

Efficient discovery of emerging patterns in heterogeneous spatiotemporal data from mobile sensors

Francisco Neves
INESC-ID and IST, Univ. de Lisboa
francisco.neves@tecnico.ulisboa.pt

Anna C. Finamore
INESC-ID and IST, Univ. de Lisboa
anna.couto@tecnico.ulisboa.pt

Rui Henriques
INESC-ID and IST, Univ. de Lisboa
rmch@tecnico.ulisboa.pt

ABSTRACT

Heterogeneous sensor networks, including traffic monitoring systems and telemetry systems, produce massive spatiotemporal data. Geolocated time series data and timestamped trajectory data are generally produced from fixed and mobile sensors in these systems, offering the possibility to detect events of interest. Events of interest generally comprise emerging and gradual changes in the behavior of those systems, including patterns of congestion in road, utility and communication networks. However, the comprehensive discovery of these actionable events is challenged by the: i) inherently spatiotemporal and heterogeneous nature of data produced by different sensors; ii) difficulty of detecting emerging patterns not yet markedly noticeable at early stages; and iii) massive data size.

This work proposes E2PAT, a scalable method to comprehensively detect emerging patterns from heterogeneous sources of spatiotemporal data generated by large sensor networks. We combine simplistic time differencing and spatial intersection principles to identify all emerging patterns distributed along geographies of interest. We show that the use of these principles guarantee a linear-time efficiency of E2PAT on the size of the input data. In addition, we propose an integrative score to measure the relevance of emerging patterns and show its role to support pattern retrieval, promote usability, and guarantee the actionability of the found patterns. These contributions are comprehensively assessed in the context of the Lisbon’s road traffic monitoring system, a large-scale network of mobile and fixed sensors. The gathered results confirm the actionability of the found patterns and the scalability of E2PAT. E2PAT is provided as an open-source tool available at <https://github.com/francisconeves97/emerging-patterns>.

KEYWORDS

heterogeneous sensor networks, spatiotemporal data, emerging patterns, geolocated time series, timestamped trajectory data, sustainable mobility

ACM Reference Format:

Francisco Neves, Anna C. Finamore, and Rui Henriques. 2020. Efficient discovery of emerging patterns in heterogeneous spatiotemporal data from mobile sensors. In *Proceedings of MobiQuitous*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/1122445.1122456>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MobiQuitous, 2020, Cyberspace

© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00
<https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

Heterogeneous sensors placed within complex systems – whether urban, physiological, mechanical, digital, geophysical, or communication systems – generate spatiotemporal data that offer the possibility to acquire comprehensive views of systems’ behavior along time. Discovering emerging patterns in such heterogeneous sensor networks is essential to identify important changes that reveals needs for actuation [2]. Illustrating, increased utility, processing and communication needs along certain routes of a supply network may reveal future bottlenecks, propelling dynamic rebalancing initiatives. In urban mobility, emerging patterns reveal ongoing changes in city traffic dynamics, whose growth along time may indicate the establishment of new congestion trends with impact on the normal traffic flow [23, 31]. Those trends can evolve to create traffic bottlenecks if timely precautions are not taken [7]. As such, the early detection of emerging patterns offers urban planners the opportunity to make the necessary provisions to urban mobility.

Despite the relevance of emerging pattern discovery in heterogeneous sensor networks, this task is hindered by the major challenges. First, the inherent spatiotemporal nature of the available data. In the context of road traffic, massive timestamped trajectory data produced by vehicles with mobile sensors, as well as geolocated time series data produced by inductive loop counters, are often available [17]. Second, the need to combine multiple views from the heterogeneous data produced by a network of mobile and fixed sensors. In the context of road traffic, this implies combining distinct views on traffic flow, including speed limits, congestion size, congestion recurrence, experienced delay, or vehicle throughput from speed and loop-counter sensors. Third, the need for a robust and timely detection of emerging patterns [1]. In the context of road traffic, this implies a comprehensive geographical traversal of changing dynamics at an earlier stage. Fourth, the massive size of signal data produced by large sensor networks. In the context of road traffic, cities can be equipped with hundreds of loop counters and millions of active mobile devices [45], producing abundant data along time. Fifth, and finally, the need to guarantee the actionability, interpretability, and navigability of emerging pattern’s solutions.

This work proposes a linear-time method to comprehensively discover emerging patterns from heterogeneous spatiotemporal data that is able to address the listed challenges. To this, we combine three simplistic yet effective operations – time series differencing, spatial intersection and regression calculus – for the efficient discovery of all emerging patterns observed along geographies of interest. In addition, we propose an integrative score to measure the relevance of emerging patterns that yield statistical properties of interest.

The proposed method, referred as E2PAT (Emerging Event Pattern miner), offers four major contributions. First, E2PAT is able

to comprehensively mine both emerging patterns and trends from heterogeneous spatiotemporal data collected along a geography of interest. In particular, we show its efficacy in dealing with timestamped trajectory data and geolocalized time series. Second, E2PAT has a linear computational complexity on the size of the input data. Third, E2PAT combines growth rate, support and fitting-error criteria to guarantee pattern actionability and minimize the presence of both false positive patterns (retrieved yet not relevant) and false negative patterns (not retrieve yet relevant). Fourth, and finally, we show that the proposed scoring schema has a key role in assessing the relevance of an emerging pattern, which guides the navigation along pattern solutions. In particular, E2PAT implements modal, spatial and temporal zoom facilities to guarantee the usability of the pattern analysis process.

To assess the significance of the proposed contributions, E2PAT is extensively evaluated against the Lisbon's road traffic monitoring system, a system that combines fixed sensors – inductive loop detectors at the major road junctions in the city – and sensors from mobile devices – geolocalized speed detectors from citizens and visitors. The gathered results confirm the relevance of the proposed E2PAT method, highlighting the aforementioned contributions, and further illustrating the relevance of the found patterns to support mobility decisions.

E2PAT is provided as both a graphical and programmatic tool satisfying strict usability criteria.

The manuscript is structured as follows. Section 2 provides essential background on spatiotemporal data analysis. Section 3 surveys relevant work on detecting actionable changes patterns in complex systems. Section 4 proposes E2PAT, introducing principles, Section 5 gathers and discusses the results from experimentally assessing E2PAT in urban systems. Finally, Section 6 offers the major concluding remarks and implications.

2 BACKGROUND

This section provides essential background to understand the target task: emerging pattern discovery from spatiotemporal data. Section 2.1 describes the spatiotemporal data structures that can be inputted for this task, with a focus on georeferenced time series from stationary devices, and spatiotemporal event data from mobile devices. Section 2.2 introduces the concept of emerging pattern, describing the desirable properties to be pursued during their discovery.

To illustrate and validate the contributions, we consider road traffic monitoring systems as the study case. In the context of this work, a road traffic monitoring system is an heterogeneous network composed of: i) inductive loop counters, stationary devices typically found in major road junctions in a city; and ii) geolocalized speed meters from active mobile devices used along private and public road trips.

