Thematic Spatiotemporal Associations for Images" (MiTSAI), which is an extractor of Thematic Spatiotemporal Association Rules (TSARs) from Solar SITS. Here, a description is given about the details of the modern algorithm MiTSAI, which is an extractor of Thematic Spatiotemporal Association Rules (TSARs) from solar Satellite Image Time Series (SITS). In addition, its adaptation to the Space Weather and discussion about the specific use in favor of forecasting activities are presented. Finally, some results of its application specifically to solar flare forecasting are also presented. MiTSAI has to extract interesting new patterns compared with the art-state algorithms.

KEYWORDS

Satellite Image Time Series; Thematic Spatiotemporal Association Rules; Space Weather Patterns.

1. INTRODUCTION

Daily several sources generate a massive amount of spatiotemporal data. Satellites are considered one of these sources and their data are called Satellite Image Time Series (SITS). SITS encompass a complex and interdisciplinary domain composed of a series of spatiotemporal images and their respective semantic data.

In this paper, SITS are used to support the analysis of solar data. The Solar SITS is composed of several solar images (acquired at different wavelengths) that can present Solar Active Regions (called sunspots) and semantic data that classify the image sunspots. The semantic data also gives the sunspots' location (solar coordinates) and the date when the images were collected (spatiotemporal characteristics).

Motivated by providing a better understanding of the Solar SITS and also supporting forecasting of solar behavior, we propose the Miner of Thematic Spatiotemporal Associations for Images (MiTSAI) algorithm. A high-intense solar event can cause various problems with telecommunication and navigation systems. The prediction of solar behavior can help in preventive measures avoiding the instabilities of these systems.

Our proposal is supported by the hypothesis that the extraction of Thematic Spatiotemporal Association Rules (TSARs) can aid in the analysis of the solar data. TSARs consider the temporal evolution of the sunspots and also the relationships among the sunspots.

MiTSAI is an algorithm that extracts TSARs from Solar SITS, considering their visual features and their semantic information. In our experiments, MiTSAI was validated by a domain expert. The domain expert examined whether or not the mined rules were interesting and the algorithm performance is acceptable.

The remainder of this paper is organized as follows: Section 2 presents the concepts, background, state of the art, and related works; Section 3 presents our proposed algorithm, the Miner of Thematic Spatiotemporal Associations for Images (MiTSAI); Section 4 presents the performed experiments to validate MiTSAI, and; Section 5 presents the conclusions and future works.

2. THEORETICAL FRAMEWORK

Solar data have a myriad of sources, being generated by detectors on both ground and space. Those space-based detectors are almost exclusively located on scientific satellites, which take part in the long-term programs of the international space agencies (NASA, ESA, JAXA, etc). The database of solar phenomena used for this work is <https://www.solarmonitor.org/>. There are available solar data from the following instruments: Magnetogram and 6173 /AA images from the Helioseismic and Magnetic Imager (HMI) onboard the satellite Solar Dynamics Observatory (SDO), Fe ix/x lines at 174 /AA Sun Watcher using Active Pixel System detector and Image Processing (SWAP) onboard PROBA2 scientific mission, H-alpha from Kanzelhoehe Observatory, X-Ray Telescope (XRT) onboard Hinode satellite, and Fe XII, 193 /AA line from Atmospheric Imaging Assembly (AIA) onboard SDO

Spatiotemporal data are characterized by space and time properties [9, 17, 22, 34]. In a formal definition: the D-database is spatiotemporal only if its items have spatiotemporal characteristics. I.e. let i be an item of D-database, we define i as a quintuple $\{x, y, z, t, F\}$, where x, y and z are coordinates in Cartesian space; t is temporal coordinates; and, F is a set of thematic attributes (non-spatiotemporal).

Examples of spatiotemporal data are: meteorological data [16, 28], sensor data [3, 31, 32], network traffic [24, 25], among others. An example of a spatiotemporal sensor data is $\{S11, W65, 0, 20150824, 0.123\text{mm}\}$, where the space coordinates are 11 to the south, 65 to the west, at sea level, the time coordinate is August 24, 2015, and at this place and time, it was recorded 0.123 milliliters of rain (thematic attribute).

Mining algorithms should consider space and time constraints. In the literature, these constraints can appear in the pre-processing time [7, 35] and in the post-processing [27, 38]. Applying the constraints to the pre-processing reduces the search space and the mining execution time.

However, the post-processing constraints can take advantage of previous mine results to tune-up the constraint values.

It is also possible to apply spatiotemporal constraints during the mining process [20]. Figure 1 shows the application of spatiotemporal constraints during the frequent itemsets generation. Since the item sets satisfy the spatiotemporal constraints, they are called spatiotemporal item sets.

The algorithm starts generating the itemset candidates. This step combines the frequent itemsets generating itemset candidates that are larger than the seeds. If it is the first iteration, the frequent itemsets are extracted from the database. The candidate itemsets are filtered using the spatial constraints (set by the user) resulting in spatial itemsets candidates. Those itemsets are filtered using the time constraints (set by the user) resulting in spatiotemporal candidates itemsets. The frequency of these candidates is calculated and the frequent itemsets are filtered. The frequent spatiotemporal itemsets are used as the seed to the generation of the candidate itemsets in the next iteration.

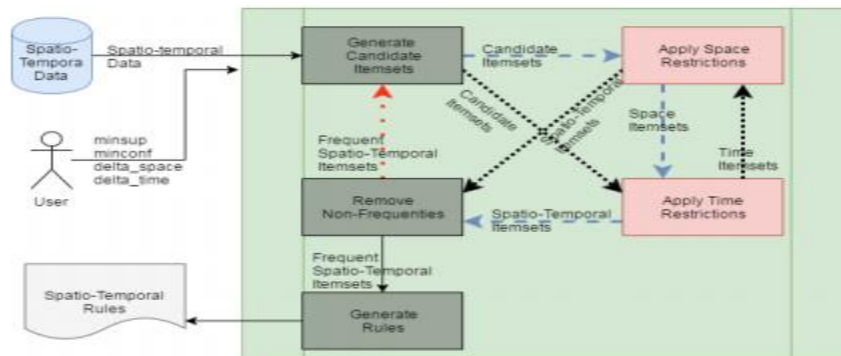


Figure 1. Spatiotemporal rule extraction based on Pillai et al. [20].

When it is not possible to generate frequent spatiotemporal itemsets, the found frequent itemsets are used to generate the spatiotemporal rules. To do so, the sub-itemsets are combined and the confidence of the rule is calculated. The rules that satisfy the minimum of confidence, defined by minconf are mined as the spatiotemporal rules.

It is also possible to invert the order in which spatiotemporal constraints are applied as also presented in Figure 1. In this way, the temporal constraints and then the spatial constraints are initially applied, thus generating the space-time candidate's itemset. In addition, regardless of the application order of the spatiotemporal constraints, it is also possible to iterate over the set of constraints.

In the literature, there are three distinct types of association rules for spatiotemporal domains. Each type of association rules can be used to achieve different goals; they are [23].

