A review of unsupervised statistical learning and visual analytics techniques for non-residential building performance analysis

Clayton Miller

Architecture and Building Systems (A/S), Institute of Technology in Architecture (ITA), ETH Zürich
John-von-Neumann-Weg 9, Zürich, Switzerland

Abstract

Measured and simulated data sources from the building design phase, control and energy management systems, and advanced metering systems ~~is~~ increasing rapidly. Mechanistic, manual analysis of such data is time consuming using conventional techniques. Thus, a significant body of literature has been generated to analyze these data sets using unsupervised statistical learning techniques designed to quickly uncover structure and information with ~~less~~ input parameters or meta data about the buildings collected. Further, visual analytics techniques are developed as aids in this process for a human analyst to utilize and interpret the results. This paper reviews 98 publications ~~from the last ten years~~ of unsupervised learning techniques as applied to non-residential building performance applications. The categories of analytics techniques covered include clustering, novelty detection, motif and discord detection, rule extraction, and visual analytics. The publications apply these techniques in the domains of smart meters, portfolio analysis, operations and controls optimization, and anomaly detection and they are from the building energy analysis, building simulation, computer science and electrical engineering domains.

Keywords: building performance analysis, data mining, unsupervised learning, visual analytics, clustering, novelty detection, smart meter, portfolio analysis, review

---

1. Introduction

The architectural design and building construction domains are characterized by the complexity and number of stakeholders in involved in the creation of their products. Most non-residential buildings have a relatively unique design financed by a unique owner or developer and are constructed by a group of contractors with slightly varying techniques. These factors contribute to the fact that the energy consuming systems within each building are slightly different in numerous ways. A large increase in the amount of temporal sensor data from these systems is being created and made accessible at rapid rate. Analysis of these data

Email address: `miller.clayton@arch.ethz.ch` (Clayton Miller)
URL: `http://systems.arch.ethz.ch/` (Clayton Miller)

is stifled by the inability to scale techniques across the non-heterogeneous building stock.

Analytics of non-residential buildings is strong in its investigative nature; an analyst's first task is quite often to collect raw temporal data from a building energy management system (BEMS) or building automation system (BAS) and discover opportunities for performance improvement. Conventionally this process then requires a certain amount of in-depth data collection about the physical and systems meta data from the building itself. Non-residential buildings are differentiated in the complexity and control of heating, ventilation, air-conditioning, and lighting systems as compared to residential. They are not as occupant-behavior driven in most cases and, thus require their own set of approaches.

Unsupervised statistical learning techniques are advantageous due to their ability to characterize measured or simulated performance data quickly with less analyst intervention, meta data, and ground truth labeled data. These types of techniques are generally used in conjunction with visual analytics approaches as a first pass in analysis of a large dataset.

1.1. Previous Reviews

Various reviews have been completed in the past that overlap with a part of the scope of this paper. Most of them are designed to focus on a single core domain of research; the main two areas being building operations analysis and smart grid optimization. One of the earliest reviews of artificial intelligence techniques for buildings was completed in 2003 by Krarti [1] and covered both supervised and unsupervised methods. Dounis [2] updated this work and focused on outlining specific techniques in detail. Reddy's seminal book [3] about a large variety of analysis techniques for energy engineers includes chapters on clustering and unsupervised methods specifically. Lee et al. [4] describe~~s~~ a variety of retrofit analysis toolkits which incorporate unsupervised and visual analytics approaches in a practical sense. Ioannidis et al. [5] created a broad ontology of data mining and visual analytics for building performance analysis, however with a strong focus on the techniques and not examples of works using them. From the utility and power grid side, Morias et al. [6] created a general overview of various data mining techniques as focused on power distribution systems. Chicco [7] covered clustering methods specifically focused on load profiling tasks. Zhou et al. [8] covered the concept of customer load classification approaches. Ruparathna et al. reviewed a wide array of building performance technologies, including several analytics techniques [9].

1.2. This Review

This review paper discusses the analytics of the building industry more holistically by combining key publications from a larger set of category applications and by specifically targeting unsupervised learning and visual analytics techniques. Additionally, there is the contribution to aggregate publications from the research domains of building energy analysis, building simulation, computer science and electrical engineering. These domains have often remain isolated within the previously outlined review studies. This paper compiles a set of 98 publications released since 2005 that focus on the use of unsupervised statistical learning techniques on temporal or spatial data from non-residential buildings.

The process was started through a selection of unsupervised analytics categories outlined by authoritative sources from the machine learning community [10, 11, 12, 13]. These categories are clustering, novelty detection, motif and discord detection, and rule extraction. The field of visual analytics was added to these categories in order to cover the presentation layer of many of these types of techniques. An initial search of publications was then selected for inclusion through a Google Scholar search of the combination of the technique categories and the terms “building energy”, “building performance analysis”, and “building energy analysis”. From this initial list of publications, a set of application categories and sub-categories was developed as seen in Figure 1. A more detailed search of each application category was then completed to account for the more unique analytics techniques used in those domains. Only publications with a majority of the focus on utilization of unsupervised techniques and with a focus only on non-residential buildings are reviewed. Only works completed since 2005 are included in order to review only the most contemporary work and due to the relatively recent development of most of the techniques reviewed. A cutoff date of December 31, 2015 is applied for inclusions of publications in this review.

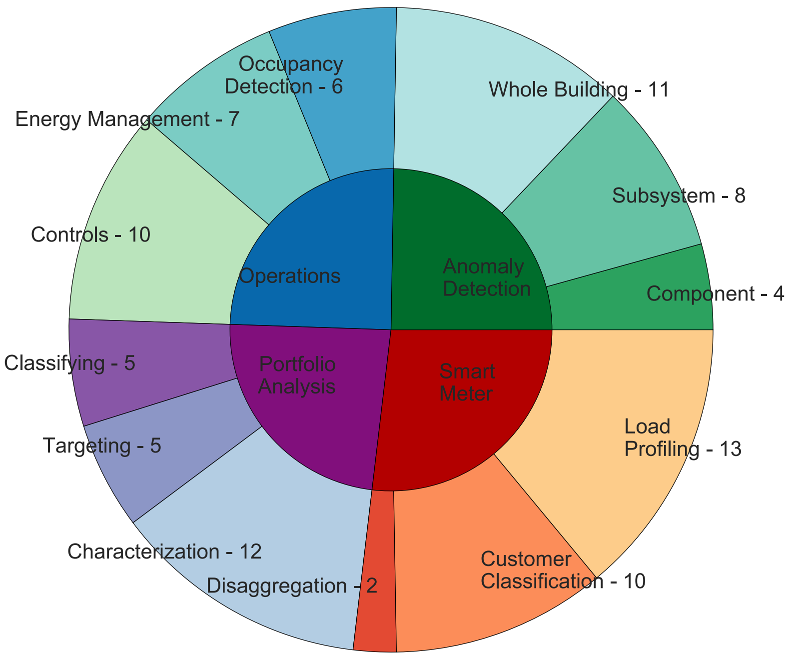

Figure 1: Categories and sub-categories (including number of publications) of building performance analysis applications of unsupervised learning and visual

~~In this paper~~, Section 2 creates an outline of the analytical techniques discussed in these publications and the subsequent sections focus on various categories of applications in the industry. Sections 3-6 evaluate these application categories in-depth. Section 7 discusses an overview of the unsupervised learning techniques used in the publications, the yearly time line of publication, and which research domains and specific journal and conference publications contain them.

