Meter Level Electrical Load Anomaly Detection using Contextual Matrix Profile

Roberto Chiosaa, Marco Savino Piscitellia, Alfonso Capozzoli a[[1]](#footnote-1)\*, Cheng Fanb

a Department of Energy “Galileo Ferraris”, TEBE Research Group, BAEDA Lab, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

b affiliation

Abstract

The rapidly growing electrification of buildings energy systems and appliances lead to an increasing electricity demand. On a global scale, direct and indirect CO2 emissions

*Keywords:* Anomaly Detection, Matrix Profile, Energy Information System

1. Introduction

The rapidly growing electrification of buildings energy systems and appliances lead to an increasing electricity demand. On a global scale, direct and indirect CO2 emissions from buildings energy use reached its all-time high in 2019 \cite{IEEA2020}. In Europe \cite {EU2019} estimates that buildings sector account for 40% of final energy use and 36% of CO2 emissions. The targets imposed by the European community to reduce greenhouse emissions by at least 55% by 2030 \cite{ ﻿EU2018} highlight the critical role played of buildings. Considering that \cite{ ﻿Ramesh2010 } estimates that almost 90% of the total energy consumed during the life cycle of a building depends on the building operation, reducing energy consumption, increasing appliances efficiency and prevent energy wastes through an effective energy management is the key to meet climate change goals.

In the last few years, the increasingly widespread use of IoT sensors for building energy monitoring led to an unprecedented acquisition of reliable and accessible real-time data, supporting the transition from the so-called “smart buildings” into more complex energy ecosystems called “cognitive buildings” \cite{﻿Rinaldi2020}. Although a great deal of research has been done, the increasing volume of collected building energy data still overwhelms end-users, making it hard to spot energy reduction opportunities, find the root cause of energy anomalies or simply be aware of energy usage in buildings and systems. \cite{ ﻿Erhan2021 } estimates that building sector accounts for. reaching zettabyte of data making buildings not only and energy intensive but information intensive \cite{﻿Fan2021}. Building data are heterogeneous and reflects the complex interaction that occurs between occupants, energy systems, the building envelope, and external conditions. Managing those data is not trivial \cite{ ﻿Molina-Solana2017 }, however if properly managed ingested and analyzed, provide the opportunity to gain insight on the building operational behavior discovering opportunities for savings.

A robust coupling of IoT sensors data, Machine learning approaches and energy domain has been proved to be effective in terms of energy savings opportunities to variety of tasks \cite{Molina-Solana2017}: pattern recognition, analysis for

performing essential tasks in building energy management such as energy consumption forecasting, anomaly detection and diagnosis, advanced benchmarking, load profiling, and schedule optimization of building energy systems.

In this paper, we focus anomaly detection of electrical loads in buildings, which is a key application to aid decision makers to reduce wastes and promote sustainable behavior of end users.

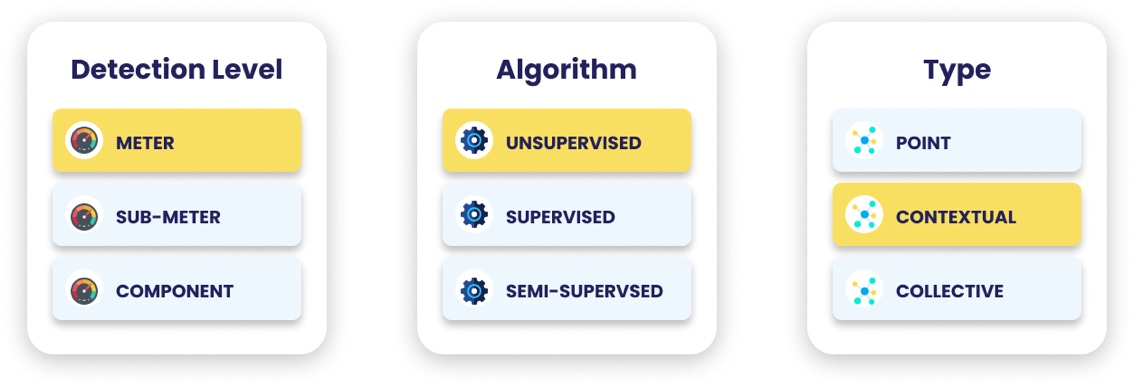
* 1. Anomaly detection: related work

Anomaly detection can be employed in detecting abnormal behavior of end users, detection of faulty appliance or energy subsystem and spotting technical and non-technical energy loss \cite{ ﻿Himeur2020}.

An anomaly is a region of data with significantly different behavior from other data and that do not conform to expected values \cite{}. It can be referred as discord, outlier or exception and its definition is significantly different depending on the field of application.

In fact, this classical definition does not take in consideration other form of anomalies that could exist in buildings energy consumptions such as, abnormal occupation patterns, wrong occupants behavior, incorrect functioning of energy systems, abnormal sub loads consumption and so on \cite{ ﻿Himeur2020 }. Therefore, when performing anomaly detection in this field is of paramount importance to take into consideration other information sources related to the internal and external environmental conditions, level of detection, occupancy patterns, and join the domain knowledge.