Moving Objects: Describes the movement of objects between regions. Such type of rule means that an object satisfying a given condition c has migrated from one region $r1$ to another $r2$ in a given time period $[t1, t2]$. The rule format is $(r1, t1, c) \rightarrow (r2, t2) < sup, con >$, where sup is support and con is the association rule confidence. For this type of rule, support is the number of objects that migrated from region $r1$ to region $r2$ in the period between $t1$ and $t2$ divided by the number of objects satisfying c in the same period. Confidence is the number of objects that

migrated ("divided" ?) by the total of objects in the region r1 in time t1. Examples of works using 'moving objects' are Kong et al. [14], Mohan and Revesz [18] and Alamri et al. [2].

Topological Relations: Rules that involve space topologies and predicates such as overlap, intersection, touching. . . , as well as temporal predicates such as sequential, parallel, and so on. The datasets usually need to be pre-processed to find the topological relationships and to organize the data in the function of the found relations. Only after the pre-processing, it is possible to apply the techniques of data mining. The rule format is $R1(obj1, obj2, t1) \rightarrow R2(obj3, obj4, t2) < sup, con >$, where R1 and R2 are space relations, obj1...4 Characteristics of the objects that differentiate them in time periods t1 and t2. For example, $overlapping(New York, rain, summer) \rightarrow neighbor(New York, high flow rivers, autumn)$. An example of work that uses topological relations is Burbey and Martin [6].

The thematic Rules: Are association rules that involve space and time properties and attributes not necessarily related to spatiotemporal properties. Often extracting such rules requires a preprocessing of the database. The pre-processing aims to expose spatiotemporal properties and associate them with no spatiotemporal attributes (thematic attributes). Thematic rules usually have the following format: $a1(R1, t1) \rightarrow a2(R2, t2) < sup, con >$, where a1 and a2 are attributes of the domain to which the mining is being applied (e.g. temperature and atmospheric pressure in the climate domain), R1 and R2 are regions whose attributes are related to t1 and t2 the period. E.g. $Rainfall(New York, summer) \rightarrow Rainfall(New Jersey, autumn) < sup, con >$, this example shows that if it rains in New York in summer, it will rain in New Jersey in the autumn, with support and confidence of sup e con, respectively. An example of work using thematic rules is Landgrebe et al. [15].

An example of spatiotemporal association rules extractor algorithm is found in Compieta et al. [8]. This work is based on Apriori to mine spatiotemporal association rules. In its spatiotemporal data model, each item i is associated with a set of spatial points se_i in which the item occurs at a given time period t . A virtual point vp is defined with a spatial point that supports an itemset. An itemset is considered frequent, if and only if it is frequent in the set of virtual points vp 's – called spatial itemset. The virtual point vp has its existence associated with one or more time periods. Based on this, the idea of the algorithm is to avoid unnecessary processing performed by the Traditional Apriori algorithm. Thus, that approach only processes data with a significant spatiotemporal relation –virtual points. The result is a set of frequent itemsets and association rules. Another change made to the Traditional Apriori is: during the joining of two itemsets, $i1$ and $i2$, (performed for candidate generation) it is necessary to check whether the intersection of the associated virtual point sets vp_{i1} and vp_{i2} is not empty, $vp_{i1} \cap vp_{i2} \neq \emptyset$

Correlated Work

Kawale et al. [13] apply time-series extraction to climate data, precipitation, and temperature data, to determine anomalies that occur in one region and if that anomaly will happen in another region after some time. In this approach, the authors performed the extraction of positive and negative patterns using a graph-based approach (representing spatial constraints through edges). The idea is to group the locations such that members of a group have characteristics more similar than the members of other groups. That algorithm aims to find pairs relating groups with different characteristics.

Yoo and Bow [37] proposed a framework to find correlated patterns composed of two algorithms for mining correlated neighbors. By using the approach of filtering spatiotemporal relations and refinement of the shape of objects, the approach reduces the number of candidates compared to the traditional data mining approaches. Different ways of determining closest neighbors were also evaluated by the use of estimates of the relative distance of each edge of the objects.

The work of Hana et al. [10] processes neighborhood relations between objects over a period of time by the use of spatial queries with temporal parameters. The algorithm developed in this work was called START and had three phases: i) calculation of spatiotemporal predicates; ii) generation of frequent itemsets, and; iii) extraction of spatiotemporal rules (based on Apriori). In this work, the spatiotemporal objects are characterized by a quadruple $\{ai, gi, pi, ti\}$. The STAR shows the spatiotemporal evolution of geographic objects, $(X, aiti) (X, gi, ti) (X, pi, ti) \rightarrow (R, cr, ti +)$; X is the reference to an object in the database, ai is the attribute that characterizes X in the time ti , gi is the geometric feature of X in the time ti . pi is a topological relation that X may have with an object characterized by cr that has R probability of occurring close to X in time $tt+$. Example, $(\text{Rain}, 0.15\text{mm}, \text{fall})(\text{Rain}, 20\text{km}^2, \text{autumn})(\text{Rain}, \text{neighbourhood}, \text{autumn}) \rightarrow (0.8, \text{increase}, \text{Autumn} + 1 \text{ month})$. That approach did not consider the influence of multiple events in others, what can occur in images. Also, a pre-processor is required to define the spatiotemporal data predicates, e.g. a definition of the overlapping neighborhood. It may cause data loss.

The work presented in Huo et al. [12] finds co-occurring spatiotemporal rules by applying a sliding window whose increments of data is dynamic, and the data importance reduces overtime. The proposed algorithm, DIAD, makes the use of hash trees for the storage and access to the standards, through this approach, the DIAD presented a gain of performance compared to the other algorithms. The DIAD distributes events into partitions and calculates the spatial distance between the events. As new data are added, these are distributed between the partitions and the events on the affected partitions are updated. The decay of importance is an adapted technique that aims to capture changes in the flow of events dynamically. The domain used is derived from social networks, which have spatial references. This approach has a limitation in relation to the work proposed in this paper: It does not mine complex data, e.g., images. To mine images, it is necessary to extract the image features. However, the domain that DIAD is very dynamic and to process the images would cause a delay in the processing and performance of the algorithm.

Pillai et al. [20] presented a new algorithm to find spatiotemporal rules through the application of filters; these filters are used to restrict standards that satisfy spatiotemporal constraints. The algorithm also applies the refinement of geometric shapes to take into account the topology of events. This work used a dataset of solar images; through the application of this technique, it was possible to find forms of events that moved. The algorithm, based on Apriori, can handle a relatively large amount of data and is called FastSTCOPs-Miner. Pillai et al. [21] presented an evolution of the previous work on the following aspects: a new framework for mining co-occurring patterns; spatial events are modeled as 3D objects and the evolution of their shapes are captured; an algorithm for the discovery of co-occurring rules based on the evolution of spatial relations is presented. However, both works have an important limitation: they do not extract association rules, but rather sequential patterns for the evolution of an event. In this way, the approaches do not consider the influence among events. In addition, none of the approaches consider thematic attributes of the domain.

3. MINING THEMATIC SPATIOTEMPORAL ASSOCIATION RULES IN IMAGES

In this paper, we propose the Miner of Thematic Spatiotemporal Association for Images (MiTSAI), a new algorithm to extract thematic spatiotemporal association rules from spatiotemporal series of images and textual data.

MiTSAI considers the relationship between itemsets that happen at the same moment and obeys a spatial constraint. It also considers the evolution of the itemsets obeying a time constraint. Both constraints are set by the user.