2.1 Spatiotemporal data from sensor networks

Georeferenced time series. Signals produced by *fixed sensors* are generally represented as *time series*, an ordered set of observations $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$, each observation \mathbf{x}_t being recorded at a specific time point t . Time series can be *univariate*, $\mathbf{x}_t \in \mathbb{R}$, or *multivariate*, $\mathbf{x}_t \in \mathbb{R}^m$, where $m > 1$ is the multivariate order (number of variables). Time series recorded at a particular location are

mentioned as georeferenced. A *georeferenced time series* is a tuple $GT = (\phi, \mathbf{x})$, where ϕ is a pair (*latitude, longitude*) describing the location where the series \mathbf{x} is being recorded.

Time series can be decomposed into *trend*, *seasonal*, *cyclical*, and *irregular components* using additive or multiplicative models [4]. Classical approaches for time series analysis generally rely on statistical principles, including *auto-regression*, *differencing* and *exponential smoothing* operations [43]

In the context of road traffic monitoring systems, *inductive loop detectors* (ILDs) are equipment installed under roads pavements for detecting vehicle passages. Depending on the type of ILD, these equipments are able to detect variables such as the volume, speed and class of vehicles passing (more information in [24]). ILD raw signal data are thus often aggregated to provide frequentist views on the cumulative number or average speed of different classes of vehicles on a given road along specific time intervals, i.e. georeferenced multivariate time series data.

Spatiotemporal event data. A *trajectory* is a sequence $\langle \phi_1, \phi_2, \dots, \phi_n \rangle$, where ϕ_i is a pair (*latitude, longitude*). A timestamped trajectory, $\langle (\phi_1, t_1), \dots, (\phi_n, t_n) \rangle$, has its coordinates, ϕ_i , annotated with a timestamp, t_i .

Floating car data are paradigmatic examples of timestamped trajectory data produced from mobile devices with active global positioning systems (GPS), gathering the position of vehicles along time. Methods for producing floating car data from GPS information generally produce rather sparse trajectories that need to be completed within the constraints of the road network mesh [3, 5, 6]. These individual trajectories produced by mobile devices can be aggregated, and specific features of interest (such as speed) extracted from devices circulating throughout the same trajectory segments at similar time periods [33]. Applications, such as [GOOGLEMAPS](https://www.google.com/maps)¹, [WAZE](https://www.waze.com/en-GB/)² or [TOMTOM](https://www.tomtom.com/en_gb)³, installed in some of the mobile devices, offer localization and navigation facilities, providing an aggregate view of the ongoing traffic dynamics within the city.

In this context, an *event* is a tuple $E = (\mathbf{x}, s, \tau)$, where:

- $\mathbf{x} = (x_1, \dots, x_m)$ is the observation, either *univariate* ($m=1$) or *multivariate* ($m > 1$) depending on the number of monitored variables. For instance, given speed (y_1) and throughput (y_2) variables, an illustrative observation is $\mathbf{x}=(x_1=15\text{km/h}, x_2=10\text{cars/min})$;
- s is the *spatial extent* of the observation \mathbf{x} . The spatial extent s can be any spatial representation associated with the event, such as a *geographic coordinate* or a *trajectory*;
- τ is the *temporal extent* of the observation \mathbf{x} , either given by a time instant or a time interval.

A *spatiotemporal event dataset* is a collection of events, $E = \{e_1, e_2, \dots, e_n\}$, each event producing a (multivariate) observation recorded along specific spatial and temporal context.

In the context of road traffic monitoring systems, an event generally corresponds to a significant change to the regular speed along a given road segment (trajectory) or location (coordinate). Geolocalized speed data can thus be seen as a collection of events, where

¹<https://www.google.com/maps>

²<https://www.waze.com/en-GB/>

³https://www.tomtom.com/en_gb

each event, $e_i = (x_i, s_i, \tau_i)$, is a traffic jam event that occurred at time t_i in a trajectory (road segment) s_i . The set of observations x_i contains traffic information – such as the recorded speed, delay, severity level or road type – that characterizes the occurring jam.

2.2 Emerging pattern discovery

Spatiotemporal pattern mining. Given a spatiotemporal dataset (section 2.1), a **pattern** is a spatially correlated set of frequent, periodic or coherently changing observations along time. Illustrating, periodic patterns describe recurrent behavior over regular time intervals at certain locations or trajectories. Graph patterns are sets of trajectories or locations within the target sensor network that are frequently co-associated with an event of interest (such as co-occurring congestions). Temporal association rules define hypothetical causal relationships between correlated frequent events at nearby locations or trajectories.

For each of these pattern solutions, different criteria of interest can be measured: i) *pattern support*, the number of observations satisfying the pattern; ii) *pattern length*, the multivariate order and spatiotemporal extension of the given pattern; and iii) *pattern strength*, including confidence, lift and interestingness, defines the association strength among the elements composing a pattern.

Emergin patterns. Emerging Patterns (EPs) were firstly introduced by Dong et al. [8] in the context of multivariate observations collected from two periods/datasets. An emerging pattern was in this context defined as a multivariate pattern whose support suffered a significant change between the two given periods.

This work extends this early notion of emerging pattern to encompass an arbitrary number of time periods and to further incorporate spatial information. Given a spatiotemporal dataset, an **emerging pattern** is a set of spatially correlated observations whose values satisfy specific *growth*, *fitness* and *support* criteria along time.

The *growth* criterion defines the rate at which observations change along time. For instance, given a specific location and periodicity, a growth rate of 1% indicates that the values of a given observation increase 1% on every period under assessment.

Given a specific growth rate, the *fitness* (error) criterion defines how well observations follow (deviate) from the given expectations. For instance, fluctuations of the observed values around the expected values produce residues that can be used to characterize the fitness (error) of a given emerging pattern. Emerging patterns below a given accuracy threshold may be spurious findings and should therefore be discarded.

Finally, *support* criterion defines the number of observations (temporal extent) satisfying the given growth and accuracy criteria. In this context, emerging pattern discovery can be applied under minimum growth, accuracy and support thresholds.

In the context of traffic monitoring systems, a pattern of road traffic is a coherent form of traffic behavior that satisfies a specific criterion of periodicity, frequency, or growth. An illustrative emerging pattern of road traffic is:

$\{(speed\ limit\ decrease\ at\ weekly\ growth\ rate\ 2\% \mid trajectory\ s_A),$
 $(traffic\ throughput\ increase\ at\ weekly\ growth\ rate\ 3\% \mid location\ \phi_B)\}$

where = Areeiro, when = ([10h,11h[\wedge Mondays)
 satisfying $r^2 > 0.5 \wedge support > 10$

where the coefficient of determination is used as the fitness criterion and at least 10 observations (support criterion) necessary to infer the observed growth rates, 2% and 3%, from congestions on a segment s_A and location ϕ_B within the Areeiro region at Mondays, 10h.

In alternative to speed limits and traffic flow, emerging patterns of road traffic may further capture growing mobility restrictions associated with the congestions' extent, recurrence, average delay per distance, and severity.

Pattern quality. In addition to the introduced *growth*, *quality* and *support* criteria, emerging patterns should further satisfy the following properties of interest:

- *non-triviality* (novelty) and *actionability* (support decisions);
- *robustness* (bounded noise tolerance);
- *statistical significance* (excluded spurious patterns that occur by chance);
- *interpretability*;
- *coverage* (complete solutions spanning different geographies and time periods);
- *efficiency* of the pattern retrieval process.