2. Overview of Unsupervised Learning and Visual Analytics Techniques

This section gives an overview of the five categories of unsupervised learning and visual analytics techniques discussed in this paper: Clustering, Novelty Detection, Motif and Discord Detection, Rule Extraction, and Visual Analytics. The general nature of the techniques is discussed and the specific methods are briefly listed along with the abbreviations that are used in Tables 1-4. Statistical learning can be divided into two major categories: supervised learning and unsupervised learning. According to the formative text on the subject by James et al. [11] and Hastie et al. [10], supervised statistical learning techniques involve a set of observations $x_i, i = 1, ..., n$ and an associated response variable y_i. The goal of a supervised process is to predict y_i using the features generated from x_i; this objective is easy to verify according to the algorithms accuracy in predicting that response. Unsupervised learning, in contrast, lacks the associated response y_i because it seeks to simply to understand the relationship between the observations in x_i, generally in an exploratory fashion. These references go on to explain that unsupervised techniques are, thus, often more subjective in their application and are considered more challenging in their utilization in practical applications [13, 11, 10]. Despite the challenge, various instances of their application are found in the literature with relation to building performance data as outlined in the subsequent sub-sections.

2.1. Clustering

Clustering is the most common general unsupervised approach applied to building performance data. It is used to automatically generate *subgroups* of similar types of observations. James et al. [11] describe this process as the grouping of n observations into K groups, or clusters, according to a set of generated p features. The two most common types of clustering are K-means and Hierarchical clustering. A wider array of techniques have been developed to optimize the objectives of separating subgroups in more effective and efficient ways. An example of the clustering of daily energy consumption profiles can be seen in Figure 2. This example shows three and half years of hourly energy data for a building in which the diurnal profiles are grouped according to similarity and plotted across the time range of the data set in addition to the typical clustered profiles plotted across a 24 hour period. This application of clusering is known as load profiling and it is covered in Section 3. Specific types of clustering techniques found in publications in this review are K-means, Principal Component Analysis (PCA), Maximum Likehood (ML), Hierarchical, Self-Organizing Maps (SOM), Empirical Mode Decomposition (EMC), Fuzzy clustering, Support Vector Machines (SVM), and Ant Colony clustering.

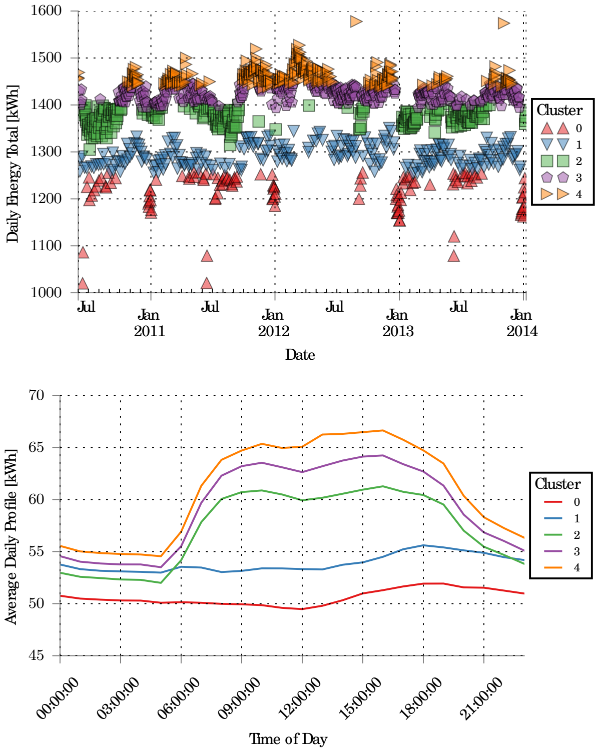

Figure 2: Load profiles clusters from a building electricity consumption data set (adapted from [14])

2.2. Novelty Detection

Novelty detection is a category of techniques that finds behavior or observations within a data set that are generally unique and often do not conform to expected behavior [15]. Concepts such as outlier and anomaly detection fit within this category. An example of an outliers detection process can be seen in Figure 3 in which a set of *quality analysis metrics* were implemented on a single data stream including an outliers detection component that is designated as quality level 2. Specific types of novelty detection techniques in this review include Principal Component Analysis (PCA), Generalized Additive Models (GAM), Classification and Regression Trees (CART), Wavelets, Fourier Transforms, Linear Discriminant Analysis (LDA), and Nearest Neighbor methods.

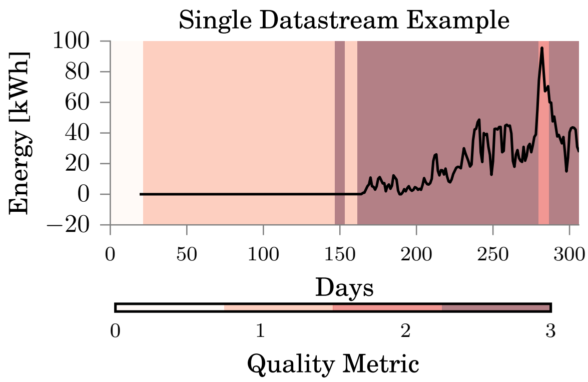

Figure 3: Example of novelty detection through data quality metrics in which a value of 2 signifies and statistical outlier (adapted from [14])

2.3. Motif and Discord Detection

Motif and discord detection and analysis is sub-domain specific to the investigation of time-series datasets [16]. A motif is a subsequence of data that exhibits a pattern that occurs frequently in a data stream. A discord is a subsequence that occurs rarely and is considered anomalous amongst the rest of the dataset. Figure 4 illustrates the concept of motif and discord candidate identification on a two week sample data set. This example is of thirteen days of electrical consumption where the diurnal patterns were extracted using Symbolic Aggregate Approximation (SAX) and divided into motif and discord candidates according to their frequency amongst the sample set. Other approaches found in the literature related to motif and discord detection include Markov Models, Maximum Likelihood, Dirichlet Process Gaussian Mixture Models (DPGMM), and Incremental Summarization and Pattern Characterization (ISPC).