Many categorization have been proposed in literature \cite{ ﻿Xu2019} and some are specific for building environments \cite{ Himeur2020, ﻿Erhan2021}. The scope of this paper is not to go deep into categorization; thus, we adopted an anomaly classification based on type, level and algorithm as reported in figure.



**Fig. 1.** Classification of anomaly detection method depending on: (a) detection level (b) algorithm (c) anomaly type.

Classification based on type implies a comparison between the observation and the rest of the data. A *point anomaly* is one individual instance or observation that can be considered anomalous when compared to the remaining data. On the other side, a *Collective anomaly* is an instance does not represent an anomaly per se, but only if considered within the collection of all the other events instances. Finally, *context anomalies* are anomalies only if considered in a certain context (i.e., boundary conditions) and may not be considered an anomaly if it happens in a different context.

Depending on the detail of electrical load monitored the anomaly detection can be performed at different levels. The *meter level* detection analyses the whole building electrical load, without having any information on the disaggregation of that load among the different sub loads or appliances. *Sub-meter level* detection analyses the disaggregated total electrical load and is usually referred to a specific energy system. Finally, *component level* detection consists in identifying anomalies referring to a given appliance/sensor.

The third is an algorithmic centric classification is based on data-driven anomaly detection techniques. *Supervised* anomaly detection requires to train a machine learning algorithm using labelled dataset (i.e., ground truth) to classify anomalous consumption or not. Although supervised anomaly detection can achieve high-accuracy identification results as demonstrated in academic frameworks, its adoption in real-world is still limited compared to unsupervised methods, due to the absence of power consumption annotated datasets \cite{Himeur2020}. Examples of supervised algorithms are deep learning, ANN, Regression, Probabilistic models, Traditional classification. On the other side, *unsupervised* anomaly detection consists in detecting rare and unknown anomalous energy patterns without any a priori knowledge. It usually consists in modelling the normal behaviour and then identify patterns that deviates, under the assumption that the number of anomalies is low compared to the observations. Examples of unsupervised algorithms are Clustering, \cite{﻿Voltage2016} performs anomaly detection on smart grid though the use of clustering. Finally, there are some semi-supervised algorithms that ..

* 1. - Anomaly detection using Matrix Profile

One of the most promising technique for unsupervised anomaly detection in time series is Matrix Profile (MP). Introduced by \cite{﻿Yeh2017d} it is a novel algorithm that performs *all-similarity-join-search* among two time series, i.e. it finds the nearest neighbor for each object of a data collection. Trivial implementation result in excessive computation al time even for modest datasets. Common variants of this problem involve the search ok k-nearest neighbor by setting a threshold, which is critical as well as difficult to set \cite{ ﻿Yeh2018 }. Others perform siliarity search by reducing the dimensionality of dataset through PAA in order to speed up computation, however, this method causes loss of valuable information.

MP proposes a ultra-fast similarity search under the z-Euclidean distance that does not reduce dimensionality, but calculates the full join, eliminating the need of setting a threshold making the method almost parameter free and exact. The exact and scalable algorithm allows the method to be incrementally maintainable, deterministic in time and so parallelizable on multicore processor to speed up even further the computations.

Given two time series and a given subsequence length, the MP algorithm produces two new series: the MP and Matrix Profile Index (MPI) index. MP is a one-dimensional timeseries that stores the z-normalized Euclidean distance between each subsequence of the first series and the closest matching subsequence (i.e. nearest neighbor) of the second time series. MPI is a one-dimensional timeseries that contains the index of where the nearest neighbor is located in the second timeseries.

By joining information of MP and MPImany insights could be extracted. In particular, finding the minimum value of the Matrix Profile is possible to find the best matching subsequence in a series (i.e., motif discovery) on the other side by finding the maximum value of the MP it is possible to find the subsequence with the largest distance to its nearest match, (i.e., discord discovery). In this sense discord discovery may be interpreted as an anomaly detection method that discovers the most unique subsequences in a dataset. Discord discovery using MP as anomaly detection method has been employed with success in different fields.

In medical field \cite{﻿Alshaer2020} proposes an unsupervised real time anomaly detection method based on continuous learning of time series shaplets extracted though Matrix Profile algorithm. Those shaplets are extracted and stored in an anomaly library and then used for anomaly detection in an electro-cardiogram (ECG) time series (﻿MIT-BIH database [32]), using a in a sliding window.

An industrial application of anomaly detection is presented in\cite{﻿Anton2020} which combines the classical approach of Matrix profile with the hamming distance to automatically detect intrusions in the network of a water processing facility.

\cite{ ﻿Madrid2019 } applies a generalization of MP algorithm called Pan MP find different length anomalies in ﻿automated pedestrian counting system developed in Taipei.