Target problem. Given the introduced sources of spatiotemporal data in section 2.1 (input), as well as the desirable pattern properties in section 2.2 (output), the problem targeted in this work is the efficient and effective discovery of emerging patterns.

3 RELATED WORK

Emerging pattern discovery. The concept of emerging patterns (EP) can be traced back to two different research streams. In time series data analysis, the discovery of emerging behaviors generally corresponds to the modeling of non-linear trends within a time series [16]. In this field, emerging behaviors are generally approximated using non-linear regressive or auto-regressive models, including regime switching models and neural network models, approximated on the original time series or on a decomposed series after removing seasonal and cyclical components [11].

In the pattern mining field, emerging behaviors were in 1999 coupled with the pattern concept, implying the satisfaction of well-defined frequency criteria. An emerging pattern (EP), as firstly introduced by Dong et al. [8], is a set of data instances whose characteristics entail significant changes between two (or more) timestamped datasets. Since then, this original notion of EPs has been extended and mostly applied in bioinformatic domains [8, 27, 28, 32, 34]. Nevertheless, and to our knowledge, EPs have not yet been extended towards spatiotemporal data structures neither applied in the context of sensor networks.

Dong et al. [8] observed that, given the nature of the early-formulated EPs (larger in size and small in support), naïve algorithms are costly and the Apriori property does not hold for EPs, proposing dedicated EP searches, and further extending these searches to build a classifier. Similarly, a substantial number of following works proposed the use EPs in classification tasks [26, 27, 27, 32]. Liu et al. [27] monitored gene expression under

varying conditions, aiming at detecting trends under these conditions across genes for cells of the same type in order to predict the class of cells from the underlying EPs. Li et al. [28] extended these contributions using more efficient search variants. Fan et al. [9] propose a hybrid version of previous EP classifiers and Naive Bayes, yielding interpretation facilities.

Novak et al. [34] presented a survey on supervised descriptive rule discovery, a framework combining contrast set mining (CSM), emerging pattern mining (EPM), and sub-group discovery (SD). They explain that "while all these research areas aim at discovering patterns in the form of rules induced from labeled data, they use different terminology and task definitions, claim to have different goals, claim to use different rule learning heuristics, and use different means for selecting subsets of induced patterns".

In terms of efficiency, Fan et al. [10] in an effort addressed the issue of a large number EPs being generated by EP mining approaches by proposing an algorithm which considers only interesting EPs. This interestingness score is based on: support, growth rate, and a relationship between EPs and statistical measures. Soulet et al. [40] further proposed condensed representations of EPs based on the classic concept of frequent closed pattern.

More recent contributions extend EP discovery towards large-scale data [13, 14] by combining evolutionary fuzzy systems with the MapReduce paradigm; three-dimensional data via triclustering algorithms [22]; collections of events using both generative and deterministic approaches [20, 21]; as well as streaming data [41] by combining evolutionary algorithms with batch strategies.

Song et al. [38] developed a methodology to detect and assess emerging, unexpected and added/perished changes in customer behavior taking into consideration customer profiles and sales data along time. In the same domain, Chen et al. [44] propose association rule discovery along different time periods. They extend the early Song et al. [38] concepts towards emerging, unexpected and added rules and propose corresponding evaluation measures of growth, difference, and modified difference. Li et al [25] considered data from online reviews to identify EPs of hotel features in order to give hotel managers' insights about travellers' interests and expectations. For additional contributions and applications on EP discovery, the reader is invited to consult the work of Garcia-Vico et al. [12]. Again, and despite the relevance of the surveyed contributions on EP discovery, its applicability towards spatiotemporal data structures and mobile sensor domains remains unexplored.

Spatiotemporal patterns of road traffic. Although general principles for spatiotemporal pattern mining have been identified [2, 30, 42], the existing contributions highly depend on the application domain, considerably varying in accordance with the inputted data structure and pursued patterns of interest. In the context of traffic monitoring systems, the discovery of actionable spatiotemporal mobility patterns has been considerably researched [31, 36, 37]. Classic approaches make use of statistics, parametric models and visualization principles to understand spatiotemporal traffic dynamics, with particular focus on highlighting discrepancies between origin-destination matrices [18, 31] and establishing views on changing traffic flows [17, 19, 29]. Clustering has been also applied to identify road segments with similar traffic load over time [33], detect vulnerabilities along the road network (recurrent jams over

time termed congestion pockets) [35] and other spatial associations assessed against external factors of influence [39]. Classic pattern mining algorithms have been also successfully extended to detect trajectory patterns (T-Patterns) [15] – location precedences with timing constraints that occur frequently among trajectory instances such as *railway station* $\xrightarrow{15min}$ *town square* $\xrightarrow{2h15min}$ *museum*; patterns of daily congestion and spatial propagation [7, 23]; amongst other patterns of urban mobility [45]. Despite their relevance, the concept of emerging behavior is to our knowledge absent from the existing approaches. In addition, current approaches are unable to provide consolidated pattern views from heterogeneous spatiotemporal data structures.

4 SOLUTION

As introduced, our work aims at efficiently discovering emerging patterns from heterogeneous sources of spatiotemporal data produced by sensor networks. In particular, and as the motivating study case, we focus on emerging patterns of road mobility from georeferenced time series data produced by stationary loop counters, and multivariate event collections produced by GPS sensors. This task is challenged by the inherently complex spatiotemporal nature, heterogeneity, and massive size of the target sensor data. To address these challenges, we propose a linear-time method to comprehensively discover emerging patterns, termed E2PAT (Emerging Event PATtern miner).

E2PAT combines three simplistic yet effective principles: i) spatial intersection and time windowing operations for the comprehensive traversal of search space (section 4.1); ii) combined use of time series differencing operations with linear regressors (section 4.2); and iii) integrative scoring to measure the relevance of emerging patterns and control the amount of false positive and false negative discoveries (section 4.3). E2PAT is available at [Github](#).

4.1 Spatiotemporal data mappings

E2PAT is a two-step process. First, transformation procedures are applied to consolidate the original spatiotemporal data sources and map them into new data structures more conducive to the subsequent mining task. To this end, spatial and temporal constraints can be inputted at this stage to guide the discovery.

Second, emerging patterns are discovered from the transformed data by combining differencing, regression and integrative scoring principles.

4.1.1 Spatial constraints. For handling trajectories of arbitrary length, there is the need to fix an adequate spatial granularity. E2PAT offers two major possibilities. First, E2PAT can rely on an already established categorization. For instance, street names in the context of road trajectory data or every segment between two junctures/nodes from a sensor network are supported criteria.

Second, the categorization can be automatically produced using a geographical mesh/grid for segmenting the set of all possible trajectories. The granularity of the input mesh can either create coarser or finer spatial views in comparison with the first option.

Under the selected spatial granularity, events are then linked to one or more segments in accordance with their spatial extent. To this end, simplistic yet efficient trajectory-mesh indexation and

trajectory-segment intersections are applied to associate events to segments in linear time (section 4.4).

Finally, the events associated with each one of the identified segments are temporally ordered to compose a *sparse (multivariate) time series* from each event’s timestamp, t , and (multivariate) observation, x .