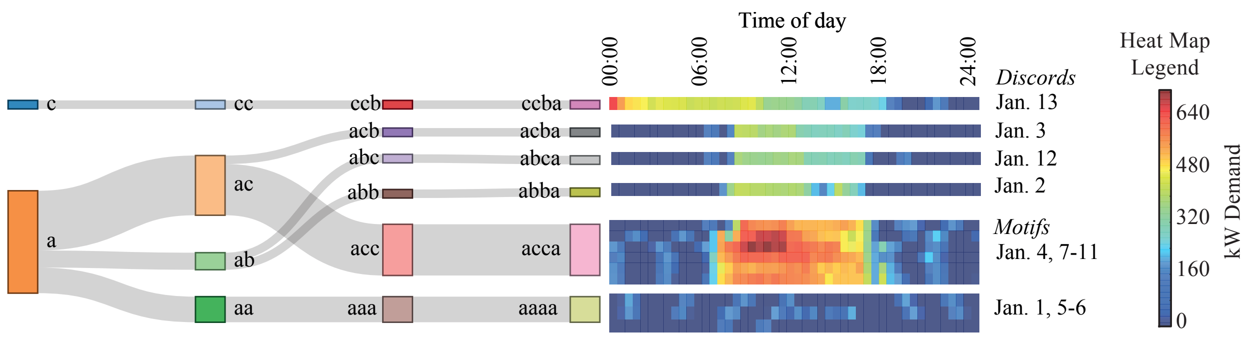

Figure 4: Example of motif and discord candidate extraction from a two week example data set (adapted from [17])

2.4. Rule Extraction

Rule extraction techniques focus on the ability to automatically find relationships between variables within a data set. Hastie et al. [10] describe the goal of these processes as attempting to find joint values of the variables $X = (X_1, X_2, ..., X_p)$ that appear most frequently in the data base. These rules are often used to support control sequences and fault detection approaches at the component level. The general field of creating Data Association Rules (DAC) is the most common form of this technique. Other methods found in this review include Markov Models and Support Vector Machines (SVM).

2.5. Visual Analytics

The domain of visual analytics is relatively new and rapidly developing. Keim et al. define the field [18] as a medium of a semi-automated analytical process, where humans and machines cooperate using their respective distinct capabilities for the most effective results. This domain focuses on the use of visualization and various human interactions such as the Schniederman mantra of Overview First, Filter and Zoom, and Details-on-Demand [19]. Visual analytics is developing in the building industry through the proliferation of energy management system dashboards. This review divides the visual analytics for building energy into two categories: approaches focused on finding patterns in energy data and those on more general energy management with focuses on performance indicators.

Many of the publications in this review indicate a dominant technique that is determined and included in Sections 3-6 and their associated tables. Some of the publications don't specify a single technique or implement numerous techniques for comparison; these publications are classified under the category of *Multiple Techniques*.

3. Smart Meter Analytics

Advanced Metering Infrastructure (AMI), also known as smart meter systems, is a network of energy meters, most often focused on electrical power measurement of a whole building. These systems are generally implemented and utilized by electrical utility providers. The growth of these meters has been rapid in many parts of the world. There are close to 52 million smart meters alone installed in the United States [20]. Conventional metering infrastructure only facilitates monthly data collection for billing puruposes, while the new AMI framework allows for sub-hourly electrical demand readings. This data is primarily used for demand characterization and billing, however many additional uses are being discovered. A wide-range of studies have been completed in recent years to focus on a range of issues related to automatically extracting information from these data using unsupervised techniques. In this section, three sub-categories of application are discussed: load profiling, account classification, and disaggregation. Table 1 lists these publications and their general attributes.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2015	Load Profiling	Computer Science	Clustering	K-Means	[21]
2015	Customer Class.	Computer Science	Visual Analytics	Patterns	[22]

Continued on next page

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2015	Load Profiling	Elec. Engineering	Clustering	PCA	[23]
2015	Load Profiling	Elec. Engineering	Clustering	Multiple	[24]
2015	Customer Class.	Computer Science	Novelty Detection	Markov	[25]
2014	Disaggregation	Elec. Engineering	Motif Detection	SAX	[26]
2014	Load Profiling	Building Energy	Clustering	K-Means	[27]
2014	Load Profiling	Elec. Engineering	Clustering	K-Means	[28]
2014	Load Profiling	Computer Science	Visual Analytics	Patterns	[29]
2014	Customer Class.	Elec. Engineering	Visual Analytics	Multiple	[30]
2013	Disaggregation	Computer Science	Motif Detection	DPGMM	[31]
2013	Load Profiling	Elec. Engineering	Clustering	Ant Colony	[32]
2013	Customer Class.	Building Energy	Clustering	Multiple	[33]
2013	Load Profiling	Elec. Engineering	Clustering	Multiple	[34]
2012	Load Profiling	Elec. Engineering	Clustering	Multiple	[35]
2012	Customer Class.	Building Sim.	Clustering	Bayesian	[36]
2011	Load Profiling	Elec. Engineering	Motif Detection	ISPC	[37]
2010	Customer Class.	Elec. Engineering	Clustering	Multiple	[38]
2009	Load Profiling	Computer Science	Clustering	Multiple	[39]
2009	Load Profiling	Elec. Engineering	Clustering	Multiple	[40]
2009	Load Profiling	Elec. Engineering	Clustering	SVM	[41]
2009	Customer Class.	Elec. Engineering	Clustering	Multiple	[42]
2008	Customer Class.	Computer Science	Clustering	SOM	[43]
2006	Customer Class.	Elec. Engineering	Clustering	SOM	[44]
2005	Customer Class.	Elec. Engineering	Clustering	K-Means	[45]

Table 1: Publications from the Smart Meter category

3.1. Load Profiling

Load profiling is the process of grouping temporal subsequences of measured energy data for the purpose of characterizing the typical behaviour of an individual customer. It generally involves time-series clustering and feature extraction. Chicco et al. [41] provides an initial example in our review of this process using support vector machine clustering. Gullo et al. [39] and Räsänen et al. [40] took the process further by introducing a framework of various clustering processes that were implemented on case studies. Ramos et al. [35], Iglesias et al. [34] and Panapakidis et al. [24] tested various conventional and new clustering methods and similarity metrics in order to determine those most applicable to electrical load profiling. Chicco et al. [32] explored new clustering techniques based on ant colony grouping while Pan et al. [23] discovered the use of kernal PCA for the same purpose. Several groups of reseachers such as Lavin and Klabjan [27] and Green et al. [28] have found effective use in using the basic K-Means clustering algorithm for load profiling. Shahzadeh et al. [21] discussed the use of profiling as applied to forecast accuracy of temporal data. Two studies diverge from the standard profile development using clustering paradigm. The first is by De Silva et al. [37] who uses Incremental Summarization and Pattern Characterization (ISPC) instead of clustering to find load profiles. The other is the visual analytics-based approach of creating a smart meter analytics dashboard by Nezhad et al. [29] to create and inspect typical load profiles.

3.2. Customer Classification

Automated account classification is the next sub-category that utilizes unsupervised learning technques within the smart meter domain. These methods often employ load profile clustering as a first step, but differentiate themselves in using those features to classify accounts, or buildings, that fit within various categories. Therefore, account classification is a type of manual semi-supervised analysis utilizing load profiling as a basis. The study by Figueiredo et al. [45] harnessed K-Means and a labeled sample from accounts in Portugal to showcase this concept. Verdu et al. [44] and Räsänen et al. [43] applied self-organizing maps (SOM) to accomplish a similar study that classifies accounts according to applicability of several demand response scenarios. Vale et al. [42] gives an overview of a general data mining framework focused on characterizing customers. Florita et al. [36] diverge from the use of measured data by creating a massive amount of simulation data of load profiles in order to quantify energy storage applications for the power grid. Fagiani et al. [25] use Markov Model novelty detection to automatically classify customers who potentially have leakage or waste issues. Çakmak et al. [30] and Liu et al. [22] test new visual analytics techniques within more holistic analysis framework for analyzing customers. Borgeson used various clustering and occupancy detection techniques to analyze a large AMI data set from California [33]. Bidoki et al. tested various clustering techniques to evaluate applicability for customer classification [38].