MiTSAI mines association rules of the form: $r : i_1 \dots i_n \rightarrow j_1 \dots j_m \langle \text{sup}, \text{conf}, \text{space}, \text{time} \rangle$, where $i_1 \dots i_n$ is a set of items that happens at the same time and obeys the spatio constraint. In solar context, the item is defined as an attribute of the sunspot (solar event) that presents spatiotemporal characteristics, e.g., the visual feature of a sunspot is an item, its classification is an item. A spatial constraint is an input parameter set by the user; it limits the distance between sunspots to consider be considered to join the itemsets. I.e. i_1 is closer to the other items $i_2 \dots i_n$ than the limit given by the spatial constraint. The same is valid for $j_1 \dots j_m$. The average of the space values is given by space. The ruling period varies from i to j using the time unit, showing the time evolution of the itemset. sup and conf are the support and confidence values. The sup and conf calculus are showed later.

Algorithm 1 presents the Miner of Thematic Spatiotemporal Association for Images (MiTSAI) algorithm. The implementation of MiTSAI is available in Silveira-Junior [30]. MiTSAI inputs are: a horizontal spatiotemporal database, DB; minimum values of support and confidence, minSup and minConf, respectively; the spatial restriction, distance, and; the time-variation restriction, period. The MiTSAI process is divided into two general steps: (i) finding the spatial itemsets in the database DB, as presented in Line 2, and; (ii) generating the spatiotemporal association rules based on the spatial itemsets generated at (i)-step, as presented in Line 3.

Data: *DB*: database; *minSup*: minimum support value; *minConf*: minimum confidence value; *distance*: spatial restriction; *period*: time restriction.

Result: *R*: Set of spatiotemporal rules.

```

1 begin
2   itemsets  $\leftarrow$  genItemset( $\emptyset$ , DB, minSup, distance);
3   R  $\leftarrow$  genRules(itemsets, minConf, period);
4 end

```

Algorithm 1: Miner of Thematic Spatiotemporal Association rules for Images (MiTSAI) – Overview.

Algorithm 2 presents the first step of MiTSAI. It is responsible for finding frequent spatial itemsets. An itemset I is a set of items, it is formally defined as $I = \{i_1 \dots i_n\}$ for $n \in \mathbb{N} \mid n \geq 1$ and $i_a = i_b$ for $0 < a, b \leq n$ only if $a = b$. A spatial itemset $SI = si_1 \dots si_n$ is an itemset whose items have spatial characteristics such as $si_a.location$ and its items obey the spatial restriction given by the user. An example of a spatial restriction for the item $ia, ib \in SI$ can be stated using Euclidean distance among these items ia, ib . The spatial restriction is characterized by the Euclidean distance being not higher than a parameter. In MiTSAI, the distance is received as input.

Data: *base*: Itemset; *db*: Database; *minSup*: minimum support value; *distance*: spatial restriction.

Result: *R*: Frequent spatial itemsets.

```

1 begin
2   R ← ∅ ;
3   fi ← frequentItems(db) ;
4   for item i ∈ fi do
5     si ← base ⊕ i ;
6     si.support ←  $\frac{|i.closeToDistance(base)|}{|DB|}$  ;
7     if si.support ≥ minSup then
8       p ← projection(db, si);
9       R ← R ∪ {si} ∪ genItemset(si, p, minSup, distance) ;
10    end
11  end
12 end

```

Algorithm 2: MiTSAI – genItemset Function.

A frequent spatial itemset F SI is a spatial itemset that often happens in the database. The frequency of a F SI is limited by the support value, which is calculated as support $(FSI) = \frac{|FSI|}{|DB|}$, where F SI is an itemset, |F SI| is the number of occurrences of F SI's itemset at the same moment, obeying the spatial restriction, and |DB| is the number of tuples in the database. A tuple is a quintuple {x, y, z, t, F} where x,y and z are coordinations in Cartesian space; t is the time coordinate, and; F is a thematic attribute. An itemset is considered frequent if its support is greater than the minSup, set by the user.

In Algorithm 2 at Line 2, the result set is returned to Algorithm 1, and R is initialized as empty. In-Line 3, f i is initialized with the set of frequent spatiotemporal items in the database db. For the first iteration of genItemset, db shall be the whole database DB, however, for the recursive iteration of genItemset, db will be a projection of the database DB. This procedure is further detailed in the explanation of Lines 8 and 9.

After that, the Algorithm 2 at Line 4 makes a loop for each item in f i. In the loop scope, the spatial itemset is created at Line 5, by adding base (itemset received as the input parameter) and the f i's item called i. If it is the first iteration of genItemset Function, base is an empty itemset i.e. the operation base ⊕ i creates an itemset composed only with the item i. In the other iteration of genItemset, base contains the frequent spatial itemset found on the database projecting project db.

In-Line 6, si-itemset support is calculated considering the spatial restriction, distance, set by the domain expert. The spatial restriction is considered by the |i.closeT odistance(base)|. This operation shall return the number of occurrences when i happens at the same time that the base's items happen, and the Euclidean distance between i.location and base.location in the limit value.

Since base itemset can have items in different spatial positions, base.location shows the center of this items and si.location receives the base.location updated, considering the i.location.

Figure 2 presents an example of spatial restriction applied during support calculation. For this figure, consider A, B and C as spatial items and e is a spatial restriction; the distance between A.location and B.location are smaller than e-value, i.e. A.closeT odistance(B) < e. So, as {A, B}-itemset obey the spatial restriction, it is considered a valid spatial-itemset occurrence, and it will be count for support({A, B}) calculation. The A.closeT odistance(C) > e, that way, {A, C}-itemset does not obey the spatial restriction; it is not considered a valid spatial itemset occurrence and it will no be counted during support({A, C}) calculation.

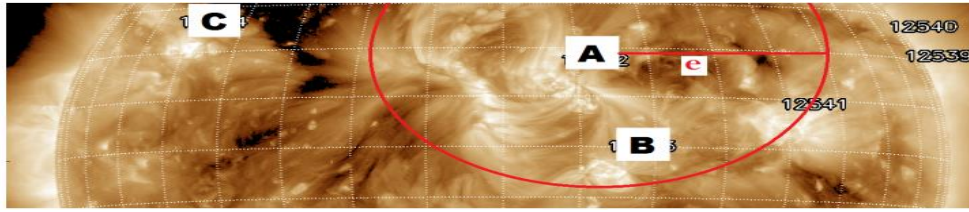


Figure 2. Example of support calculation considering the spatial data characteristic.

In-Line 7, it checks and filters the si -frequency from the DB-database perspective. If si is not frequent, it is discarded, and the loop continues for the next item $i \in f_i$. If si is frequent, Line 8, a projection p of the db-database is generated based on si . The p -projection is composed only of the database registries where si happens (valid occurrences of si). Furthermore, R -set is updated adding to itself si and $getItemset(si, p)$ result. $getItemset(si, p)$ is a recursive call to $genItemset$, using si as base-input parameter and p as db-input parameter. By that recursively way, it is possible to find all combinations of si with the frequent items in the projected database. The recursive loop ends when no more generated candidate itemset si is frequent (passed by $si.support \geq minSup$ condition at Line 7).