4.1.2 Temporal constraints. Three major types of temporal constraints can be placed. First, calendrical constraints can be placed to segment the available data in (possibly overlapping) data chunks, and data transformations applied on each chunk. By default, the day of the week, weekdays, holidays, and on/off-academic period calendars are considered. Emerging patterns are discovered in linear-time for each one of these calendars (section 4.4).

Second, a time granularity (e.g. 15-minute, hour or on/off-peak intervals) can be optionally specified to guide the discovery of emerging behaviors. In its absence, the proposed pattern discovery is iteratively performed using multiple time aggregations. Note that emerging patterns are not detected over a continuous timeline due to daily traffic cycles. Instead, they are discovered on these time windows throughout the days of the previously fixed calendar.

Given a specific time granularity, georeferenced time series can then be resampled using aggregators (e.g. *sum* of vehicles per time interval, *average* vehicle speed per time interval). In the context of spatiotemporal event data, the sparse series produced under the principles introduced in previous section 4.1.1 are resampled using aggregation procedures in accordance with the target variables (e.g. event with *maximum* spatial extension per time interval, or severity *mode* from the occurring events).

Third, larger time windows, spanning a fixed number of weeks, can be optionally specified to guide the discovery of patterns whose emerging behavior was only recently elicited (spanning just a partial period). By default, a single window spanning all the available data is considered since late-occurring patterns can still be detected under the proposed differencing operations.

4.1.3 Data transformation. Once these constraints are fixed, data mappings are applied to transform the original spatiotemporal data structures into multivariate time series data, more conducive to the subsequent step. In the target structure, each observation corresponds to a day from a specific calendar at a specific time (see section 4.1.2) and each variable measures some aspect of the target system at a specific location. Considering road traffic monitoring systems, the transformed ILD data measures the number of cars passing over a single loop detector during a specific time interval for each calendar day. For the geolocalized speed data, multiple measurements are taken per event and principles from section 4.1.1 applied to aggregate events per road segment. Figure 1a and 1b show the original structures of ILD and geolocalized speed data.

The integration of the previous mappings is a simple concatenation of the time series variables resulting from the transformation of each data source, as shown in Figure 1c.

4.2 Series differencing and emergence

4.2.1 Comprehensive emerging pattern discovery. The proposed mapping generates as many multivariate time series as the number of calendars, time intervals per day, and fixed spatial granularity

date	cod	ct_id	latitude	longitude	count
2018-11-19 23:00:00	21	4	38.722828	-9.16808	142
2018-11-19 23:15:00	21	4	38.722828	-9.16808	140
2018-11-19 23:30:00	21	4	38.722828	-9.16808	381
2018-11-19 23:45:00	21	4	38.722828	-9.16808	328
2018-12-06 00:00:00	3	24	38.722702	-9.168654	390
2018-12-06 00:15:00	3	24	38.722702	-9.168654	869
2018-12-06 00:30:00	3	24	38.722702	-9.168654	345
2018-12-06 00:45:00	3	24	38.722702	-9.168654	345
2018-12-06 01:00:00	3	24	38.722702	-9.168654	342

(a) Original ILD data structure

date	street_coord	speed	street_name
2018-09-03 16:10:24.02	("type": "LineString", "coordinates": [[-9.145424, 3...	3	Av. da República (lateral)
2018-09-03 16:10:23.831	("type": "LineString", "coordinates": [[-9.204668, 3...	4	Av. Dr. Mário Soares
2018-09-03 16:10:25.042	("type": "LineString", "coordinates": [[-9.152298, 3...	1	R. Mq. de Sá da Bandeira
2018-09-03 16:10:23.965	("type": "LineString", "coordinates": [[-9.098841, 3...	0	Ac. IC2 Lisboa
2018-09-03 16:10:23.771	("type": "LineString", "coordinates": [[-9.134226, 3...	1	Av. Almirante Reis
2018-09-03 16:10:24.06	("type": "LineString", "coordinates": [[-9.146396, 3...	0	Av. da República (lateral)
2018-09-03 16:10:23.976	("type": "LineString", "coordinates": [[-9.198936, 3...	1	Av. 25 de Abril
2018-09-03 16:10:23.992	("type": "LineString", "coordinates": [[-9.210107, 3...	4	Pç. do Império

(b) Original geolocalized speed data structure

date	delay_R. Bernardim Ribeiro	spatial_extension_Av. da Liberdade	cod_3-id_9	cod_4-id_1
2018-11-09 22:00:00	0	0	4868	84
2018-11-10 22:00:00	0	0	2466	65
2018-11-11 22:00:00	0	0	0	0
2018-10-10 23:00:00	0	0	1973	352
2018-10-11 23:00:00	0	0	0	0
2018-10-12 23:00:00	0	0	0	0
2018-10-13 23:00:00	0	0	0	0
2018-10-14 23:00:00	0	0	0	0
2018-10-15 23:00:00	0	0	1863	73
2018-10-16 23:00:00	0	0	765	609
2018-10-17 23:00:00	0	0	3388	591

(c) Integrative series data structure

Figure 1: Original and transformed spatiotemporal data structures.

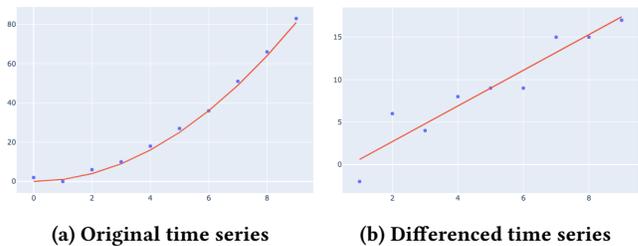


Figure 2: Role of time series differencing operations for detecting emerging trends using linear models.

(section 4.2), i.e. number of regions in the context of mobile sensors and sensorized locations in the context of stationary sensors.

E2PAT is then applied for each variable of each multivariate time series in order to detect isolated emerging behavior in accordance with differencing and regression principles (sections 4.2.2 and 4.2.3). The found emerging behaviors are then comprehensively assessed using integrative scoring principles (section 4.3) and combined to produce the target emerging pattern profiles.

4.2.2 Differencing. Given the fact that heterogeneous sensor networks produce massive data, learning non-linear (auto-)regressive models is a computationally expensive task, intractable to the target end. To tackle this observation, we make use of simple yet effective time series differencing operations.

Time series differencing is the act of subtracting consecutive observations from a time series, $x_{t+1} - x_t$. Time series differencing

has been traditionally applied in time series analysis to stationarize non-stationary time series. In the context of our work, we use this principle with a different end, to approximate emerging trends using linear regressions as illustrated in Figure 2. When a time series follows an exponential trend explained by a certain growth factor (e.g. Figure 2a), the differenced time series will have a linear behavior (e.g. Figure 2b). This turns time series differencing a robust candidate for the targeted task.

4.2.3 Regression. E2PAT allows the search for three types of patterns: simple, emerging and abruptly changing patterns. *Simple patterns* are simple trends in the original time series data, approximated by a linear regression with the dependent variable being a variable (e.g. speed limit or jam spatial extent) and the independent variable the calendar day.

Emerging patterns are trends when the target series variable is differenced, allowing us to capture exponential trends.