3.3. Disaggregation

The last area of smart meter data analysis is the field of meter disaggregation. Disaggregation attempts decompose a measurement signal from a high level reading to the individual loads being measured. This domain is well-researched from a supervised model perspective but recent attempts at unsupervised pattern-based disaggregation were developed to facili-ate implementation on unlabelled smart meter data. Shao et al. [31] use Dirichlet Process Gaussian Mixture Models to find and disaggregate patterns in sub-hourly meter data. Reinhardt and Koessler [26] use a version of symbolic aggregate approximation (SAX) to extract and idenfity dissaggregated patterns for the purpose of prediction. These studies are also unique in that few of the disaggregation studies focus on commercial buildings as opposed to residential buildings.

4. Portfolio Analytics

Portfolio analysis is a domain in which a large group of buildings, often located on the same geographical area or owned or managed by the same entity, are analyzed for the purpose of managing or optimizing the group as a whole. Each subsection covers the publications reviewed in this domain that fall into three categories: characterization, classification, and targeting. Table 2 illustrates an overview of all the publications in this section in addition to their attributes.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2015	Characterization	Building Sim.	Clustering	Multiple	[14]
2015	Characterization	Building Energy	Visual Analytics	EIS	[46]
2015	Characterization	Building Energy	Motif Detection	SAX	[17]
2015	Characterization	Elec. Engineering	Visual Analytics	EIS	[47]
2015	Classifying	Building Energy	Clustering	Multiple	[48]
2014	Classifying	Elec. Engineering	Novelty Detection	GAM	[49]
2014	Characterization	Building Sim.	Visual Analytics	Patterns	[50]
2014	Classifying	Building Energy	Clustering	K-Means	[51]
2013	Characterization	Computer Science	Visual Analytics	Patterns	[52]
2013	Targeting	Building Energy	Rule Extraction	DAR	[53]
2012	Classifying	Elec. Engineering	Clustering	Multiple	[54]
2012	Targeting	Building Energy	Clustering	ML	[55]
2012	Characterization	Building Sim.	Clustering	K-Means	[56]
2011	Targeting	Computer Science	Novelty Detection	Markov	[57]
2011	Characterization	Building Energy	Visual Analytics	EIS	[58]
2010	Characterization	Building Energy	Visual Analytics	EIS	[59]
2010	Targeting	Building Energy	Clustering	Multiple	[60]
2009	Targeting	Computer Science	Clustering	ML	[61]
2009	Characterization	Building Energy	Visual Analytics	EIS	[62]
2009	Characterization	Computer Science	Visual Analytics	EIS	[63]
2008	Characterization	Building Energy	Novelty Detection	PCA	[64]
2007	Classifying	Building Energy	Clustering	Fuzzy Clust.	[65]
2005	Characterization	Building Energy	Clustering	Hierarchical	[66]

Table 2: Publications from the Portfolio Analysis category

4.1. Characterization

Publications that address the characterization of a portfolio of buildings include unsupervised techniques meant to evaluate and visualize the range of behaviours and performance of the group. A majority of the techniques utilized are either clustering or visual analytics that provide a mode of exploratory analysis that enable further steps. Seem [66] produced an influential study that extracts days of the week with similar consumption profiles. Further clustering work was completed by An et al. [56] to estimate thermal parameters of a portfolio of buildings. Lam et al. [64] used Principal Component Analysis to extract information about a group of office buildings. Approaches focused on visual analytics and dashboards were completed by Agarwal et al. [63], Lehrer [62], and Lehrer and Vasudev [58]. Granderson et al. [59] completed a case study-based evaluation of energy information systems, in which some methods combine some unsupervised approaches with visualization. Diong et al. [47] completed a case study as well focused on a specific energy information system implementation. Morán et al. [52] and Georgescu and Mezic [50] developed hybrid methods that employed visual continous maps and Koopman Operator methods respectively to visualize portfolio consumption. Miller et al. [14, 17] completed two studies focused on the use of screening techniques to automatically extract diurnal patterns from performance data and use those patterns to characterize the consumption of a portfolio of buildings.

Yarbrogh et al. used visual analytics techniques to analyze peak demand on a university campus [46].

4.2. Classifying

The concept of classifying buildings within a portfolio supplements the characterization techniques by assigning individual buildings to a subgroups of relative performance for the purpose of benchmarking or decision-making. Santamouris et al. [65] produced a report using clustering and classification to assign schools in Greece to subgroups of similar performance. Nikolaou et al. [54] and Pieri et al. [48] further extended this type of work to office buildings and hotels. Heidarinejad et al. [51] released an analysis of clustered simulation data to classify LEED certified office buildings. Ploennigs et al. [49] created a platform for monitoring, diagnosing and classifying buildings and operational behaviour within a portfolio in addition to quickly visualizing the outputs.

4.3. Targeting

Targeting is a concept that builds upon characterization and classification to identify specific buildings or measures to be implemented in a portfolio to improve performance. These publications are differentiated in that specific measures are identified in the analysis. Sedano et al. [61] uses Cooperative Maximum-Likelihood amongst other techniques to evaluate the thermal insulation performance of buildings. Gaitani et al. [60] used PCA and clustering to target heating efficiency in school buildings. Bellala et al. [57] used various methods to find lighting energy savings on a campus of a large organization. Petcharat et al. [55] also found lighting energy savings on a group of buildings. Cabrera and Zareipour [53] used data association rules to complete a similar study to find wasteful patterns.

5. Operations, Optimization, and Controls

Unsupervised techniques focused on individual buildings themselves are placed in the category for building operations, optimization and control. This category contains the largest number of publications and it incorporates a wider range of applications. It is differentiated from Section 6 in that the applications are not as focused on detecting and fixing anomalous behaviour. This section evaluates publications within the sub-categories of occupancy detection, retrofit analysis, controls, and energy management. Table 3 outlines the publications in this section and their key attributes.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2015	Occupancy Detection	Computer Science	Clustering	K-Means	[67]
2015	Occupancy Detection	Building Energy	Rule Extraction	Markov	[68]
2015	Energy Management	Building Energy	Rule Extraction	DAR	[69]
2015	Energy Management	Building Energy	Visual Analytics	EIS	[70]
2015	Occupancy Detection	Building Energy	Rule Extraction	Multiple	[71]