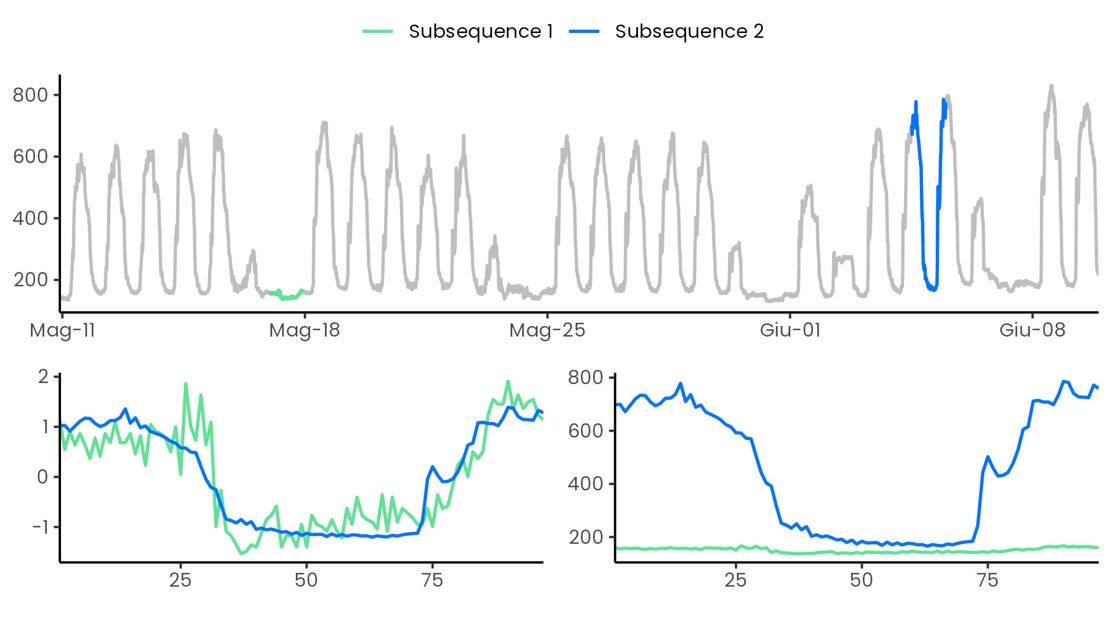
MP is an effective method used to identify anomalies in IT field. \cite{DinalHerath2019} introduces ﻿a real time anomaly detection framework based on matrix profile called RAMP (Real-Time Aggregated Matrix Profile), that is able to identify anomalies in scientific workflows. (building block). \cite{ ﻿DePaepe2020a} Applies a noise elimination technique based on Matrix Profile on real Yahoo! internet traffic metrics to detect anomalous behaviors; \cite{ ﻿DePaepe2019 } demonstrate how the elimination of noise can help in anomaly detection of noisy date by testing the algorithm on Numenta Benchmark \cite{ ﻿Ahmad2017 }.

In the energy field there are few implementations of MP algorithm. \cite{ ﻿Nichiforov2020 } Identifies anomalous patterns though a basic application of Matrix Profile on public building energy traces and then classifies the pattern.\cite{ ﻿Zhu2020} demonstrates how Matrix Profile can be useful in detecting anomalies in different fields in particular in meter swapping and earthquake monitoring. \cite{﻿DePaepe2020b} applied an implementation of the classic Matrix Profile, called Contextual Matrix Profile, in detection of anomalous energy consumption on a ventilation units of three households. \cite{ ﻿Park2020a } applies MP as a part of a automated load profile discord identification ( ALDI ) based on statistic comparison between normal and anomalous patterns in a large portfolio of buildings.

* 1. Implication MP on energy domain

From the previous literature review it emerges that the MP method have been employed successfully in different fields for anomaly detection and the authors have proposed different implementations according to the field of interest. In fact, even if MP is an unsupervised method useful for discord discovery every field have different boundary conditions and different restrictions that cannot be overlooked. In the field of building, the most interesting timeseries are those related to energy consumption. Those timeseries are strictly correlated to many different variables such as occupation, weather conditions, energy systems and so on. A completely unsupervised method may fail to consider the relation with those variables and extract ineffective or trivial results, not useful for anomaly detection.

In buildings, an anomaly is an unexpected behaviour that consists in a strongly different energy consumption. The classic MP, by performing with z-score normalization, searches for each subsequence the nearest neighbour based on shape similarity, however, anomalous shapes not always correspond to anomalous energy consumption, as well as similar shapes in z score not always reflect similar behaviour. With reference to Figure 2 it is possible to appreciate how three subsequences of electrical load timeseries are considered similar in z score while without normalization the amplitudes are very different reflecting very different energy consumption.



**Fig. 2.** Effect of z-score normalization on two subsequence of electrical load timeseries: (a) electrical load timeseries; (b) comparison between z-score normalized subsequences; (c) not normalized subsequences.

The effect of z-normalization not only does not consider the magnitude of the time series but also tends to enhance any fluctuation of the timeseries. By comparing two relatively flat subsequences under z-score normalization the resulting Euclidean distance is higher compared to non-flat subsequences (see Figure), this results into higher values of MP in flat regions of the timeseries. This issue have been largely analysed in \cite{ ﻿DePaepe2020a } where a smoothing is proposed as possible solution to this issue, beside the trivial solutions of discard flat regions or change the sub sequence length.