Finally, *abruptly changing patterns* are trends observed in time series after two differencing operations (i.e. second-order differenced series).

To identify the linear trends on the (differenced) time series, a simple linear regression is estimated on each variable using the least squares method, with slope, d ,

$$d = \frac{\sum_{i=1}^n x_i t_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n t_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{\sum_{i=1}^n x_i^2}{n}}, \quad (1)$$

and a coefficient of determination, r^2 ,

$$r^2 = \frac{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2)$$

Linear-time decomposition of the time series can be optionally applied to remove seasonal and cyclical components for a more correct approximation of the determination coefficient.

4.3 Integrative scoring

An integrative scoring, yielding statistical properties of interest, is proposed to quantify the relevance of patterns. The score is further used to filter the outputted emerging patterns in order to minimize the presence of both false positive patterns (retrieved yet not relevant) and false negative patterns (not retrieve yet relevant).

The proposed score function is influenced by four major attributes:

- the slope of the linear regression, measuring the growth rate of the pattern;
- the r^2 of the regressive model, measuring the accuracy term of the pattern (1 when optimal and near 0 when random);
- the relative support of the pattern (i.e. the number of observations in the pattern divided by the maximum number of observations found within the same data source);
- the differencing order (section 4.2.2), a term to favour emerging patterns against simple trends.

The proposed scoring function is given by:

$$\text{score}(d, r^2, \text{sup}, p) = (\alpha_1 \times d + \alpha_2 \times r^2 + \alpha_3 \times \text{sup})^{\alpha_4(1+p)} \quad (3)$$

where d is the slope of the fitted linear regression (1), r^2 is the coefficient of determination (2), sup is the relative support, and p is differencing order (0 if absent). $\alpha_1, \alpha_2, \alpha_3$ and α_4 are parameterizable weights for each factor that can obtain under a sensitivity analysis. Compelling empirical evidence from the Lisbon's road traffic system suggests $\alpha_1=0.3, \alpha_2=0.5, \alpha_3=0.2$ and $\alpha_4=1$ as default values.

The d term is bounded between -1 and 1, while r^2 and sup terms are bounded between 0 and 1. As a result, the proposed score function is also bounded between -1 and 1.

Because the sign of the score is dictated by the slope d , it easily informs whether we are in the presence of a congestion pattern (positive score) or a decongestion pattern (negative score). In the context of urban mobility, some of the monitored variables including speed limits, jam spatial extent, jam recurrence, jam delay, jam severity, and car frequencies. With the exception of speed limits, all the values measured for the remaining road traffic variables increase when congestion levels increase. As such, we change the sign of the slope of the speed limit variables, so that patterns with different road traffic variables can be interpreted seamlessly.

In addition to its easily interpretable bounds, the statistical distribution of the scores computed for the found patterns from spatiotemporal traffic data reveals that the observed values approximately follow a centered Gaussian distribution, passing tests of normality at $\alpha=0.05$ significance level.

4.4 Computation complexity of E2PAT

Theorem. E2PAT has linear time complexity on the input data size.

Proof. First, considering georeferenced time series data. Let us assume the presence of r stationary devices, each measuring m variables along T steps. Input data in this context has $O(rmT)$ size. Now consider the presence of k_1 calendars and the presence of k_2 time periods per day. The proposed transformation process will lead to the formation of $k_1 k_2 r$ time series, each with a m multivariate order. This leads to time series data of size $O(k_1 k_2 r m \frac{T}{k_2}) = O(k_1 r m T)$. Note that the number of calendars is always a small constant. For instance, considering the days of the week, where weekdays are further decomposed according to on and off-academic periods, $k_1=2+5 \times 2=12$. As such, the produced time series have size $O(rmT)$. As the data transformation step is just based on linear-time segmentation and resampling operations, the computational complexity of this step is in fact $O(rmT)$.

Three differencing operations, $k_3=3$, are applied to allow the discovery of emerging and abruptly changing patterns. As differencing is a linear operation, this step takes $O(k_3 r m T) = O(rmT)$ time. Finally, linear regressions are learned for each variable of each time series. Since the calculus of formulae (1) and (2) is also accomplished in linear time, the time complexity of this step is $O(k_3 r m T) = O(rmT)$.

Let us now consider spatiotemporal event data, and the presence of massive number of q events with varying timestamps and trajectories spanning different geographies. Let us consider the presence of a total number of k_4 trajectory segments. The production of these segments against a user-defined mesh can be understandably performed in linear time using a simple spatial data structure. In addition, the indexation of the q events into these segments has time complexity of $O(k_4 q)$. Since k_4 is also a constant, the time

complexity of the transformation stage is $O(q)$. Similarly as described for georeferenced time series data, the subsequent steps on the transformed event series with total size $O(qm)$ (where m is the multivariate order of the events) can be computed in linear time, yielding $O(qm)$ complexity. \square

5 RESULTS

Considering the Lisbon city as a study case, we applied the proposed approach to comprehensively discover emerging traffic patterns of road mobility from geolocalized speed data provided by WAZE and inductive loop detector (ILD) data collected during a two month period in central junctures of the city (Figure 3). To illustrate the enumerated potentialities, the gathered results from consolidated ILD-WAZE data sources are discussed.

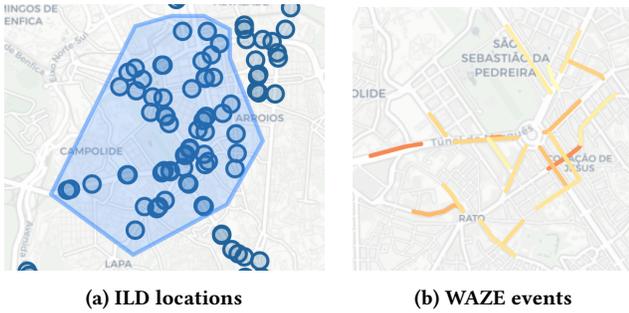


Figure 3: Map visualization of the two sources of urban traffic data along the studied area: a) ILD sensor placement; b) WAZE events.

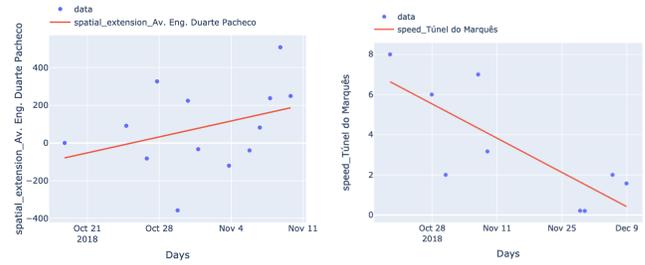
Experimental setting. The score parameters were chosen by empirically experimenting different values and validating the relevance of the best scored patterns with mobility experts from LNEC⁴ and CML⁵. Upon experimental analysis, a higher weight was allocated to the r^2 term to ensure that high scored patterns corresponded to regressions with a good degree of fitness. Since we are analyzing consolidated sensor data from ILD and WAZE event data, a lower weight was given to the support terms so that patterns from the sparser event data are not penalized in relation to the more dense stationary ILD data. The fixed score parameter values were: $\alpha_1 = 0.3$; $\alpha_2 = 0.5$; $\alpha_3 = 0.2$.

