Astronomy and Computing 13 (2015) 136–144

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

Mining spatiotemporal co-occurrence patterns in solar datasets

B. Aydin [a,*], D. Kempton [a], V. Akkineni [a], R. Angryk [a], K.G. Pillai [b]

[a] *Georgia State University, Department of Computer Science, United States*

[b] *Montana State University, Department of Computer Science, United States*

ARTICLE INFO

Article history:
Received 2 February 2015
Received in revised form
29 August 2015
Accepted 14 October 2015
Available online 23 October 2015

Keywords:
Spatiotemporal co-occurrence
Frequent pattern mining
Solar data mining
Spatiotemporal indexing

ABSTRACT

We address the problem of mining spatiotemporal co-occurrence patterns (STCOPs) in solar datasets with extended polygon-based geometric representations. Specifically designed spatiotemporal indexing techniques are used in the mining of STCOPs. These include versions of two well-known spatiotemporal trajectory indexing techniques: the scalable and efficient trajectory index and Chebyshev polynomial indexing. We present a framework, STCOP-MINER, implementing a filter-and-refine STCOP mining algorithm, with the indexing techniques mentioned for efficiently performing data analysis.

1. Introduction

We have observed the emergence of massive solar datasets with the launch of NASA's Solar Dynamics Observatory. Solar data can be considered to have two major parts: (1) raster data, which comprise the images; and (2) vector data, which are the polygon-based representations of solar events. In this work, we address the problem of mining spatiotemporal co-occurrence patterns (STCOPs) from massive solar event datasets (namely, vector data). Existing data access methods, used in spatiotemporal knowledge discovery techniques for solar data mining, do not fully fulfil the requirements for the following reasons: (1) complex data representations of spatial and temporal data extensions; and (2) implicit semantics lying in the spatiotemporal objects, such as co-occurrences, sequences, and periodicity. Recently, many data mining algorithms have been proposed for various types of spatiotemporal patterns. Notable algorithms include discovery of spatial co-location patterns by Shekhar and Huang (2001), discovery of co-location episodes by Cao et al. (2006), mining mixed-drove STCOPs by Celik et al. (2006), and mining STCOPs from evolving polygon-based datasets by Pillai et al. (2012, 2013). Nevertheless, none of them, to our knowledge, used specifically tuned spatiotemporal indexing mechanisms to increase the performance of their algorithms.

The discovery of STCOPs can help us predict the relationships among event types, and also us help perform a large-scale verification of science questions in various scientific fields, such as astronomy, meteorology, and geophysics. An important application area of STCOP mining is the discovery of co-occurring solar events. Spatiotemporal co-occurrences frequently appear among various types of solar event instances. Identifying STCOPs on the Sun can help us better understand the relationships among solar event types and lead to better modeling and forecasting of important events, such as coronal mass ejections and solar flares, which impact radiation in space, can reduce the safety of space and air travel, disrupt intercontinental communication and GPS, and even damage power grids (Langhoff and Straume, 2012).

The performance of spatiotemporal data mining algorithms is contingent on efficient data access methods. Because of the massive datasets that are needed for the generation of significant outcomes from data mining applications, use of a memory-based model is not applicable in almost all cases; hence, disk-based storage is more widespread. Improving the performance of data access mechanisms in the spatiotemporal pattern mining field is not the primary area of interest for the data mining research community, although it immensely effects the overall performance. A very similar situation regarding the input–output (I/O) intensiveness of data mining applications applies to STCOP mining. In addition to the basic data retrieval challenges, computationally more expensive tasks (e.g., spatial overlay operations) are necessary for mining spatiotemporal data. Retrieving spatiotemporal data without a proper indexing structure has higher I/O costs, because of the lack of a natural ordering on spatiotemporal dimensions. Additionally, the solar event datasets being used in STCOP mining algorithms are historical and persistent. In other words, the current and future positions are not necessary for the mining process.

* Corresponding author.
E-mail address: baydin2@cs.gsu.edu (B. Aydin).

http://dx.doi.org/10.1016/j.ascom.2015.10.003
2213-1337/

In this work, we present our framework, Stcop-Miner, for efficiently mining STCOPs in solar event datasets. The Stcop-Miner framework has three loosely coupled modules: Miner, Indexing, and Data Access. The Data Access module is responsible for data I/O operations, whereas the Indexing module is in charge of indexing spatiotemporal data. The Miner module performs the actual pattern mining–related operations. In the Indexing module of the framework, we adopt two well-studied trajectory indexing techniques: *the scalable and efficient trajectory index* (*SETI*) (Chakka et al., 2003) and *Chebyshev polynomial indexing* (*CPI*) (Cai and Ng, 2004). We purposely selected trajectory-based spatiotemporal indexing techniques, since STCOP mining principally considers trajectory data (Pillai et al., 2012, 2013; Aydin et al., 2014b). The indexing techniques are modified to handle polygon-based spatiotemporal data. Moreover, the algorithm presented in Pillai et al. (2013) is improved by our further exploiting the computational efficiency.