Continued on next page

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2015	Controls	Elec. Engineering	Motif Detection	SAX	[72]
2014	Energy Management	Building Energy	Rule Extraction	DAR	[73]
2014	Controls	Elec. Engineering	Rule Extraction	SVM	[74]
2013	Energy Management	Elec. Engineering	Visual Analytics	Patterns	[75]
2013	Controls	Building Energy	Clustering	Multiple	[76]
2013	Energy Management	Building Sim.	Rule Extraction	DAR	[77]
2013	Controls	Building Sim.	Rule Extraction	DAR	[78]
2013	Controls	Building Energy	Clustering	K-Means	[79]
2013	Controls	Computer Science	Clustering	EMC	[80]
2012	Energy Management	Elec. Engineering	Visual Analytics	Patterns	[81]
2012	Occupancy Detection	Elec. Engineering	Motif Detection	ML	[82]
2011	Controls	Building Energy	Rule Extraction	DAR	[83]
2011	Energy Management	Building Energy	Visual Analytics	EIS	[84]
2011	Occupancy Detection	Computer Science	Clustering	Multiple	[85]
2011	Occupancy Detection	Building Sim.	Motif Detection	Markov	[86]
2011	Controls	Computer Science	Visual Analytics	Patterns	[87]
2010	Controls	Computer Science	Motif Detection	SAX	[88]
2009	Controls	Computer Science	Motif Detection	SAX	[89]
2008	Controls	Elec. Engineering	Clustering	Multiple	[90]

Table 3: Publications from the Operations, Optimization, and Controls category

5.1. Occupancy Detection

Occupancy detection using unsupervised techniques infers human presence in a non-residential building without a labeled ground truth data set or as part of a semi-supervised approach using a subset of labeled data. This occupancy detection is then used for analysis or as inputs for control of systems. Augello et al. [85] used multiple techniques to infer occupant presence on a campus in Italy. Dong and Lam [86] used Hidden Markov Models to detect occupancy patterns that were then used in simulation. Thanayankizil et al. [82] developed a concept called Context Profiling in which occupancy was detected temporily and spatially. Mansur et al. [67] used clustering to detect occupancy patterns from sensor data. The newest studies by Adamopoulou et al. [68] and D'Oca and Hong [71] use a range of techniques to extract rules related to occupancy.

5.2. Controls

Controls optimization is an enduring field of study aimed at creating a state of the best operation and energy performance for a building system such as heating, cooling, ventilation or lighting. Kusiak and Song [90] created a means of optimally controlling a heating plant with clustering as a key step. Patnaik et al. [88, 89] completed studies focused on using motif detection to find modes of chilled water plant operation that proved most optimal. Hao et al. [87] built upon these concepts to create a visual analytics tool to investigate these motifs. May-Ostendorp et al. [83, 78] used rule extraction as a means of enhancing a model-predictive control process of mixed-mode systems. Bogen et al. [76] used clustering to detect usage patterns for building control system evaluation. Fan et al. [79] used clustering to enhance chiller power prediction with the ultimate goal of control optimization. Hong et al. [80] used Empirical Mode Decomposition to spatially optimize the placement of sensors in a building. Domahidi et al. [74] used support vector machines (SVM) to extract optimized rules for supervisory control. Habib and Zucker use SAX to identify common motifs of an absorption chiller for the purpose of characterization and control [72].

5.3. Energy Management

Energy management and analysis of an individual building using unsupervised techniques is becoming common due to the increasing amounts of raw building management (BMS) and energy management system (EMS) data. Users of these techniques are often facilities management professionals or consultants who undertake the process in order to understand how the building is consuming energy. Duarte et al. [84] uses visual analytics to process data from an EMS along with various pre-processing techniques. Lange et al. [81, 75] created two overview studies focused spatio-temporal visualization of building performance data and its interpretation in various case studies. Gayeski et al. [70] completed a recent survey of building operations professionals on their use of graphical interfaces of BMS and EMS dashboards. Outside of the visual analytics realm, Fan et al. [69], Xiao and Fan [73], and Yu et al. [77] completed studies of an entire data mining using framework using data association rules to improve operational performance.

6. Anomaly Detection

Anomaly detection for buildings generally focuses on the detection and diagnostics of problems occuring within a building, its subsystems and components. This field is most often focuses on the use of novelty detection or clustering approaches to find anomalous behaviour. The sub-categories for this section are divided according to the spatial hierarchy of systems within a building; the highest level being whole building consumption, down to the subsystems such as heating, cooling or lighting and then to the individual components within those systems. Table 4 provides an overview of the publications in this category.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2015	Whole Building	Building Energy	Novelty Detection	DAR	[91]
2015	Whole Building	Building Energy	Novelty Detection	Multiple	[92]
2015	Subsystem	Elec. Engineering	Novelty Detection	DAR	[93]
2014	Subsystem	Building Energy	Novelty Detection	PCA	[94]
2014	Whole Building	Elec. Engineering	Novelty Detection	GAM	[95]
2014	Whole Building	Building Energy	Clustering	K-Means	[96]
2014	Subsystem	Computer Science	Novelty Detection	Multiple	[97]
2014	Subsystem	Building Sim.	Novelty Detection	PCA	[98]
2013	Whole Building	Elec. Engineering	Novelty Detection	GAM	[99]

Continued on next page

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2013	Whole Building	Computer Science	Visual Analytics	Patterns	[100]
2013	Component	Computer Science	Novelty Detection	EMC	[101]
2013	Whole Building	Computer Science	Novelty Detection	Multiple	[102]
2012	Subsystem	Elec. Engineering	Novelty Detection	Nearest Neighbor	[103]
2012	Component	Building Energy	Novelty Detection	Wavelets	[104]
2012	Whole Building	Elec. Engineering	Novelty Detection	Fourier	[105]
2012	Component	Building Energy	Rule Extraction	DAR	[106]
2010	Whole Building	Computer Science	Novelty Detection	CART	[107]
2010	Whole Building	Building Sim.	Clustering	Hierarchical	[108]
2010	Subsystem	Building Energy	Novelty Detection	PCA	[109]
2010	Subsystem	Computer Science	Visual Analytics	Patterns	[110]
2008	Subsystem	Elec. Engineering	Novelty Detection	LDA	[111]
2006	Whole Building	Building Energy	Novelty Detection	GESD	[112]
2005	Component	Building Energy	Novelty Detection	PCA	[113]

Table 4: Publications from the Anomaly Detection category

6.1. Whole Building

Whole building anomaly detection uses the electricity or heating and cooling energy supply incoming to a building to determine sub-sequences of poor performance. This category is complimentary to many of the Smart Meter solutions as they both focus on the use of a single data stream for a building. Seem [112] had an early work again in this category with his work in using novelty detection to find abnormal days of consumption in buildings. Liu et al. [107] used classification and regression trees (CART) and Wrinch et al. [105] frequency domain analysis for the same purpose. Jacob et al. utilized hierarchical clustering to use as variables in regression models for whole building monitoring [108]. Fontugne et al. [102] created a process known as the *Strip, Bind, and Search* method to automatically uncover misbehavior from the whole building level and subsequently detects the source of the anomaly. Janetzko et al. developed a visual analytics platform to highlight anomalous behavior in power meter data [100]. Chou and Telaga created a hybrid whole building anomaly detection process using K-means [96]. Ploennigs et al. [99] and Chen et al. [95] created similar systems that uses generalized additive models (GAM). In the most recent work, Capozzoli et al. [92] and Fan et al. [91] use various techniques as part of a framework to detect and diagnose performance problems.