Emerging patterns of road traffic

Table 1 presents the best scored congestion and decongestion traffic patterns on the consolidated ILD-WAZE data. The results gathered from our solution capture a wide variety of simple and emerging patterns on different traffic variables, spanning different road segments at different periods of the day. By having a quick overview of the table results, we can remark some interesting aspects: (i) the discrepancy between the support of ILD and WAZE patterns did not prevent the discovery of varied patterns, showing the importance of a proper score parameterizations able to handle arbitrarily-high data sparsity levels; (ii) the r^2 of emerging patterns from the stationary ILD sensors is generally lower than those produced from mobile

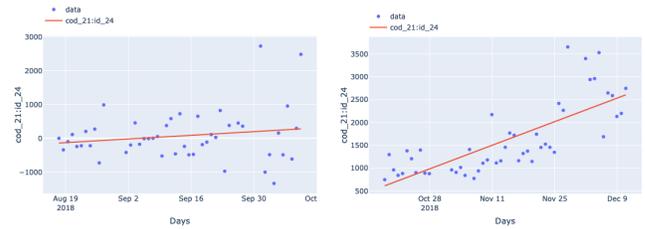
⁴<http://www.lnec.pt/en/departamental-units/transportation-department/>

⁵<https://www.lisboa.pt/cidade/urbanism>



(a) Emerging congestion pattern (growing jam extension) with score: 0.25 = $(\alpha_1 * 0.24 + \alpha_2 * 0.12 + \alpha_3 * (13/15))^2$

(b) Simple congestion pattern (decreasing speed limit) with score: 0.28 = $(\alpha_1 * -0.94 + \alpha_2 * 0.61 + \alpha_3 * (9/31))^1$



(c) Emerging traffic throughput trend with score: 0.21 = $(\alpha_1 * 0.06 + \alpha_2 * 0.03 + \alpha_3 * (47/48))^2$

(d) Simple traffic throughput trend with score: 0.21 = $(\alpha_1 * 0.31 + \alpha_2 * 0.58 + \alpha_3 * (48/48))^1$

Figure 4: Illustrative set of patterns found by E2PAT in consolidated ILD-WAZE data.

sensors. A minimum threshold on the r^2 term can thus be placed to guarantee the absence of false positive discoveries; (iii) among the highest and lowest scored patterns, we find an emerging sharing the same street (R. Castilho) and time (13:00) due to an increase of jam extensions but, simultaneously, a higher throughput of cars.

To guide the interpretation of the found patterns and the associated scores, Figure 4 visually depicts four of the found patterns (two emerging and two simple patterns from each data source). The red line corresponds to the resulting regression and the blue dots are the data points after the undertaken transformation procedures. Fig.4a presents an emerging pattern associated with an increased traffic queue. The pattern has a considerably good score, which is visually justified by the accentuated slope and moderate fitness to the data points. Fig.4b depicts a speed congestion pattern, which for this variable has negative slope (decreasing limit). Fig.4c shows an ILD emerging pattern with a relatively low r^2 , but with a small moderately positive tendency from many available data points. Finally, Fig.4d is one of the top simple ILD patterns, showing an accentuated slope and a good fitness term.

E2PAT visualization tool

A visualization tool was developed to support the analysis and guide the navigation throughout the outputted pattern solutions. The E2PAT tool is integrated within a decision support system that is currently being deployed in the Lisbon city Council to support urban mobility reforms.

An overview of the tool is shown in Figures 5 and 6. The tool provides a user friendly interface for querying the desirable sources

Time	Location	Attribute	Pattern Type	Score	R^2	Slope	Support
13:00	R. Castilho	Spatial Extension	Emerging	0.51	0.81	0.50	5
22:00	Av. da Liberdade	Spatial Extension	Simple	0.39	0.75	1.00	6
08:00	R. Castilho	Spatial Extension	Emerging	0.36	0.68	0.31	5
13:00	R. Braamcamp	Spatial Extension	Simple	0.33	0.98	0.20	6
22:00	Túnel do Marquês	Speed	Simple	0.28	0.61	-0.94	9
03:00	Av. da Liberdade	Spatial Extension	Emerging	0.27	0.55	0.26	6
10:00	R. Sousa Martins	Delay	Emerging	0.24	0.60	0.10	5
18:00	R. Sol ao Rato	Spatial Extension	Simple	0.24	0.74	0.61	5
16:00	R. Joaquim António de Aguiar	Delay	Emerging	0.24	0.54	0.20	5
20:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	0.22	0.01	0.07	48
19:00	ILD cod ₂₁ :id ₁₂	Car Frequency	Emerging	0.22	0.02	0.05	48
19:00	ILD cod ₃ :id ₂₃	Car Frequency	Emerging	0.22	0.02	0.05	48
12:00	ILD cod ₃ :id ₁₄	Car Frequency	Emerging	0.22	0.03	0.06	47
12:00	ILD cod ₂₁ :id ₂₄	Car Frequency	Emerging	0.22	0.03	0.06	47
13:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	0.21	0.00	0.06	47
19:00	ILD cod ₂₁ :id ₂₄	Car Frequency	Emerging	0.21	0.01	0.04	48
19:00	ILD cod ₃ :id ₁₄	Car Frequency	Emerging	0.21	0.01	0.04	48
16:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	0.21	0.00	0.05	47
22:00	ILD cod ₂₁ :id ₂₄	Car Frequency	Simple	0.21	0.58	0.31	48
22:00	ILD cod ₃ :id ₁₄	Car Frequency	Simple	0.21	0.58	0.31	48
18:00	R. Sol ao Rato	Delay	Emerging	-0.21	0.39	-0.34	5
07:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	-0.22	0.00	-0.07	47
12:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	-0.22	0.00	-0.07	47
08:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	-0.22	0.00	-0.08	47
22:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	-0.22	0.00	-0.08	48
18:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	-0.22	0.00	-0.08	48
11:00	R. Silva Carvalho	Speed	Emerging	-0.23	0.24	0.56	5
10:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	-0.23	0.00	-0.11	47
11:00	R. Viriato	Spatial Extension	Simple	-0.24	0.89	-0.17	6
09:00	ILD cod ₃ :id ₉	Car Frequency	Emerging	-0.24	0.00	-0.15	47
04:00	Av. da Liberdade	Speed	Simple	-0.27	0.78	0.57	6
23:00	Av. da Liberdade	Spatial Extension	Emerging	-0.31	0.15	-0.90	5
03:00	Av. da Liberdade	Speed	Simple	-0.34	0.76	0.87	6
08:00	R. das Amoreiras	Speed	Emerging	-0.34	0.14	1.00	5
13:00	R. Castilho	Speed	Simple	-0.35	0.72	1.00	5
08:00	R. Castilho	Speed	Simple	-0.36	0.83	0.75	5
18:00	R. Sol ao Rato	Speed	Simple	-0.39	0.77	0.99	5
22:00	Av. da Liberdade	Speed	Simple	-0.45	0.81	1.00	6
08:00	R. Castilho	Speed	Emerging	-0.49	0.76	0.56	5
13:00	R. Castilho	Speed	Emerging	-0.65	0.82	0.94	5

Table 1: Top 20 congestion and decongestion traffic patterns in the studied area under a weekday calendar.

of spatiotemporal data by selecting the desirable types of sensors spread across the Lisbon’s city and variables of interest. The data can be queried by date, under different calendrical and time granularity constraints, as well as filtered spatially using the geometric selection tool in the map.