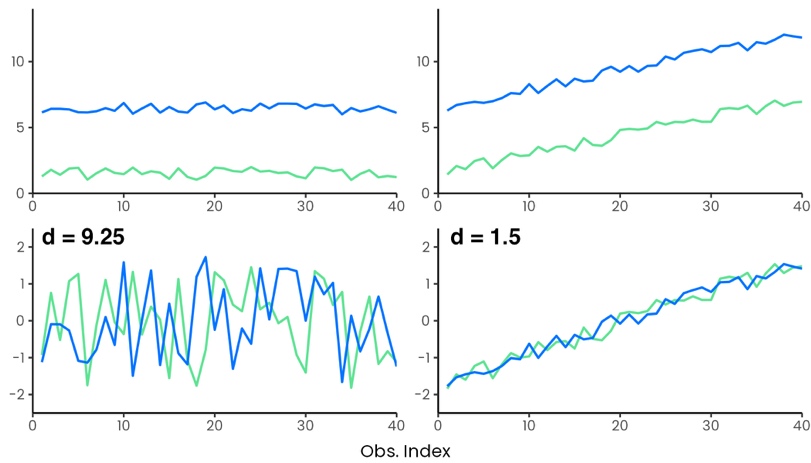


Fig. 3. Effect of z-score normalization of relatively flat subsequences on Matrix Profile values.

Moreover the energy consumption pattern changes between weekdays and weekends/holiday, so it would be unfair comparison to compare subsequences pertaining to these groups, the same is to compare subsequences of night hours and daily hours. Introducing domain knowledge to find discords only in sone subgroups of the timeseries became important. \cite{Dau2017} introduces the concept of annotation vector used to introduce domain knowledge in the process of motif and discord discovery, which allows to find results that follows users defined constraint and produce better results, closer to expectations of the analyst. This method has been proved to be effective to solve different issues: simplicity bias, actionability bias. However, this method is a posteriori method that does not modify the way MP is calculated, in particular all-pairs similarity search is always performed and then some regions are excluded form motif/discord search. Sometimes it can be useful to exclude some region or to group subsequences into different groups and then perform the similarity search in order to discover anomalies by comparing only the interesting regions and excluding others. A solution to this problem have been proposed by \cite{DePaepe2020b} where Contextual Matrix Profile (CMP) algorithm permits to define ranges along two timeseries and look for the best matching subsequence among these ranges. This permits different a priori grouping of the time series observations so that MP calculation can provide novel and more interesting insights.

* 1. Contribution of the paper

The prompt and accurate discovery of anomalies in building electrical load is the key to reduce energy wastes and enhance energy management in buildings. To this aim the objective of this work is to detect anomalous electric load at building level in quasi real time through the unsupervised anomaly detection methods based on matrix profile method. According to the previous literature review and excursus on implication of MP as anomaly detection method, this paper intends to address the following issues by contributing as follows:

* Unsupervised selection of the only parameter of the MP algorithm the time window, Couple unsupervised automatic data driven methods, without setting any parameters
* Overcame issue of z-score in energy with the use of Euclidean distance
* Apply a contextual anomaly detection by applying an advanced contextual matrix profile method with the definition of context and groups through an unsupervised fashion
* Propose an almost real time anomaly detection method,
* Robust anomaly score definition based on majority voting
* Define a clever anomaly score that talked into account only positive anomalies (higher power consumption)

The rest of the paper is organized as follows: in Section 2 in section data analytics methods are presented, in Section 3 the methodological framework, in Section 4 results and Section 5 discussions.

1. Description of the Data Analysis Methods
   1. Matrix Profile

With reference to Fig 4 some fundamental concepts and definitions need to be introduced before explaining the method. First of all a time series $T \in R^n$ is a sequence of real-valued numbers $t\_i \in R : T = \{t\_1,t\_2 \dots t\_n \}$ with $1<i\leq n$ where $n$ is the length of $T$

Since we are interested in local properties of time series, we define portion of time series as subsequence $T\_{i,m} \in R^m$ as a continuous subset of values from $T$ of length $m$ starting from position $i$; dormally defined as $t\_i \in R : T\_{i,m} = \{t\_i,t\_{i+1} … t\_{i+m-1} \}$ with $1<i<n-m+1$.

An ordered set of all possible subsequences of $T$ obtained by sliding a window of length $m$ across $T$ is called all-subsequences set $A$ of a time series $T$ and is formally defined as follows: $A = \{T\_{1,m} , T\_{2,m} \dots T\_{n-m+1,m} \}$ where $m$ is a user-defined subsequence length. We use $A[i]$ to denote $T\_{i,m}$

By computing the distance between a given query (subsequence $T\_{i,m}$) and each subsequence in an all-subsequence set $A$ it is possible to define is a vector of distances called distance profile $D\_i$ of a timeseries $T$. Formally, $D\_i = [d\_{i,1}, d\_{i,2},\dots, d\_{i,n-m+1}]$, where $d\_{i,j} = dist(T\_{i,m}, T\_{j,m})$ for all $j \in [1,2,\dots, n-m+1]$ where $i\neq j$ and $dist$ is the distance metric applied. It is possible to adopt different kind of distances to compute the distance profile \cite{Gharghabi2020, ﻿Zhu2020} but the original method uses the Euclidean distance between the z-normalized subsequences.

If the distance profile is calculated between a query in $T\_{i,m}$ and the all subsequence set of T (i.e. self-join), by definition the $i^{th}$ location of the distance profile $D\_i$ is zero since the distance is calculated between the query and itself ($d\_{i,i} = dist(T\_{i,m}, T\_{i,m}) = 0$). Moreover, the distance is close to zero just before and after this location. Those matches are called \emph{trivial matches} and are usually avoided during similarity search by imposing an \emph{exclusion zone} (as function of $m$, usually set to $m/4$) before and after this location.