6.2. Subsystems

Subsystem anomaly detection focuses on the use of a broader data set to detect and diagnose faults from a lower level. Yoshida et al. [111] provided a semi-supervised approach that seeks to determine which variables within a building are most influential in contributing to overall building performance. Wang et al. [109] uses PCA to diagnose sensor failures. Forlines and Wittenberg visualized multi-dimensional data using what they call the Wakame diagram [110]. Linda et al. [103] and Wijayasekara et al. [97] use various techniques to diagnose system faults and visualize them spatially. Le Cam et al. [98] use PCA to create inverse models to detect problems in HVAC systems. Li and Wen [94] created a similar process using PCA in conjunction with wavelet transform. Sun et al. [93] used data assocation rules to create fault detection thresholds for finding anomalies.

6.3. Components

Component level anomaly detection is a bottom-up fault detection approach that focuses on determining faults in individual equipemnt. Wang and Cui [113] use PCA to detect component faults in chilled water plants. Yu et al. [106] and Fontugne et al. [101] both compliment their work at the whole building level to find associated component performance anomalies automatically. Zhu et al. [104] use wavelets to diagnose issues in air handling units (AHU).

7. Discussion

Analysis of the publications reviewed in this paper at a high level reveals insights regarding the frequency and types of techniques used, the number of publications and trend over time, and the popularity of specific publications within the domain. This section outlines each of these considerations.

7.1. Specific Techniques

[H] Figure 5 illustrates the frequency of use of various unsupervised and visual analytics methods in the publications reviewed. The largest specific category are papers that use numerous techniques in conjuction to achieve a complex goal. Amongst the papers that have a dominant technique, K-Means clustering, data assocation rules, and visual analytics-based pattern analysis and energy information systems round out the top five.

7.2. Research Sectors and Trends

Figure 6 illustrates the breakdown of publications based on the year published since 2005. They are further divided into four general research domains: building energy analysis, building simulation, computer science and electrical engineeering. These research domain categories were subjectively determined for each paper through evaluating a combination of which university department the authors were from and which publication the research was published in. Building energy analysis pertains to researchers who predominantly focus on measured data analysis from buildings, while building simulation experts generally research forward modelling and simulation of building and urban systems. Both fields of study most often exist within architecture or mechanical engineering departments. Electrical engineering and computer science are two well-established domains and generally exist in their own departments. It is noticed that there is a gradual increase in the number of publications over the last ten years with electrical engineering and building energy analysis being the most common in the first few years and computer science and building simulation picking up since 2008.

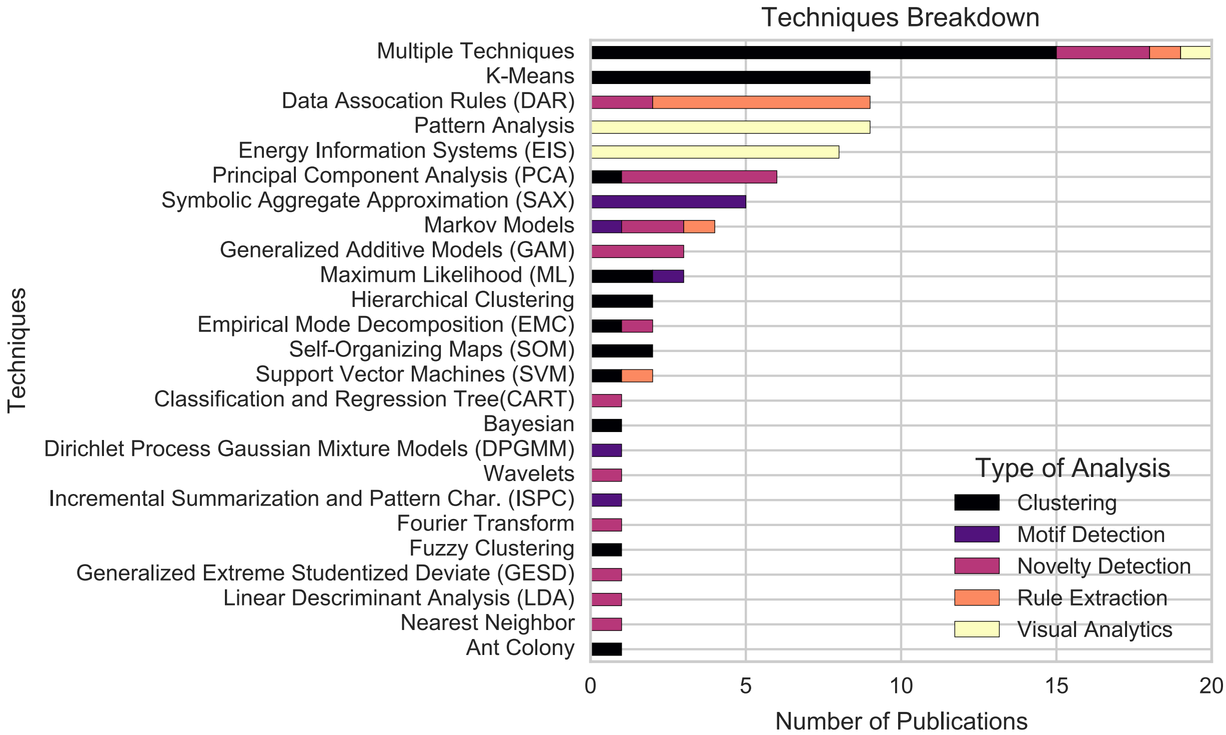

Figure 5: Breakdown of specific and general analytics techniques utilized

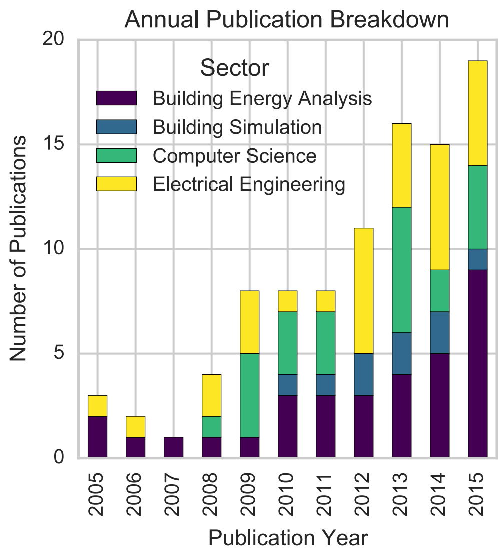

Figure 6: Breakdown of publications by year published and research domain

7.3. *Publications Venues*

This section analyzes the prevalence of certain publication venues within this review. Figure 7 illustrates the breakdown of the publication venues represented. The Energy and Buildings Journal from the building energy analysis research domain dominates this list with 17 publications. Building simulation and energy analysis research domains publish most often in this journal as well as Applied Energy and Energy Efficiency. Several IEEE conferences and journals are also dominant as most of the papers from the electrical engineering domain are in these venues..