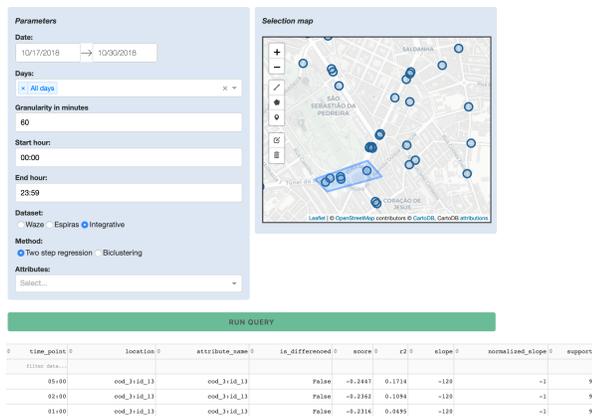
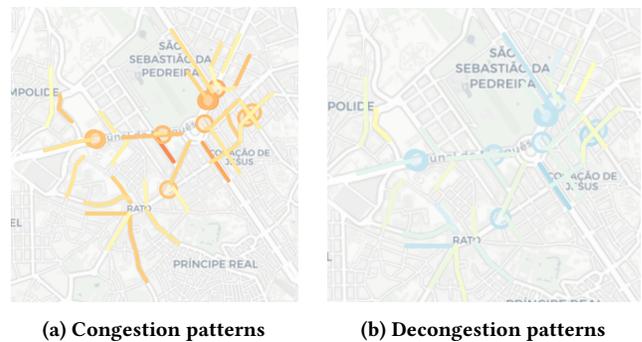


Figure 5: Overview of the user dashboard for querying the road traffic data sources.

The pattern solutions are presented using both interactive tables and interactive maps. The listed emerging patterns can be sorted by spatial and temporal criteria in order to aggregate potentially correlated patterns; as well as by the final score or by each one of the constituent terms (growth, fitness and support). The E2PAT further



(a) Congestion patterns

(b) Decongestion patterns

Figure 6: Map visualization of the found patterns from both ILLD-WAZE data sources using score-based coloring of point-based and trajectory-based emerging patterns.

supports different importation-exportation facilities, allowing the queried data and pattern solutions to be exported into a CSV format to perform further analyzes.

The outputted maps offer a wide-range of possibilities for users to comprehensively explore emerging patterns in accordance with their relevance and spatiotemporal properties. The map has a time selector which allows the user to select specific time points or aggregate results produced over a time range. The visualizations can also vary in accordance with the selected variables of higher interest. The count, which represents the number of times a trajectory was congested between a certain period, is summed, and for the speed and delay the mean is calculated.

6 CONCLUDING REMARKS

This work proposed the E2PAT method to discover emerging patterns from heterogeneous sensor networks in linear time. E2PAT combines spatiotemporal data mappings with simple yet effective time series differencing operations to find emerging behaviors. Differencing orders are explored to further find regular trends and abruptly changing behaviors. E2PAT further provides statistical guarantees of pattern growth, support and accuracy, as well as visualization and navigation facilities, to safeguard the soundness and usability of the pattern analysis process.

An integrative score is also proposed to measure the relevance of emerging patterns, offering a sound criterion to control the false positive and negative discovery rates. Remarkably, we show that the proposed score yields statistical properties of interest: bounded, easily interpretable, and passes normality tests for the found pattern solutions.

To assess the significance of the proposed contributions, E2PAT was applied over the Lisbon's road traffic monitoring system. Results from geolocalized speed and loop counter data confirm the ability to fully retrieve all the emerging congestions, spanning diverse city regions and time periods of the day in accordance with the inputted spatial criteria and calendrical constraints.

The found emerging patterns of urban mobility explore the multivariate nature of the gathered data, covering different jam-related views, such as speed limits, vehicle passage frequencies, and the spatial extent of congested road segments. Results further evidence the ability to unveil actionable, interpretable of road mobility, thus providing a trustworthy context with enough feedback to support mobility reforms.

Future work. As future work, we expect to: 1) extend the conducted analysis towards utility supply and communication networks; 2) extend E2PAT to combine sources of situational context (weather, public events, road interdictions); 3) find new pattern abstractions from emerging behaviors that are spatially and temporally related; and 4) further explore pattern navigation facilities.

ACKNOWLEDGMENTS

The authors thank *Câmara Municipal de Lisboa* for data provision, support and valuable feedback, particularly *Gabinete de Mobilidade* and *Centro de Operações Integrado*. This work is further supported by national funds through *Fundação para a Ciência e Tecnologia* under project ILU (DSAIPA/DS/0111/2018) and INESC-ID pluriannual (UIDB/50021/2020).