A generic example of projection is presented in Table 1. In this example, the spatial characteristic is not considered. The database DB is composed of 5 registries, which IDs are 20150101 . . . 20150105. DB shows the items that composed the registry, for instance, the items A, B, c, and d composes the registry whose ID is 20150101. The projection p of this database is performed for Itemset equals to $\{A, B\}$. The p -projection is composed of the registries whose IDs are 20150101, 20150102 and 20150105. The $getItemset(si, p)$ call will use p as db parameter, considering only the 3 items for it recursively processing.

DB's ID	DB's items	In projection
20150101	A, B, c, d	Yes
20150102	A, B, c, d	Yes
20150103	A, f, c, d	No
20150104	f, B, c	No
20150105	f, B, c, A	Yes

Table 1. Example of database DB and the projection for the itemset $I = \{A, B\}$.

Algorithm 3 presents the second step of Algorithm1, $genRules$. The $genRules$ -function receives the frequent spatial itemsets generated by Algorithm 2. As a result, it returns a set of thematic spatiotemporal association rules.

Data: *itemsets*: Set of spatial itemsets; *minConf*: minimum confidence value *period*: time restriction.

Result: *R*: Set of thematic spatiotemporal rules.

```

1 begin
2   R ← ∅ ;
3   for itemset i and j ∈ itemsets | i ≠ j do
4     r ← rules :< i → j > ;
5     r.confidence ←  $\frac{|j.closeToPeriod(i)|}{|i|}$  ;
6     if r.confidence ≥ minConfidence then
7       R ← R ∪ {r} ;
8     end
9   end
10 end
    
```

Algorithm 3: MiTSAI – $genRules$ Function.

In the Algorithm 3 at Line 2, the set of rules result R is initialized with an empty set. At Line 3, there is a loop that combines tuples of itemset, i and j being $i \neq j$. For each combination of i and j a rule r is created as $r = \langle i \rightarrow j \rangle$, Line 4. Since, all combinations is considered, both rules shall be created: $r1 = \langle i \rightarrow j \rangle$ and $r2 = \langle j \rightarrow i \rangle$.

After the rule generation, its confidence is calculated at Line 5. For the confidence calculation, the time restrictions are considered: an occurrence of the rule $i \rightarrow j$ is considered for the confidence calculation only if it obeys to $i.date < j.date \leq i.date + period$, where period is a parameter set by the data-domain expert. The $|j.closeT\ period(i)|$ -function returns the number of occurrences of i and j that obeys the time restriction. The confidence calculation is performed by $|j.closeT\ period(i)|$ divided by the number of occurrence of i-itemset. Figure 3 presents a timeline that exemplifies this scenario. For this Figure, it is employed $i = A$, $j = B$ and d is the period.

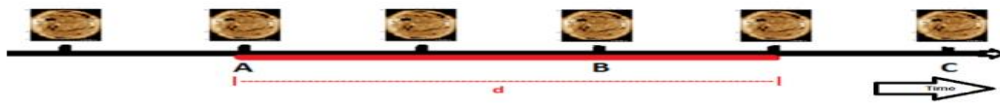


Figure 3. Example of confidence calculation considering the data temporal characteristic.

When the rule's confidence value is higher than $minConf$ (set by the data domain expert), the rule is to add to the result-set, R (see Lines 6 and 7). Otherwise, the rule is discarded. That way, it is possible to extract rules from the spatiotemporal domain.

The extracted rules has the following format: $\{ia \dots ib\} \rightarrow \{ix \dots iy\} < support, conf\ idence, average\ delta\ time, average\ delta\ spatial >$, where $\{ia \dots ib\}$ and $\{ix \dots iy\}$ are frequent spatial itemsets, support is the average of $\{ia \dots ib\}$ -support and $\{ix \dots iy\}$ - support, conf idence is the rules confidence value, average delta time are the average period between $\{ia \dots ib\}$ and $\{ix \dots iy\}$ occurrences, and average delta spatial is the spatial average in-between $\{ia \dots ib\}$ added the spatial average in-between $\{ix \dots iy\}$.

For the traditional rules $i \rightarrow j$ always obeys the property $i \cap j = \emptyset$. However, for the Spatiotemporal Rules extracted by the MiTSAI approach this property is no longer valid. It happens to increase the rule flexibility in the spatiotemporal domain. That way, it is possible to have items happens in both itemsets but occurring in a different period.

That way, the MiTSAI rules show the relation between items and their evolution during a period. Since the attributes considered during the processing are not only the spatiotemporal ones, those extracted rules are considered thematics because they use thematics attributes. Subsection 3.1 presents the Temporal Series of Solar Images used in the experiments for the MiTSAI.

Subsection 3.2 presents an example of the MiTSAI execution for the Temporal Series of Solar Images. Subsection 3.3 presents optimization implemented in the MiTSAI algorithm to get better performance results.

3.1. Temporal Series Of Solar Images

Figure 4 and Figure 5 presents the solar images of one day obtained from site Solar Monitor [19]. Each image data comes from a specific instrument and shows information recorded in different wavelengths or height of the solar atmosphere as illustrated in Bobra and Couvidat [5]. The Solar Monitor information is also composed of textual data that describe each Solar Active Region characteristics, as can be seen in Figure 5. The textual data brings information on NOAA number attributed to each Solar Active Region, as well as its corresponding latest position, Hale class, McIntosh class, area, number of spots and recently produced flares for a given date.

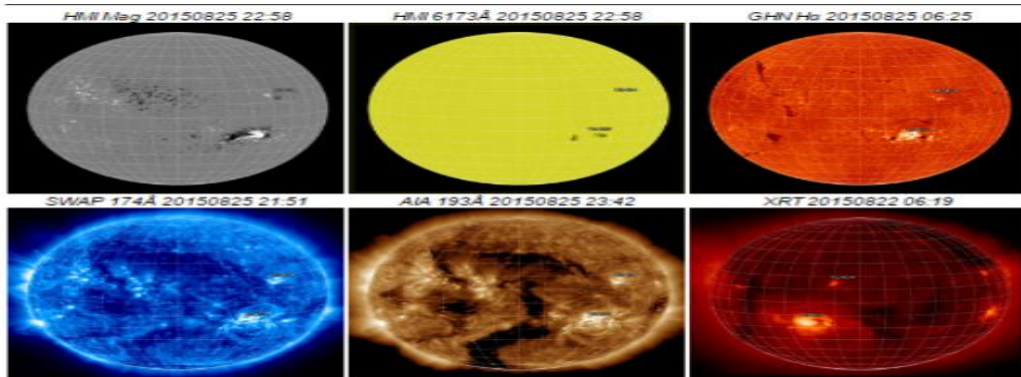


Figure 4. Example of solar images from one day taken from the NOAA [19].

NOAA Number	Latest Position	Today's/ Yesterday's Hale Class	McIntosh Class	NOAA Active Regions Sunspot Area [millionths]	Number of Spots	Recent Flares
12403	S15W32 (487°,-340°)	βγδ/βγδ	FkoEko	0930/0760	47/66	C2.2(19:23) C1.3(18:01) C1.1(14:48) C1.3(12:33) C2.3(11:55) C2.3(07:58) C4.3(06:18) C1.1(04:14) / -
12404	N14W45 (653°,148°)	ββ	Cro/Cro	0020/0030	04/07	-

Figure 5. Example of textual data that shows information of active regions corresponding to the solar images of Figure 4.

In this work, the Hale Class was excluded from the rule extraction, because it is a simplified version of a McIntosh Class.