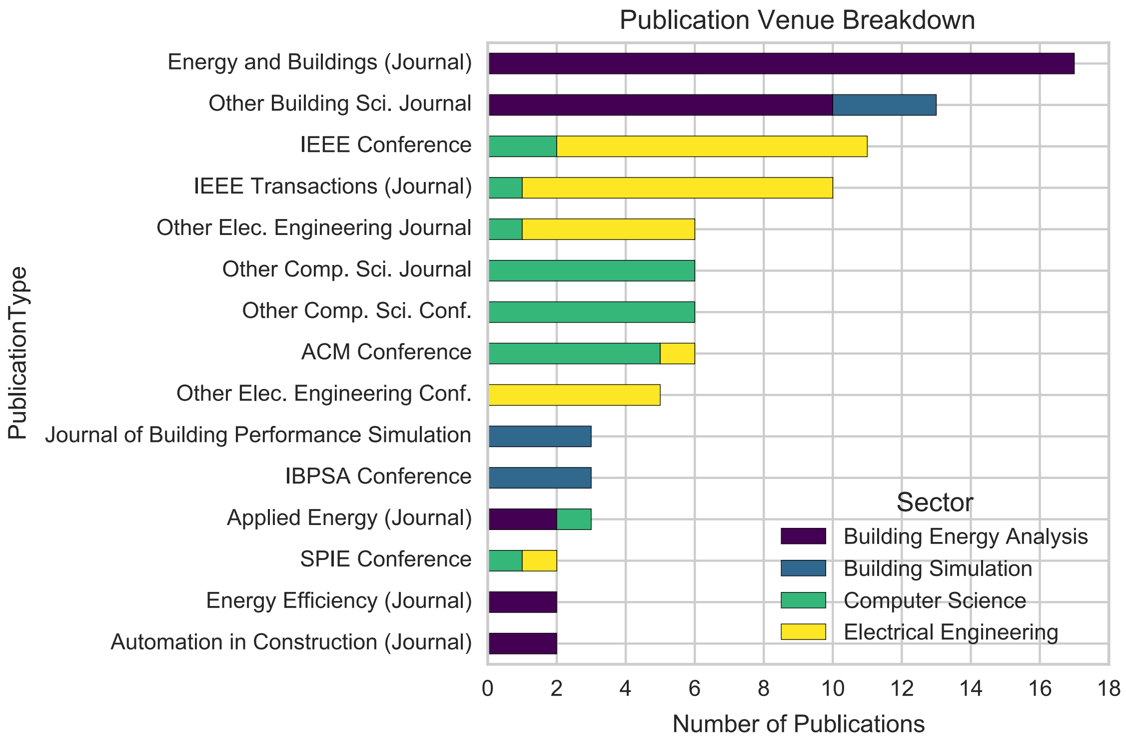

Figure 7: Breakdown of journal publications by publication name and research domain

8. Conclusion

This review provides an overview of 98 publications from several research domains that utilize unsupervised statistical learning and visual analytics as applied to non-residential building performance analysis. The publications are categorized according to machine learning technique and application and sub-application of utilization. Clustering algorithms, and especially the K-Means algorithm, stand out as popular amongst most of the application domains. Visual analytics approaches are also quite popular as generally-applied techniques. Almost half of the techniques implemented were only observed once in the publications as a dominant approach, thus there is opportunity to further explore their potential. The publications selected increase in number steadily over the last decade with computer science and electrical engineering slowly identifying the non-residential building industry as a good target for implementation of developed techniques. The dominant publication venues in this review came from the building energy analysis domain. However, many computer science and electrical engineering publication venues are quickly catching up due emerging interest in the building data analytics domain.

References

[1] M. Krarti, An Overview of Artificial Intelligence-Based Methods for Building Energy Systems, Journal of Solar Energy Engineering 125 (3) (2003) 331. doi:10.1115/1.1592186.
URL http://SolarEnergyEngineering.asmedigitalcollection.asme.org/article.aspx?articleid=1456822

[2] A. I. Dounis, Artificial intelligence for energy conservation in buildings, Advances in Building Energy Research 4 (1) (2010) 267–299. doi:10.3763/aber.2009.0408.
URL http://www.tandfonline.com/doi/abs/10.3763/aber.2009.0408

[3] T. A. Reddy, Applied data analysis and modeling for energy engineers and scientists, Springer Science & Business Media, 2011.
URL https://books.google.com/books?hl=en&lr=&id=UkWvWDFiDf4C&oi=fnd&pg=PR9&dq=+%22Applied+data+analysis+and+modeling+for+energy+engineers+and+scientists%22+T+Agami+Reddy&ots=tEUJEuedXl&sig=e41JNtpT78gpFI64XB6sNt_hQF4

[4] S. H. Lee, T. Hong, M. A. Piette, S. C. Taylor-Lange, Energy retrofit analysis toolkits for commercial buildings: A review, Energydoi:10.1016/j.energy.2015.06.112.
URL http://www.sciencedirect.com/science/article/pii/S0360544215008683

[5] D. Ioannidis, A. Fotiadou, S. Krinidis, G. Stavropoulos, D. Tzovaras, S. Likothanassis, Big Data and Visual Analytics for Building Performance Comparison, in: R. Chbeir, Y. Manolopoulos, I. Maglogiannis, R. Alhajj (Eds.), Artificial Intelligence Applications and Innovations, no. 458 in IFIP Advances in Information and Communication Technology, Springer International Publishing, 2015, pp. 421–430, dOI: 10.1007/978-3-319-23868-5_30.
URL http://link.springer.com/chapter/10.1007/978-3-319-23868-5_30

[6] J. Morais, Y. Pires, C. Cardoso, A. Klautau, An Overview of Data Mining Techniques Applied to Power Systems, Data Mining and Knowledge Discovery in Real Life Applications.
URL http://www.intechopen.com/source/pdfs/5962/InTech-An_overview_of_data_mining_techniques_applied_to_power_systems.pdf

[7] G. Chicco, Overview and performance assessment of the clustering methods for electrical load pattern grouping, Energy 42 (1) (2012) 68–80. doi:10.1016/j.energy.2011.12.031.
URL http://dx.doi.org/10.1016/j.energy.2011.12.031

[8] K.-l. Zhou, S.-l. Yang, C. Shen, A review of electric load classification in smart grid environment, Renewable and Sustainable Energy Reviews 24 (C) (2013) 103–110. doi:10.1016/j.rser.2013.03.023.
URL http://dx.doi.org/10.1016/j.rser.2013.03.023

[9] R. Ruparathna, K. Hewage, R. Sadiq, Improving the energy efficiency of the existing building stock: A critical review of commercial and institutional buildings, Renewable and Sustainable Energy Reviews 53 (2016) 1032–1045. doi:10.1016/j.rser.2015.09.084.
URL http://www.sciencedirect.com/science/article/pii/S1364032115010540

[10] T. Hastie, R. Tibshirani, J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, 2nd Edition, Springer series in statistics, Springer, New York, NY, 2009.
URL http://link.springer.com/article/10.1007%2FBF02985802