It is possible to finally define matrix profile $MP$ is a vector that stores the z-normalized Euclidean distance between each subsequence $T\_{i,m}$ and its nearest neighbor. Formally, $MP = [\min(D\_1), \min(D\_2),\dots, \min(D\_{n-m+1})]$, where $D\_i$ is the distance profile corresponding to query $T\_{i,m}$ and time series $T$. In other words, it can be generated by extracting the smallest value in each row/column of the full distance matrix. Of course, the construction of the distance metrics is the most straightforward method but event he less computational efficient, this is the reason why many algorithms has been proposed for the MP calculation, such as mass, such as stamp and stomp that computes the MP in triangular form

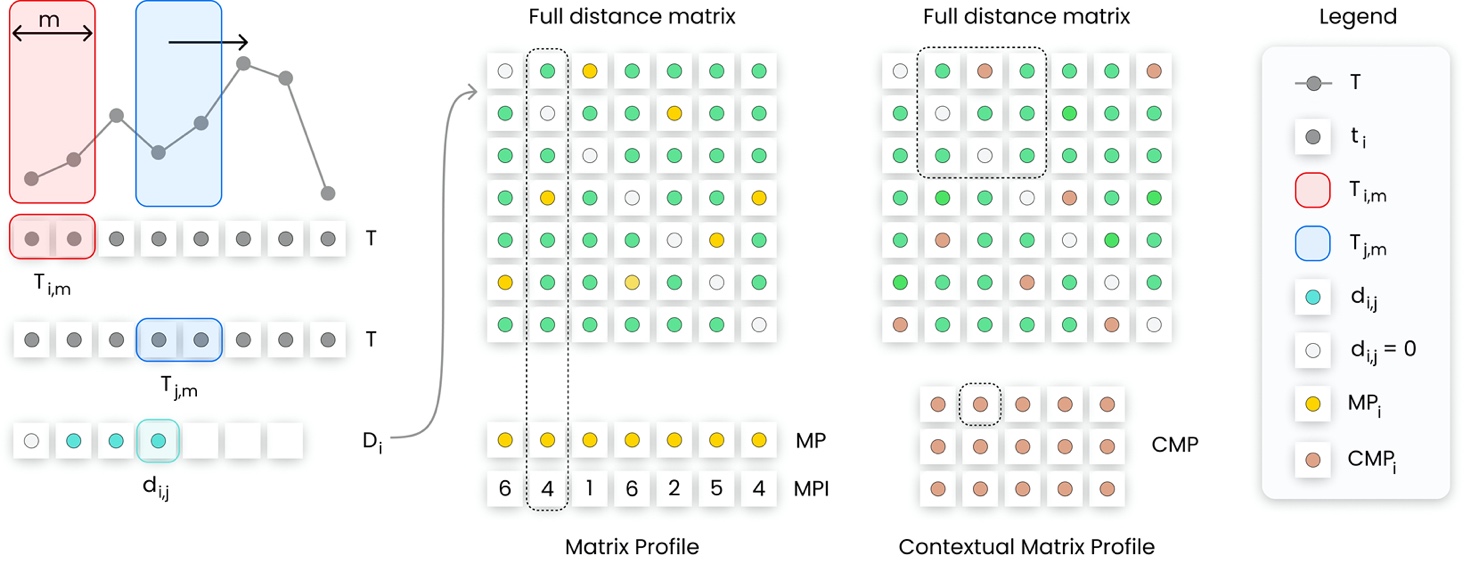


Fig. 1. step of matrix profile calculation in case of self-join. From left to right is explained the calculation of the element d\_i,j of the distance vector D\_i given the query T\_i.m. By calculating the distance vector for the all subsequence set of T, and storing them in a matrix the full distance matrix is composed. MP is the row wise minimum while the comp is the minimum over rectangular regions

The MP is the column wise minimum over the entire full distance matrix, meaning that if finds the best matching T1 subsequence (i.e. minimum distance) for any subsequence in T2. However, as already mentioned in section, comparing regions of timeseries that belongs to different context or operating conditions or different boundary conditions may result into misleading results. This is why the contextual matrix profile is introduced. Contextual matrix profile is defined as the minimum over rectangular regions of the full distance matrix, allowing to find the best matching subsequence in ranges dover T1 and T2 allowing to group data in custom quay comparing only portions of T1 with portions of T2. In this way the resulting matrix CMP will be composed with as many rows as the number of contexts.

* 1. Decision tree

Classification is the task that assigns a class label to unlabelled instances through a classification model which creates a relationship between the predicted or target variable and the predictive variables. Among the numerous classification methods used to describe and explore complex data, the most used are those based on decision trees thanks to their capability to be easily translated in graphical form are commonly used in different fields \cite{﻿Grubinger2014}. Depending on the type of target variable a decision tree can be a classification tree (categorical target variable) or regression tree (numerical target variable) \cite{ ﻿Tan2011 }.