The data from NOAA [19] were extracted and pre-processed using SETL Architecture [33]. The extraction retrieved the six images with different wavelengths for each day. Each image was processed to extract its feature vector using the SURF-algorithm [4], Haralick [11], and Histogram.

The feature vector was discretized using the Omega-algorithm [26]. This process was necessary because MiTSAI uses discrete data as input. The Omega algorithm was employed because it is designed to preprocess data for the association rule task. It reduces the number of intervals generated and also reduces the data entropy. Those characteristics facilitate the association rules extraction process.

```
[ {
  "id": "12403",
  "date": "20150825",
  "location": "417,-345",
  "mcintosh_class": "Fkc",
  "area": "0930",
  "images": [ "[0.1-0.5) [0.1-0.5) ...", ...
              "[0.5-1.0) [2.4-6.2) ..." ]
} ]
```

Figure 6. Example of a Solar Active Region tuple.

The pre-processing also separates the sunspots creating a registry for each one. Figure 6 shows an example of one sunspot after the pre-processing. It contains information about the date and active region number together with its location, area, McIntosh-class and image parameters.

Domain Constraint: It is expected that the domain expert determines if the sunspot is in the rule’s cause and consequence. This constraint is considered during the confidence calculation: the occurrence of a rule is counted using only the time restriction and if the occurrences of cause and consequence have a sunspot in common.

3.2. Example Of Mining Thematic Spatiotemporal Association Rules For Series Of Solar Images

MiTSAI starts reading the data from the database. Each sunspot (database records) are split into spatiotemporal items. A spatiotemporal item has spatiotemporal characteristics and a thematic attribute. For instance, spatiotemporal $i_{20150825,(417,-345)}^{mcintosh} = Fkc$, where 20150825 is the date, (417, -345) is the sunspot location, and F kc is the thematic attribute; and, the sunspot id is also stored.

MiTSAI counts the occurrence of each item considering only its values (not the spatiotemporal characteristic). That way is possible to determine which are the frequent items, generating the 1-size frequent itemsets. For instance, itemset $I = \{F kc\}$.

For each itemset, MiTSAI makes a database projection p. In each projection p, only the date when the itemset occurs is recorded. The recurrence call for genItemset finds the frequent items in the projection and creates one itemset for each frequent item concatenated to the base itemset (itemset whose projection is based on). If the created itemset is frequent, it is considered for the result. That way, the 2-size itemsets are created, and also, by the recursion, the larger size itemsets are generated too. For instance, the itemset $I2 = \{F kc [0.1 - 0.5][0.1 - 0.5] - 0930\}$ shows the association between F kc and a visual characteristic whose sunspot are is 0930. I.e., the F kc and the visual characteristic often happens at the same time, and they obey the spatial restriction.

The second step, genRules, combines all frequent spatial itemsets generating the rules. During the rule confidence calculation, the time restriction and the domain constraint are considered. For instance, $r : I \rightarrow I2 < 0.05 \ 0.8 \ 1 \ 0 >$ show that {F kc}-itemset and {F kc [0.1 - 0.5][0.1 - 0.5] - 0930}-itemset happen in at least 5% of the database; in 80% of the time when a sunspot is classified as F kc (Itemset I), in the next day (1-value in the average delta time), the same sunspot

keeps the classification F_{kc} , but presents the visual characteristic $[0.1-0.5][0.1-0.5]$ associated to an area of 930 one-ten-thousand of solar disc. The average delta spatial is zero because the generated pattern is regarding only one sunspot. That way, the time restriction is respected and also the domain constraint.

3.3. MiTSAI Optimizations

During the database projection, the database items are not copied, instead of it, a bitmap is created for the occurrence of the items. The projection passes the bitmap as a reference and filters the interesting items already respecting the spatial restriction. That way, the projection is composed only of the items that already is in the valid area from the initial item occurrence. This is the same strategy that is used by the ARMADA algorithm, proposed in [36].

Other optimization, it is possible to have projections that are redundancy. For instance, the itemsets A and B, it is possible that A be part of B projection and B part of A. Those redundancies are found in the recursion, and for those cases, the recursion is aborted.

Those optimizations were implemented in the original algorithm MiTSAI and brought a better performance and reduced memory usage considerably. Those optimizations made it possible to process almost ten years of solar data and images without the need for distributed processing, as it is shown in Section 4 spending adequate time as attested by the domain experts. MiTSAI complexity is $n \times \ln(n)$ being n the database input size: Each time a projection is done, it reduces the search-solution space; that way, the complexity of MiTSAI algorithm can be calculated in function of its input.

4. EXPERIMENTS, RESULTS AND DISCUSSIONS

In this section, we present three sets of experiments, for each database feature extractor (Histogram, Haralick, and SURF), with the same configuration: Support minimum of 1%; confidence minimum of 75%; maximum of space variance of 150 by 10, 000 parts of the solar disk, and; maximum of time variance of 20 days. The feature extractor has been chosen because each one extract features based on different kind of characteristics: Histogram is based on the image color, Haralick is based on texture-based feature extractor, and SURF is form-based feature extractor. MiTSAI allows that different feature extractors can be used joined; however, MiTSAI will not compare cross-feature extractor types.

The database is composed of more than 10300 sunspot records split by day for the period starting on August 25, 2007, and ending on August 24, 2016, i.e., over 70000 feature vectors was submitted to MiTSAI.

Experiment with Histogram as Feature Extractor

The rules presented in Figure 7 were mined using the database represented by the Histogram feature extractor. Figure 7 presents three rules, R1, R2 and R3. R1 shows a visual feature of a sunspot whose size is 20 parts of solar disk. The visual feature happens at the same time of Bxo-McIntosh and Cho-McIntosh. For this cause, there are two possible scenarios: (i) they are two sunspots; one has the visual feature and Bxo/Cho-McIntosh classification and there is another closer sunspot with Cho/Bxo-McIntosh classification; (ii) there are three sunspots, one

represented by the visual feature, one represented by Bxo-McIntosh and one represented by ChoMcIntosh. The distance in-between the sunspots are, on average, 116.171 parts of the solar disk. The consequent presents the visual feature that is the evolution of the cause's visual feature, in a time-variance average of 2.869 days. This pattern happens in at least 3.9% of the data and the confidence is 95.8%. Figure 8 presents an example of the occurrence of this in the solar Scenario (i).

```

(R1)
<(3584760.000;3590618.000],[3161025.000;3167250.000], 0020> Bxo
  Cho ->
<(3584760.000;3590618.000],[3154539.000;3161025.000], 0040>
  sup=0.039 conf=0.958 space=116.171 time=2.869

(R2)
<(3584760.000;3590618.000],[3161025.000;3167250.000], 0020> Bxo
  Hsx Cso ->
<(3577235.000;3584760.000],[3161025.000;3167250.000], 0070>
  sup=0.054, conf=0.767, space=135.042, time=5.023

(R3)
<(3590618.000;3597197.000],[3150223.000;3154539.000], 0060> Cro
  ->
<(3590618.000;3597197.000],[3150223.000;3154539.000], 0060>
  sup=0.046, conf=0.875, space=25.703, time=2.714
    
```

Figure 7. Rules extracted from the Histogram features database.