As the task of mining STCOPs is concerned with spatiotemporal instances with polygon-based spatial representations, we specifically tuned our indexing techniques for polygon-based trajectory data; however, indexing point-based spatial representations is beyond the scope of this article. Our primary goal is to demonstrate the efficiency of the use of indexing techniques for mining STCOPs; therefore, other spatiotemporal pattern mining algorithms are not in the scope of our work. Nevertheless, the performance of other algorithms may also benefit from spatiotemporal indexing techniques.

The rest of this article is organized as follows. In Section 2, the related spatiotemporal pattern mining algorithms are discussed. In Section 3, we explain preliminary concepts on STCOP mining. We present our Stcop-Miner framework in Section 4. In Section 5, we present our experiments on Stcop-Miner using two different indexing strategies. Lastly, we discuss future work issues and present our conclusions in Section 6.

2. Related work

In recent years, many algorithms in the field of spatiotemporal pattern mining have been introduced. "Spatial co-location" refers to the spatially clöose relationships among spatiotemporal objects (Andrienko and Andrienko, 2007), and spatial co-location pattern mining was introduced in Shekhar and Huang (2001). Collocation episodes are defined as the common sequences of regular movement occurrences. Cao et al. (2006) proposed a solution for the discovery of collocation episodes. Mixed-drove STCOPs are the common occurrences of spatiotemporal instances that are often located in both spatial and temporal proximity, and extensive algorithms to mine mixed-drove STCOPs are given in Celik et al. (2006). Other pattern mining algorithms for discovering various types of spatiotemporal semantics from large-event datasets can be found in the literature, such as spatiotemporal sequential patterns (Huang et al., 2008), cascading spatiotemporal patterns (Mohan et al., 2012), and spatiotemporal mobility patterns (Bayir et al., 2010). STCOPs, proposed in (Pillai et al., 2012), are concerned with patterns regarding exact spatiotemporal co-occurrences (meaning spatial and temporal intersections of polygon representations instead of spatially and temporally close relationships). Differently from co-locations, mixed-drove co-occurrences, or partial co-occurrences, STCOPs handle spatiotemporal event instances with evolving, polygon-based representations. In this work, we present our extension to the Apriori-based filter-and-refine algorithm for the discovery of STCOPs presented in Pillai et al. (2013) using spatiotemporal trajectory indexing techniques that are capable of and specifically tuned for handling polygon data types that evolve over time.

3. Preliminary concepts

3.1. Spatiotemporal co-occurrence patterns

As mentioned earlier, we have used the STCOP mining algorithm introduced in Pillai et al. (2013), which mines the data established on an Apriori-based algorithm that effectively prunes the candidates using a filter-and-refine strategy. Given a set of spatiotemporal feature types and spatiotemporal instances associated with these feature types, an STCOP is a subset of the set of all features, whose instances frequently intersect in both a temporal and a spatial context. In the following subsections, we provide descriptions of the preliminary concepts regarding STCOPs, and they follow the earlier work presented in Pillai et al. (2013) and Aydin et al. (2014b).

3.1.1. Instance

An instance, denoted as *Inst*, is a spatiotemporal trajectory-based object with a continuously evolving polygon-based spatial representation. Each instance is identified with an *InstanceId*, and has a feature type. Instances have start and end times, and for each valid timestamp, they have polygon-based geometric representations. The set of all instances is denoted by $\mathbb{I}$, and the set of all instances of a particular feature type (f_i) is denoted by $\mathbb{I}_{f_i}$.

3.1.2. Feature type

A feature type is a non-spatiotemporal attribute of an instance that signifies the sort (or class) of that particular instance. A feature type is denoted by f_i, and the set of all features is denoted by $\mathbb{F} = \{f_1, f_2, \ldots, f_m\}$.

3.1.3. Pattern

An STCOP is a subset of all feature types, whose instances frequently co-occur in both space and time. A pattern is denoted by P, where $P \subset \mathbb{F}$. The number of feature types in a STCOP will be referred as the *cardinality*.

3.1.4. Pattern instance

Given a k-cardinality STCOP, $P = \{f_{i_1}, \ldots, f_{i_k}\}$, a pattern instance (denoted by *PInst*) of P is a unique incidence of a spatiotemporal co-occurrence (spatial and temporal overlap) among the instances of *all* the feature types in P. (Note that similarly to instances, pattern instances are identified by a unique pattern instance identifier.) Pattern instances have start and end times, but for each valid timestamp, a pattern instance will have two kinds of spatial representation, one for the intersection and the other for the union geometries.

3.2. Measures

We used two types of objective measures in our algorithm. The first one is the co-occurrence coefficient (*cce*), which is used to assess the strength of spatiotemporal overlap in a pattern instance. The second one is the prevalence measure (p), and it is used to evaluate the prevalence of a pattern using the participation of instances of different event types.

3.2.1. Significance measures

The spatiotemporal co-occurrence coefficient is used to determine the strength of a spatiotemporal relationship. In our case, the relationship is the spatiotemporal overlap of two or more instances. To assess the strength of spatiotemporal overlap, we use the commonly used *Jaccard* (J) measure as the co-occurrence coefficient. Moreover, the computationally less expensive *OMAX* measure is used to filter pattern instances.