In this paper is employed the recursive partitioning decision tree called classification and regression tree CART. Starting from the root (all the available instances) this method proceeds through a binary decision fashion to split the instances in purer subsets (nodes) in a froward stepwise fashion maximizing at each step the purity of each node, yielding local optimum \cite{ ﻿Grubinger2014} once a stopping condition is satisfied. The purity of each node can be identified through an impurity measure \cite{ ﻿Yan2016a, ﻿Capozzoli2018a, ﻿Tan2011 }.

* 1. Clustering

Clustering is the process of creating groups (i.e., clusters) based on similarity within some attributes. Clustering algorithms can be categorized into partitional or hierarchical. In the first case, the observations are divided into non-overlapping subsets called clusters. The hierarchical clustering generates non-overlapping clusters, and each cluster can be further divided into subclusters and so on, creating a tree structure.

* 1. Anomaly detection

1. Methodological Framework

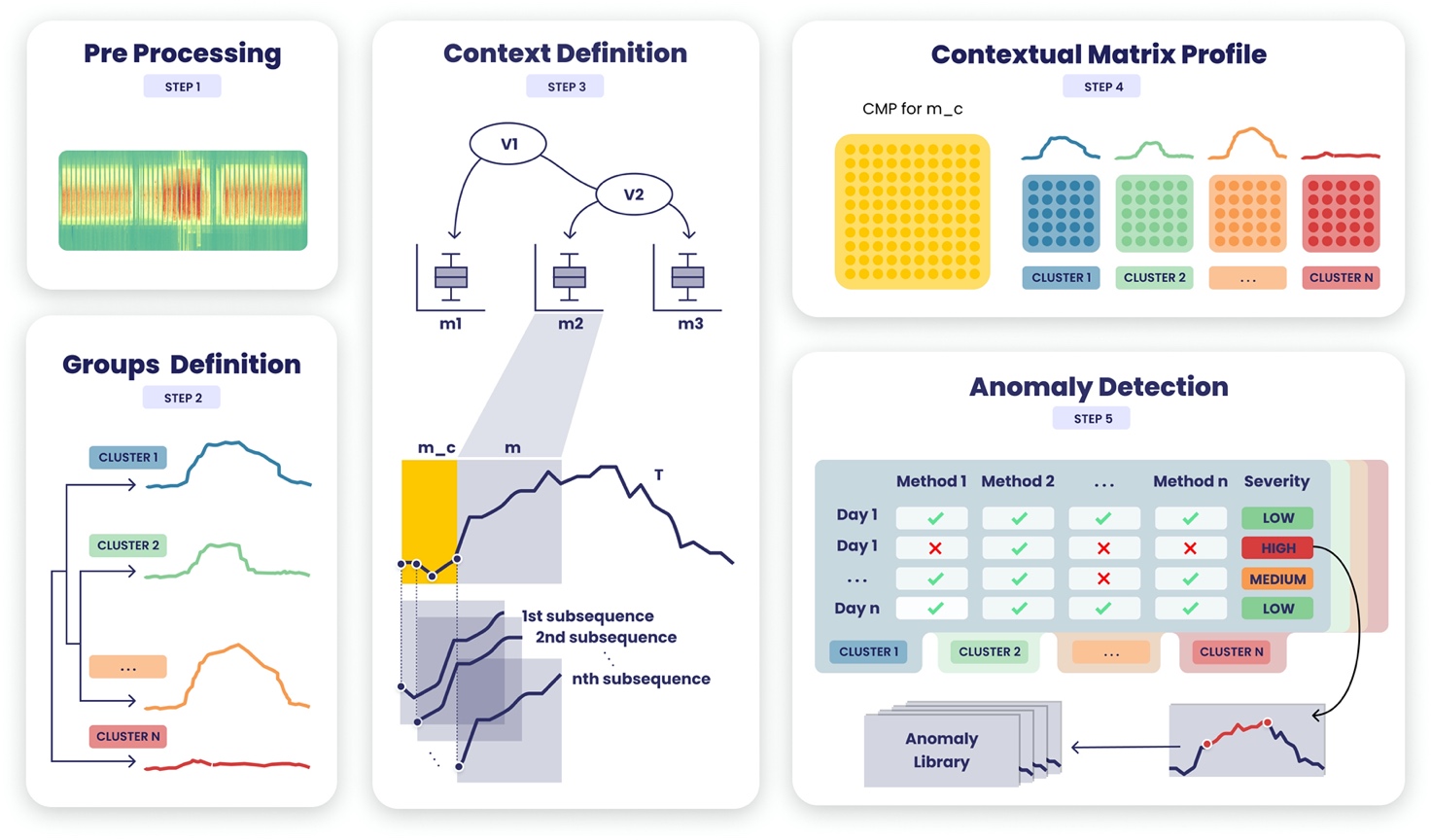


Fig. 1. (a) first picture; (b) second picture.

* 1. Pre Processing
  2. Context definition

In order to proceed with contextual matrix profile analysis two main parameters have to be set. The subsequence time windon length and the length of tha context. Those two parmaeters are usually set a priori in our analysis we use CART to define interesting time window length.

The total electrical power is predicted using hours as predictive variable.

The identified time window represent interesting regions of the power consumption in which the electrical load has a peculiar behaviour. The identification of these region have a twofold meaning, identify interesting region and identifythe length of subsequence.