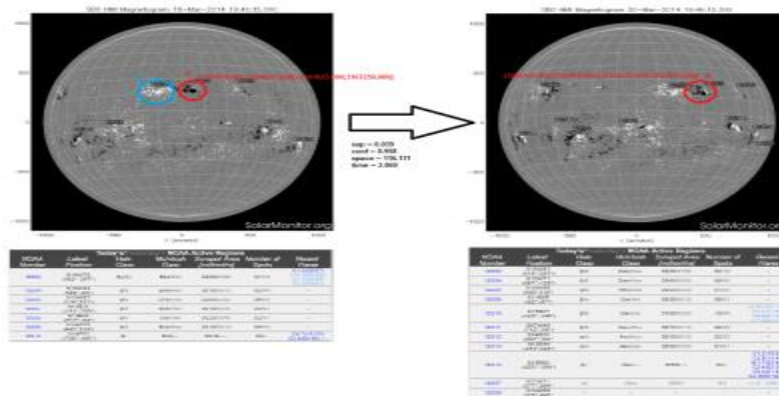


Figure 8. Example of an occurrence of R1. Image adapted from NOAA [19]

R2 exhibits the same feature vector of R1's cause associated with Bxo-McIntosh, Hsx-McIntosh, and Cso-McIntosh. In the case, there are two possible scenarios: (i) the feature vector and the of McIntosh classification is of the same sunspot (giving three possibilities), and; (ii) there are four sunspots. The average space variance in-between the sunspots is 135.042 parts of the solar disk. The R2's consequent shows a different visual feature comparing to R1's consequent. It shows that the association of the feature vector that appears in R1 and R2's produces, with different sunspots results, a different evolution of the sunspot. R3 shows two sunspots, one represented by the feature vector and the second one by the CroMcIntosh, the distance average in-between them is 25.703. The consequent shows the evolution of the sunspot associated with the feature vector, evolving to the consequent feature vector.

Experiment with Haralick as Feature Extractor

```

(R4)
<(-0.350;0.341], (9.764;23.692], (32118.140;32751.662)]... 0120> Cso
-> Hsx
sup=0.042, conf=0.758, space=1.249, time=9.227

(R5)
<(-0.350;0.341], (9.764;23.692], (45521.675;45809.851)]... 0110>
->
<(-0.350;0.341], (9.764;23.692], (45521.675;45809.851)]... 0110> Axx
sup=0.024, conf=0.875, space=40.564, time=2

(R6)
<(-0.350;0.341], (9.764;23.692], (29061.805;30339.404)]... 0300> Cao
-> Dso
sup=0.177, conf=0.772, space=92.801, time=2.588
    
```

Figure 9. Rules extracted from the Haralick features database.

Figure 9 shows the rules extracted from a database whose images were processed by the Haralick feature extractor. R4's cause presents two sunspots, the first one is represented by the visual feature vector and the second one is classified as Cso-McIntosh, the average spatial distance between them is 1.249 parts of the solar disk. The sunspot represented by the feature vector evolves to a Hsx-McIntosh in an average time of 9.227 days. The rule is found in 4.9% of the database with the confidence of 75.9%. Figure 10 presents an example of the occurrence of these rules. In that example, there is the 11204-sunspot that is close to the 11203-sunspot; 11203-sunspot is classified as Cso and presents the visual feature of R4's cause. 11203-sunspot evolves in one day to Hsx-classification. That way, we can show a real example in which the R4-rule is validated: the pattern happens, and it is obeying the spatiotemporal constraints.

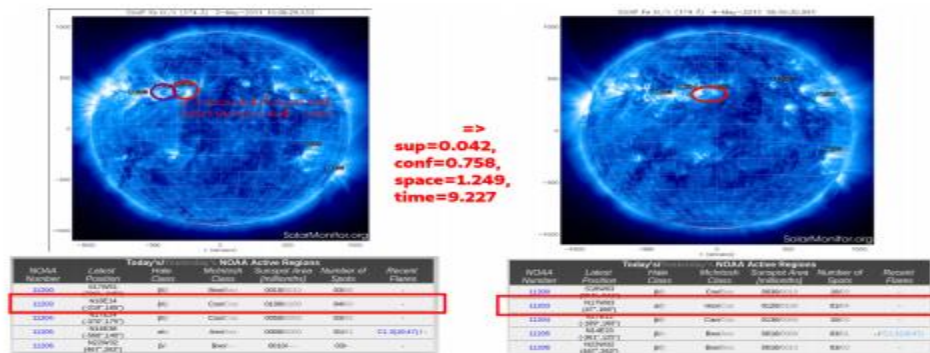


Figure 10. Example of an occurrence of R4. Image adapted from NOAA [19]

R5 shows, in its cause, a sunspot represented by a visual feature vector. In its consequence, a similar visual feature occurs in the same sunspot associated with other sunspot classified as AxxMcIntosh. The average distance between that sunspot is 40.564 parts of the solar disk. The period average between cause and consequence is 2 days. The support is 2.4% and confidence is 87.5%. R6 shows two sunspots whose average distance is 92.801 parts of the solar disk. The cause shows a visual feature associated with a sunspot classified as Cao-McIntosh. That means that the sunspots represented by the visual feature evolve to Dso-classification when they are associated with a Cao-McIntosh sunspot. The period average of this process is 2.588 days.

Experiment with SURF as Feature Extractor

Figure 11 shows the rules extracted from the database whose images were processed by SURF feature extractor. R7 shows a sunspot classified as Hsx that evolves to Dso-McIntosh when it is associated with another sunspot with the same visual feature. The distance between the sunspots is on average 33.22 parts of the solar disk. The average time for this association happen is 6.5 days. It happens in 16.4% of the data, and its confidence is 80%. Figure 12 presents a real occurrence of R7-rule, in the occurrence, the Hsx-classification below to 11895-sunspot and the visual feature below to the 11897-sunspot, in that occurrence the 11897-sunspot evolves to the Dso-classification. The R7-rule represents as i) Hsx and the visual feature below to the same sunspot and it evolves to Dso-classification, or; ii) Hsx and the visual feature below to different sunspots as presented in the Figure 12.

```
(R7)
Hsx <(665.847;666.469], (14.534;15.198], (-0.428;8.490]... 0010>
-> Dso
sup=0.164, conf=0.8, space=33.22, time=6.5

(R8)
<(406.643;407.087], (14.534;15.198], (-0.428;8.490]... 0180> Bxo
-> Dro
sup=0.107, conf=0.79, space=22.2, time=13$

(R9)
<(667.785;668.583], (14.534;15.198], (-0.428;8.490]... 0040> Axx
->
Hsx <(406.643;407.087], (14.534;15.198], (-0.428;8.490]... 0180>
<(665.847;666.469], (14.534;15.198], (-0.428;8.490]...
0010> Axx
sup=0.028, conf=1, space=30.248, time=4
```

Figure 11. Rules extracted from the SURF features database.

R8 also presents two sunspots in its cause, the distance between them is 22.2 parts of the solar disk. The sunspot that is related to the visual feature evolves to a R8-McIntosh in a time average of 13 days. It happens when it is associated with a Bxo-McIntosh sunspot. R8 has the support of 10.7% and confidence of 79%.

R9 case presents two sunspots, and the distance between them is 30.248 parts of the solar disk. The sunspot that presents the visual characteristic that appears in the cause evolves to one that is classified as Hsx-McIntosh, in an average time of 4 days. That sunspot can present one of the two visual characteristics, or it can be from other sunspots close to it in an average of 30.248 parts of the solar disk. Also, it is possible the presence of at least another sunspot associated with the previous sunspot.