REFERENCES

- [1] Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha. 2017. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing* 262 (2017), 134–147.
- [2] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–41.
- [3] Michel Bierlaire, Jingmin Chen, and Jeffrey Newman. 2013. A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies* 26 (2013), 78–98. <https://doi.org/10.1016/j.trc.2012.08.001>
- [4] Peter J Brockwell, Richard A Davis, and Stephen E Fienberg. 1991. *Time series: theory and methods: theory and methods*. Springer Science & Business Media.
- [5] Bi Yu Chen, Hui Yuan, Qingquan Li, William H.K. Lam, Shih-Lung Shaw, and Ke Yan. 2014. Map-matching algorithm for large-scale low-frequency floating car data. *International Journal of Geographical Information Science* 28, 1 (2014), 22–38. <https://doi.org/10.1080/13658816.2013.816427> arXiv:<https://doi.org/10.1080/13658816.2013.816427>
- [6] Feng Chen, Mingyu Shen, and Yongning Tang. 2011. Local Path Searching Based Map Matching Algorithm for Floating Car Data. *Procedia Environmental Sciences* 10 (2011), 576–582. <https://doi.org/10.1016/j.proenv.2011.09.093> 2011 3rd International Conference on Environmental Science and Information Application Technology ESIAAT 2011.
- [7] Z. Chen, Y. Yang, L. Huang, E. Wang, and D. Li. 2018. Discovering Urban Traffic Congestion Propagation Patterns With Taxi Trajectory Data. *IEEE Access* 6 (2018), 69481–69491. <https://doi.org/10.1109/ACCESS.2018.2881039>
- [8] Guozhu Dong and Jinyan Li. 1999. Efficient mining of emerging patterns: Discovering trends and differences. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. 43–52.
- [9] Hongjian Fan and Kotagiri Ramamohanarao. 2003. A Bayesian Approach to Use Emerging Patterns for Classification. In *ADC*.
- [10] Hongjian Fan and Kotagiri Ramamohanarao. 2003. Efficiently mining interesting emerging patterns. In *International Conference on Web-Age Information Management*. Springer, 189–201.
- [11] Philip Hans Franses, Dick Van Dijk, et al. 2000. *Non-linear time series models in empirical finance*. Cambridge university press.
- [12] AM Garcia-Vico, Cristóbal J Carmona, D Martín, Milton García-Borroto, and María José del Jesus. 2018. An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, 1 (2018), e1231.
- [13] Angel Miguel Garcia-Vico, Pedro González, Cristóbal José Carmona, and María José del Jesus. 2019. A big data approach for extracting fuzzy emerging patterns. *Cognitive Computation* 11, 3 (2019), 400–417.
- [14] Angel Miguel Garcia-Vicoa, Francisco Chartea, Pedro González, David Elizondob, and Cristóbal José Carmonaa. 2020. E2PAMEA: A fast evolutionary algorithm for extracting fuzzy emerging patterns in big data environments. (2020).
- [15] Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. 2007. Trajectory pattern mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 330–339. <https://doi.org/10.1145/1281192.1281230>
- [16] Suleyman Gokcan. 2000. Forecasting volatility of emerging stock markets: linear versus non-linear GARCH models. *Journal of forecasting* 19, 6 (2000), 499–504.
- [17] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *nature* 453, 7196 (2008), 779–782.
- [18] Diansheng Guo, Xi Zhu, Hai Jin, Peng Gao, and Clio Andris. 2012. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Transactions in GIS* 16, 3 (2012), 411–429. <https://doi.org/10.1111/j.1467-9671.2012.01344.x> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9671.2012.01344.x>
- [19] Samiul Hasan, Christian Schneider, Satish Ukkusuri, and Marta C. Gonzalez. 2012. Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics* 151 (04 2012), 1–15. <https://doi.org/10.1007/s10955-012-0645-0>
- [20] Rui Henriques and Cláudia Antunes. 2014. Learning predictive models from integrated healthcare data: extending pattern-based and generative models to capture temporal and cross-attribute dependencies. In *2014 47th Hawaii International Conference on System Sciences*. IEEE, 2562–2569.
- [21] Rui Henriques, Cláudia Antunes, and Sara C Madeira. 2015. Generative modeling of repositories of health records for predictive tasks. *Data Mining and Knowledge Discovery* 29, 4 (2015), 999–1032.
- [22] Rui Henriques and Sara C Madeira. 2018. Triclustering algorithms for three-dimensional data analysis: a comprehensive survey. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–43.
- [23] Ryo Inoue, Akihisa Miyashita, and Masatoshi Sugita. 2016. Mining spatiotemporal patterns of congested traffic in urban areas from traffic sensor data. 731–736. <https://doi.org/10.1109/ITSC.2016.7795635>
- [24] Ralph J Koerner. 1976. Inductive loop vehicle detector. US Patent 3,989,932.
- [25] Gang Li, Rob Law, Huy Quan Vu, Jia Rong, and Xinyuan Roy Zhao. 2015. Identifying emerging hotel preferences using emerging pattern mining technique.

- Tourism management* 46 (2015), 311–321.
- [26] Jinyan Li, Guozhu Dong, and Kotagiri Ramamohanarao. 2001. Making use of the most expressive jumping emerging patterns for classification. *Knowledge and Information systems* 3, 2 (2001), 131–145.
- [27] Jinyan Li and Limsoon Wong. 2001. Emerging patterns and gene expression data. *Genome Informatics* 12 (2001), 3–13.
- [28] Jinyan Li and Limsoon Wong. 2002. Identifying good diagnostic gene groups from gene expression profiles using the concept of emerging patterns. *Bioinformatics* 18, 5 (2002), 725–734.
- [29] Li Li, Rui Jiang, Zhengbing He, Xiquan Michael Chen, and Xuesong Zhou. 2020. Trajectory data-based traffic flow studies: A revisit. *Transportation Research Part C: Emerging Technologies* 114 (2020), 225–240.
- [30] Zhenhui Li. 2014. Spatiotemporal pattern mining: algorithms and applications. In *Frequent Pattern Mining*. Springer, 283–306.
- [31] Yuan Liao, Jorge Gil, Rafael HM Pereira, Sonia Yeh, and Vilhelm Verendel. 2020. Disparities in travel times between car and transit: Spatiotemporal patterns in cities. *Scientific reports* 10, 1 (2020), 1–12.
- [32] Huiqing Liu, Jinyan Li, and Limsoon Wong. 2002. A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. *Genome informatics* 13 (2002), 51–60.
- [33] Emilian Necula. 2015. Analyzing traffic patterns on street segments based on GPS data using R. *Transportation Research Procedia* 10 (2015), 276–285.
- [34] Petra Kralj Novak, Nada Lavrač, and Geoffrey I Webb. 2009. Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research* 10, 2 (2009).
- [35] Felix Rempe, Gerhard Huber, and Klaus Bogenberger. 2016. Spatio-Temporal Congestion Patterns in Urban Traffic Networks. *Transportation Research Procedia* 15 (2016), 513 – 524. International Symposium on Enhancing Highway Performance (ISEHP), June 14–16, 2016, Berlin.
- [36] F. Rodrigues, S. S. Borysov, B. Ribeiro, and F. C. Pereira. 2017. A Bayesian Additive Model for Understanding Public Transport Usage in Special Events. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 11 (Nov 2017), 2113–2126. <https://doi.org/10.1109/TPAMI.2016.2635136>
- [37] Athanasios Salamanis, Giorgos Margaritis, Dionysios D Kehagias, Georgios Matzoulas, and Dimitrios Tzovaras. 2017. Identifying patterns under both normal and abnormal traffic conditions for short-term traffic prediction. *Transportation research procedia* 22 (2017), 665–674.
- [38] Hee Seok Song, Jae kyeong Kim, and Soung Hie Kim. 2001. Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications* 21, 3 (2001), 157–168.
- [39] Jinchao Song, Chunli Zhao, Shaopeng Zhong, Thomas Alexander Sick Nielsen, and Alexander V Prishchepov. 2019. Mapping spatio-temporal patterns and detecting the factors of traffic congestion with multi-source data fusion and mining techniques. *Computers, Environment and Urban Systems* 77 (2019), 101364.
- [40] Arnaud Soulet, Bruno Crémilleux, and François Rioult. 2004. Condensed representation of emerging patterns. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 127–132.
- [41] Angel Miguel Garcia Vico, Cristobal Carmona, Pedro Gonzalez, Huseyin Seker, and Maria Jose Del Jesus. 2020. FEPDS: A Proposal for the Extraction of Fuzzy Emerging Patterns in Data Streams. *IEEE Transactions on Fuzzy Systems* (2020).
- [42] Thi Hong Nhan Vu, Jun Wook Lee, and Keun Ho Ryu. 2008. Spatiotemporal Pattern Mining Technique for Location-Based Service System. *ETRI journal* 30, 3 (2008), 421–431.
- [43] William WS Wei. 2006. Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.
- [44] Ruey-Chyi Wu, Ruey-Shun Chen, and C-CCJY Chen. 2005. Data mining application in customer relationship management of credit card business. In *29th Annual International Computer Software and Applications Conference (COMPSAC'05)*, Vol. 2. IEEE, 39–40.
- [45] Jie Yang, Xinyu Zhang, Yuanyuan Qiao, Zubair Fadlullah, and Nei Kato. 2015. Global and individual mobility pattern discovery based on hotspots. In *2015 IEEE International Conference on Communications (ICC)*. IEEE, 5577–5582.