The context is defined as the time interval in which subsequences may start. Since we are interested in subsequences falling int the interesting region denoted by the cart we decided to set to 1 h m\_c. With reference to figure with the definition of this context we analyze all the subsequences of length m that start inside the m\_c interval and ends in the m interval.

* 1. Groups definition

clusters are defined by means of hierarchical clustering with Euclidean distance of daily electrical load profiles. The resulting clusters are considered representative of some electrical load behavior.

Contextual matrix profile

Contextual matrix profile is constructed for each context generated by the cart.

* 1. Anomaly detection

The anomaly detection is performed in a given context in a give n group, since groups have similar boundary condiions. Anomaly score different methods have been apliied. The severity of the anomaly I calculated through majority voting.

* 1. Anomaly library

Once detected the anomalies are saved into an anomaly library in which context severity and profile are stored.

1. Results

The case study analysed refers to the energy consumption of a MV/LV transformer cabin identified as “substation C”, that serves a part of the main campus of Politecnico di Torino (PoliTo), an Italian university located in Turin. The measurement infrastructure provides the total electrical load with 15 min timestamps. In order to use a dataset that is large enough to capture the behaviour of the electrical load, with regular occupation patterns

* 1. Pre Processing
  2. Context definition

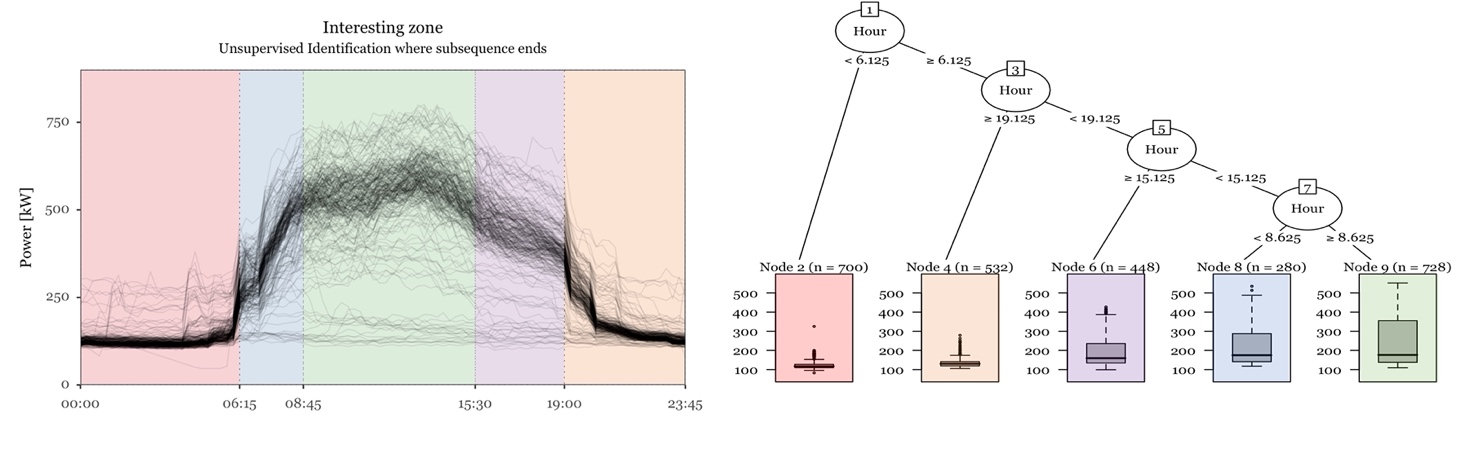


Fig. 1. (a) first picture; (b) second picture.

* 1. Groups definition

In contextual matrix profile groups are defined as homogeneous groups in which boundary conditions are similar. Four groups are defined after a cluster analysis performed on daily electrical load basis.