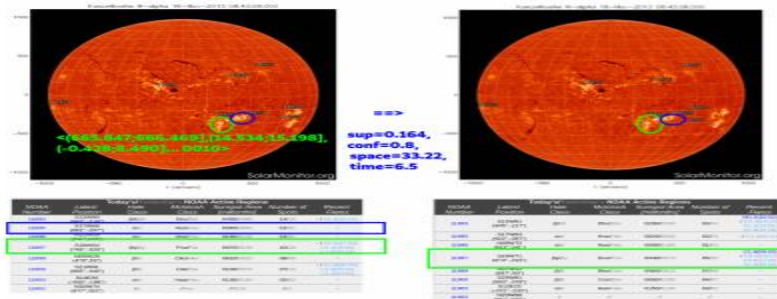


Figure 12. Example of an occurrence of R7. Image adapted from NOAA [19]

Discussions

The rules show that sunspot's behaviors can directly be connected among themselves. To verify the results and to validate the prediction results, we used a smaller database of images containing a half year of images an average of 3 sunspots each day –579 sunspots– and compared if the extracted rules appeared the non-training. That way as possible to tell which rules were true positive and false positive, and by comparing with previews rule-set results from Apriori, considering the spatiotemporal constraints, it was possible to calculate the false negatives. For histogram 75.7% of precision, for Haralick 69.1%, for SURF 78.8%. The recall values are 89.1%, 93.1%, and 87.3% for the histogram, Haralick, and SURF, respectively. The algorithm searches ever possible combination of itemsets that seems to be profitable; however, the recall is not 100% since the solar domain has a domain constraint the sunspot in the rules' cause must be the same in the rules' consequence. That way, some patterns are living out from the result. That constraint is needed since the rules will be used to generate a predictive learning model, as future work.

MiTSAI has presented an acceptable performance according to the domain expert even without distributed processing. The MiTSAI's performance variates according to the user inputs: minimum support, minimum confidence, space constraint, and time constraint. As less restrictive the constraints are, more patterns are extracted decreasing MiTSAI's performance; however, for the solar images domain, in the worse scenarios of constraints, we were able to process the 10 years of data in less than two hours.

In conclusion, we can say that MiTSAI has brought new and valuable patterns for the solar domain according to our solar domain experts, showing that MiTSAI is appropriate for SITS domain. MiTSAI can also be extended to other spatiotemporal domains.

5. CONCLUSION AND FUTURE WORKS

Satellite Images Temporal Series (SITS) is a challenge and multidisciplinary domain. Its analysis involves image processing, spatiotemporal characteristics, and additional semantic data processing. Despite the huge amount of usage possibilities, its analysis is still complex and limited, as the literature suggests.

Aiming to support the SITS analysis, MiTSAI was proposed in this paper and applied to Solar SITS. MiTSAI extracts Thematic Spatiotemporal Association Rules (TSAR) that considers the relationship in-between events that are happening at the same time and also their evolving in a period of time. In our experiments, we show that MiTSAI was able to extract the TSAR from the Solar STIS and this result was analyzed by the domain expert as new and relevant patterns for the Solar SITS.

By using the extraction of TSAR to extract patterns from the Solar STIS, we are able to extract patterns with over 75% of precision and a high value of recall (over 85%). This result is acceptable by the domain experts and also the performance result. The patterns are considered new since it is the first work that uses TSAR to extract patterns from the Solar STIS, the patterns are also considered relevant for the domain understanding. The main contribution of this working is the new way of applying the spatiotemporal constraints during the processing, it brought new valid information for the solar climatic domain and this technique can be applied to other domains that are composed by spatiotemporal images and textual series.

As future works, we have two proposals. The first proposal is to handle the visualization of the extracted Thematic Spatiotemporal Association Rules. In this proposal, a Visualizer shall read the rules and find examples that best fit the rule, as presented in Figure 8. The second proposal is to apply an Associative Classification to process the extracted association rules to employ the mined rules in a future classification.

Acknowledgment

The authors thank the SolarMonitor.org for free providing of the solar data used in this work. We also thank CAPES, CNPq and FAPESP for the financial support.

REFERENCES

- [1] T Abirami-Kongu, P Thangaraj, and P Priakanth-Kongu. Wireless sensor networks fault identification using data association. *Journal of Computer Science*, 8(9):1501–1505, 2012.
- [2] Sultan Alamri, David Taniar, and Maytham Safar. A taxonomy for moving object queries in spatialdatabases. *Future Generation Computer Systems*, 37:232 – 242, 2014. ISSN 0167-739X. doi: <https://doi.org/10.1016/j.future.2014.02.007>.
- [3] Shadi A. Aljawarneh, Radhakrishna Vangipuram, Veereswara Kumar Puligadda, and Janaki Vinjamuri. G-spamine: An approach to discover temporal association patterns and trends in internet of things. *Future Generation Computer Systems*, 2017. ISSN 0167-739X. doi: <https://doi.org/10.1016/j.future.2017.01.013>.
- [4] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. pages 404–417, 2006.
- [5] Monica G Bobra and Sebastien Couvidat. Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm. *The Astrophysical Journal*, 798(2):135–172, 2015.
- [6] I. Burbey and T.L. Martin. A survey on predicting personal mobility. *International Journal of Pervasive Computing and Communications*, 8(1):5 – 22, 2012.
- [7] M. Chen, S. Mao, and Y. Liu. Big data: A survey. *Mobile Networks and Applications*, 19(2):171–209,2014.
- [8] Paolo Compieta, Sergio Di Martino, Michela Bertolotto, Filomena Ferrucci, and T Kechadi. Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages & Computing*, 18(3):255–279, 2007.
- [9] G Fang and Y Wu. Frequent spatiotemporal association patterns mining based on granular computing. *Informatica (Slovenia)*, 37(4):443–453, 2013.
- [10] A. Hana, Y.T. Sami, and F. Sami. Mining spatiotemporal associations using queries. 2012 International Conference on Information Technology and e-Services, ICITeS 2012, 2012
- [11] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, Nov 1973. ISSN 0018-9472. doi: 10.1109/TSMC.1973.4309314.
- [12] J. Huo, J. Zhang, and X. Meng. On co-occurrence pattern discovery from spatio-temporal event stream. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8181 LNCS(PART 2):385–395, 2013.
- [13] J. Kawale, S. Liess, V. Kumar, U. Lall, and A. Ganguly. Mining time-lagged relationships in spatiotemporal climate data. pages 130–135, 2012.
- [14] Xiangjie Kong, Zhenzhen Xu, Guojiang Shen, Jinzhong Wang, Qiuyuan Yang, and Benshi Zhang. Urban traffic congestion estimation and prediction based on floating car trajectory data. *Future Generation Computer Systems*, 61:97 – 107, 2016. ISSN 0167-739X. doi: <https://doi.org/10.1016/j.future.2015.11.013>.
- [15] T.C.W. Landgrebe, A. Merdith, A. Dutkiewicz, and R.D. Mafaler. Relationships between palaeogeography and opal occurrence in australia: A data-mining approach. *Computers and Geosciences*, 56: 76–82, 2013.

- [16] Angela Lausch, Andreas Schmidt, and Lutz Tischendorf. Data mining and linked open data: New perspectives for data analysis in environmental research. *Ecological Modelling*, 295(0):5 – 17, 2015. ISSN 0304-3800. doi: <http://dx.doi.org/10.1016/j.ecolmodel.2014.09.018>.
- [17] A. Madraky, Z.A. Othman, and A.R. Hamdan. Analytic methods for spatio-temporal data in a natureinspired data model. *International Review on Computers and Software*, 9(3):547–556, 2014.
- [18] A.a Mohan and P.Z.b Revesz. Applications of spatio-temporal data mining to north platter river reservoirs. *ACM International Conference Proceeding Series*, pages 306–309, 2014.
- [19] NOAA. www.solarmonitor.org, April 2016. Last Access April 13, 2016.
- [20] K.G. Pillai, R.A. Angryk, J.M. Banda, M.A. Schuh, and T. Wylie. Spatio-temporal co-occurrence pattern mining in data sets with evolving regions. *Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012*, pages 805–812, 2012.
- [21] K.G.a Pillai, R.A.b Angryk, and B.b Aydin. A filter-and-refine approach to mine spatiotemporal cooccurrences. pages 104–113, 2013.
- [22] Vangipuram Radhakrishna, Shadi A. Aljawarneh, P.V. Kumar, and V. Janaki. A novel fuzzy similarity measure and prevalence estimation approach for similarity profiled temporal association pattern mining. *Future Generation Computer Systems*, 2017. ISSN 0167-739X. doi: <https://doi.org/10.1016/j.future.2017.03.016>.
- [23] K Venkateswara Rao, A Govardhan, and KV Chalapati Rao. Spatiotemporal data mining: Issues, tasks and applications. *International Journal of Computer Science & Engineering Survey (IJCSES) Vol, 3:39–52*, 2012.
- [24] Md. Mamunur Rashid, Iqbal Gondal, and Joarder Kamruzzaman. A technique for parallel sharefrequent sensor pattern mining from wireless sensor networks. *Procedia Computer Science*, 29(0):124 – 133, 2014. ISSN 1877-0509. doi: <http://dx.doi.org/10.1016/j.procs.2014.05.012>. 2014 International Conference on Computational Science.
- [25] Md.M. Rashid, I. Gondal, and J. Kamruzzaman. Mining associated sensor patterns for data stream of wireless sensor networks. pages 91–98, 2013.
- [26] Marcela Xavier Ribeiro, Agma J. M. Traina, and Caetano Traina, Jr. A new algorithm for data discretization and feature selection. In *Proceedings of the 2008 ACM symposium on Applied computing, SAC '08*, pages 953–954, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-753-7. doi: 10.1145/1363686.1363905.
- [27] James A. Rodger. Toward reducing failure risk in an integrated vehicle health maintenance system: A fuzzy multi-sensor data fusion kalman filter approach for ivhms. *Expert Systems with Applications*, 39(10):9821 – 9836, 2012. ISSN 0957-4174. doi: <https://doi.org/10.1016/j.eswa.2012.02.171>.
- [28] W. Sammouri, E. Cafame, L. Oukhellou, and P. Aknin. Mining floating train data sequences for temporal association rules within a predictive maintenance framework. *Lecture Notes in Computer subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, 7987 LNAI:112–126, 2013.
- [29] T. Scheffer. Finding association rules that trade support optimally against confidence. pages 9: 381–395, 1995.
- [30] Carlos Roberto Silveira-Junior. github.com/carlossilveirajr/mitsai, April 2017. Last Access April 17, 2017.
- [31] Carlos Roberto Silveira-Junior, Danilo Codeco Carvalho, Marilde Terezinha Prado Santos, and Marcela Xavier Ribeiro. Incremental mining of frequent sequences in environmental sensor data. In *The Twenty-Sixth International FLAIRS Conference (2015)*, pages 452–455, .
- [32] Carlos Roberto Silveira-Junior, Marilde Prado Santos, and Marcela Ribeiro. Stretchy time pattern mining: A deeper analysis of environment sensor data. pages 468–473, .
- [33] Carlos Roberto Silveira-Junior, Marcela Xavier Ribeiro, and Marilde Terezinha Prado Santos. A flexible architecture to integrate the solar satellite image time series data - the setl architecture. Pages 1–14, 2017. No publish by the time of this paper preparation.
- [34] Olga Spatenkovaa and Kirsi VIRRANTAU. Discovering spatio-temporal relationships in the distribution of building fires. *Fire Safety Journal*, 62, Part A:49 – 63, 2013. ISSN 0379-7112. doi: <http://dx.doi.org/10.1016/j.firesaf.2013.07.001>. Special Issue on Spatial Analytical Approaches in Urban Fire Management.

- [35] Fenzhen Su, Chenghu Zhou, and Wenzhoung Shi. Geoevent association rule discovery model based on rough set with marine fishery application. In Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International, volume 2, pages 1455–1458 vol.2, Sept 2004. doi: 10.1109/IGARSS.2004.1368694.
- [36] Edi Winarko and John F. Roddick. Armada: An algorithm for discovering richer relative temporal association rules from interval-based data. *Data & Knowledge Engineering*, 63(1):76 – 90, 2007 ISSN 0169-023X. doi: <http://dx.doi.org/10.1016/j.datak.2006.10.009>. Data Warehouse an Knowledge Discovery, 7th International Congress on Data Warehouse and Knowledge Discovery.
- [37] J.S. Yoo and M. Bow. Mining spatial colocation patterns: A different framework. *Data Mining and Knowledge Discovery*, 24(1):159–194, 2012.
- [38] B. Zaragoza, A. Rabasa, J.J. Rodriguez-Sala, J.T. Navarro, A. Belda, and A. Ramon. Modelling farmland abandonment: A study combining gis and data mining techniques. volume 155, pages 124 – 132, 2012. doi: <http://dx.doi.org/10.1016/j.agee.2012.03.019>.

AUTHORS

Carlos Roberto Silveira Junior is a System Architect at Ericsson expert in software development for telecom systems. He is a Ph.D. Student at Federal University of São Carlos (UFSCar) working with Data Mining -Artificial Intelligence- with images processing and parallelism. Holds a Master Degree in Data Mining and ontology at UFSCar, graduated Computer Science at UFSCar. Main areas of interest: data mining, programming, and distributed system.

José Roberto Cecatto holds a BS in Physics from the Pontifical Catholic University of São Paulo, a Master's and a PhD in Astrophysics from the National Institute for Space Research (INPE). He is currently a researcher at INPE. He has experience in Astronomy, with an emphasis on Radio Astronomy and instrumental development, acting mainly on the following themes: solar, radio, spectroscopy and interferometry. Currently, working in Space Climate with activities and development of tools related to the forecast of solar phenomena.

Marilde Terezinha Prado Santos is an associate professor at the Center for Exact Sciences and Technology / Department of Computing at the Federal University of São Carlos. PhD in Science with emphasis in Computational Physics from the University of São Paulo (2000). Master in Computer Science from the Federal University of São Carlos (1994) and Bachelor in Computer Science from the Pontifical Catholic University of Rio Grande do Sul (1991). Main areas of interest: engineering and applications of crisp and fuzzy ontologies, information retrieval, data mining, data integration and semantic web.