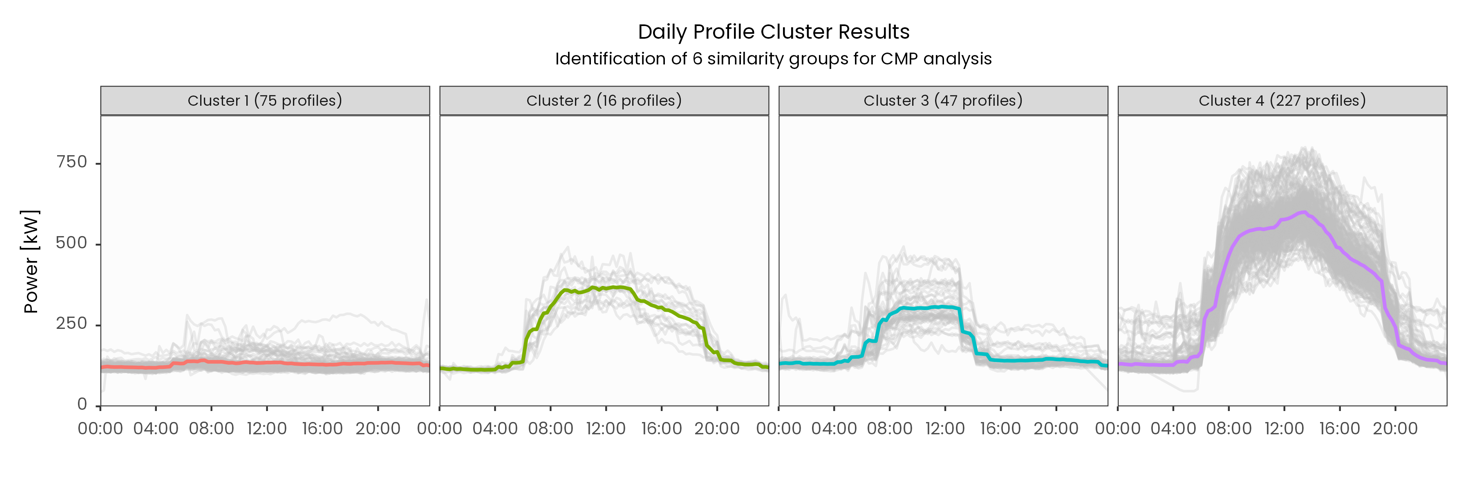


Fig. 1. (a) first picture; (b) second picture.

* 1. Contextual matrix profile

For each context a contextual matrix profile is constructed. Then CMP is divided into groups and anomaly detection performed

Immagine che contiene testo, elettronico

Descrizione generata automaticamente

Fig. 1. (a) first picture; (b) second picture.

* 1. Anomaly detection

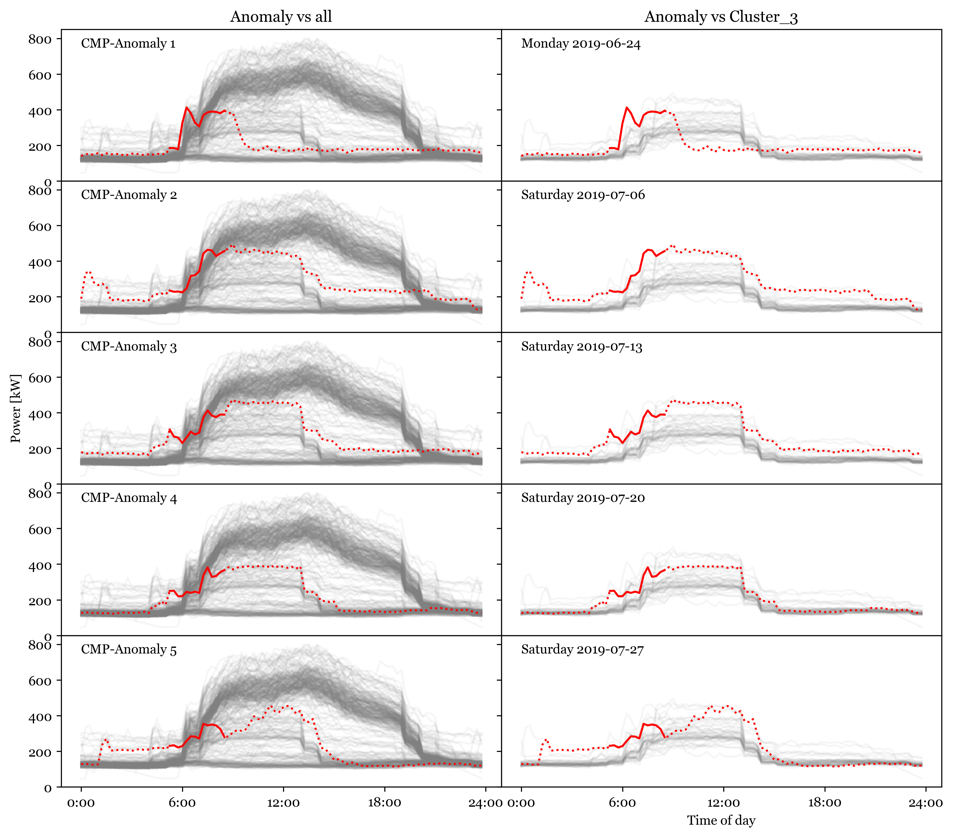


Fig. 1. (a) first picture; (b) second picture.

* 1. Anomaly library

1. Discussion

Twin freak

﻿For a given sub-sequence, Matrix Profile computes the Euclidean distance with respect to all other sub- sequences and identifies the minimum distance. Therefore, a repeated anomaly instance would cause false negatives due to the previous anomaly instance being part of the all sub- sequence set.

Specifically, frequent/rare subsequences are defined as the ones with the smallest/largest 1-nearest neighbour distance, which are also known as motif/discord. However, discord fails

the ones with the smallest/largest 1-nearest neighbour distance, which are also known as motif/discord. However, discord fails to identify rare subsequences when it occurs more than once in the time series, which is widely known as the twin freak problem.

\cite{DinalHerath2019} through a semi-supervised model permits to limits the number of subsequences compared, considering for comparison only references with no anomalies.

\cite{﻿He2020} proposes a method called “Neighbour Profile” based on sampling and density estimation to perform anomaly detection and overcame the issue of twin freak.

1. Conclusion

# Nomenclature

MP Matrix Profile

CMP Contextual Matrix Profile

CART Classification And Regression Tree

References

1. \* Corresponding author

   Email address: alfonso.capozzoli@polito.it (Alfonso Capozzoli)

   URL: www.baeda.polito.it (Alfonso Capozzoli) [↑](#footnote-ref-1)