[11] G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning, Vol. 112, Springer, 2013.
URL http://link.springer.com/content/pdf/10.1007/978-1-4614-7138-7.pdf

[12] R. O. Duda, P. E. Hart, D. G. Stork, Pattern classification, John Wiley & Sons, 2012.
URL https://books.google.com.sg/books?hl=en&lr=&id=Br33IRC3PkQC&oi=fnd&pg=PR3&dq=pattern+classification+duda&ots=2wCPIsgbDs&sig=Uq-PAZRxsyIu0Icp7zxDThyV_oU

[13] B. Mirkin, Clustering: a data recovery approach, CRC Press, 2012.
URL https://books.google.com.sg/books?hl=en&lr=&id=34-q8geA3yEC&oi=fnd&pg=PP1&dq=clustering+a+data+recovery&ots=XEGQGoSOZY&sig=lq2rKFBq1bpGKQM8gdYS9xI8hrc

[14] C. Miller, A. Schlueter, Forensically discovering simulation feedback knowledge from a campus energy information system, in: Proceedings of the 2015 Symposium on Simulation for Architecture and Urban Design (SimAUD 2015), SCS, Washington DC, USA, 2015, pp. 33–40. doi:10.13140/RG.2.1.2286.0964.
URL http://dx.doi.org/10.13140/RG.2.1.2286.0964

[15] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM computing surveys (CSUR) 41 (3) (2009) 15.
URL http://dl.acm.org/citation.cfm?id=1541882

[16] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, S. Frankenstein, M. Lerner, GrammarViz 2.0: a tool for grammar-based pattern discovery in time series, Springer, 2014.
URL http://link.springer.com/chapter/10.1007/978-3-662-44845-8_37

[17] C. Miller, Z. Nagy, A. Schlueter, Automated daily pattern filtering of measured building performance data, Automation in Construction 49, Part A (2015) 1–17. doi:10.1016/j.autcon.2014.09.004.
URL http://www.sciencedirect.com/science/article/pii/S0926580514002015

[18] D. Keim, G. Andrienko, J.-D. Fekete, C. Grg, J. Kohlhammer, G. Melanon, Visual Analytics: Definition, Process, and Challenges, in: A. Kerren, J. T. Stasko, J.-D. Fekete, C. North (Eds.), Information Visualization, no. 4950 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2008, pp. 154–175, dOI: 10.1007/978-3-540-70956-5_7.
URL http://link.springer.com/chapter/10.1007/978-3-540-70956-5_7

[19] B. Shneiderman, The eyes have it: a task by data type taxonomy for information visualizations, in: , IEEE Symposium on Visual Languages, 1996. Proceedings, 1996, pp. 336–343. doi:10.1109/VL.1996.545307.

[20] Energy Information Administration, How many smart meters are installed in the United States, and who has them? (May 2015).
URL http://www.eia.gov/tools/faqs/faq.cfm?id=108&t=3

[21] A. Shahzadeh, A. Khosravi, S. Nahavandi, Improving load forecast accuracy by clustering consumers using smart meter data, in: Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015, pp. 1–7. doi:10.1109/IJCNN.2015.7280393.

[22] X. Liu, L. Golab, I. Ilyas, SMAS: A smart meter data analytics system, in: Proceedings of the 2015 IEEE 31st International Conference on Data Engineering (ICDE), Seoul, South Korea, 2015, pp. 1476–1479. doi:10.1109/ICDE.2015.7113405.

[23] E. Pan, H. Li, L. Song, Z. Han, Kernel-based non-parametric clustering for load profiling of big smart meter data, in: Proceedings of the 2015 IEEE Wireless Communications and Networking Conference (WCNC), 2015, pp. 2251–2255. doi:10.1109/WCNC.2015.7127817.

[24] I. Panapakidis, M. Alexiadis, G. Papagiannis, Evaluation of the performance of clustering algorithms for a high voltage industrial consumer, Engineering Applications of Artificial Intelligence 38 (C) (2015) 1–13. doi:10.1016/j.engappai.2014.10.013.
URL http://dx.doi.org/10.1016/j.engappai.2014.10.013

[25] M. Fagiani, S. Squartini, M. Severini, F. Piazza, A novelty detection approach to identify the occurrence of leakage in smart gas and water grids, in: Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015, pp. 1–8. doi:10.1109/IJCNN.2015.7280473.

[26] A. Reinhardt, S. Koessler, PowerSAX: Fast motif matching in distributed power meter data using symbolic representations, in: Proceedings of 9th IEEE International Workshop on Practical Issues in

Building Sensor Network Applications (SenseApp 2014), IEEE, Edmonton, Canada, 2014, pp. 531–538. doi:10.1109/LCNW.2014.6927699.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6927699

[27] A. Lavin, D. Klabjan, Clustering time-series energy data from smart meters, Energy Efficiency 8 (4) (2014) 681–689. doi:10.1007/s12053-014-9316-0.
URL http://link.springer.com/article/10.1007/s12053-014-9316-0

[28] R. Green, I. Staffell, N. Vasilakos, Divide and Conquer K-Means Clustering of Demand Data Allows Rapid and Accurate Simulations of the British Electricity System, IEEE Transactions on Engineering Management 61 (2) (2014) 251–260. doi:10.1109/TEM.2013.2284386.

[29] A. Jarrah Nezhad, T. K. Wijaya, M. Vasirani, K. Aberer, SmartD: Smart Meter Data Analytics Dashboard, in: Proceedings of the 5th International Conference on Future Energy Systems (e-Energy '14), e-Energy '14, ACM, New York, NY, USA, 2014, pp. 213–214. doi:10.1145/2602044.2602046.
URL http://doi.acm.org/10.1145/2602044.2602046

[30] H. K. akmak, H. Maa\s s, F. Bach, U. G. Khnapfel, A new framework for the analysis of large scale multi-rate power data, Karlsruhe Institute of Technology, 2014.
URL http://www.researchgate.net/profile/Hueseyin_Cakmak2/publication/264579714_A_New_Framework_for_the_Analysis_of_Large_Scale_Multi-Rate_Power_Data/links/53e8b9820cf25d674ea8552d.pdf

[31] H. Shao, M. Marwah, N. Ramakrishnan, A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation: Applications to Residential and Commercial Buildings, in: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI 2013), Vol. 250, Bellevue, WA, USA, 2013, p. 2250.
URL http://people.cs.vt.edu/~naren/papers/temp-motif-aaai13.pdf

[32] G. Chicco, O.-M. Ionel, R. Porumb, Electrical Load Pattern Grouping Based on Centroid Model With Ant Colony Clustering, IEEE Transactions on Power Systems 28 (2) (2013) 1706–1715.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6331025

[33] S. Borgeson, Targeted Efficiency: Using Customer Meter Data to Improve Efficiency Program Outcomes, PhD, University of California, Berkeley, Berkeley, CA, USA (2013).
URL https://www.dropbox.com/s/ppi9az3z97mxx3u/Borgeson_dissertation.pdf

[34] F. Iglesias, W. Kastner, Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns, Energies 6 (2) (2013) 579–597. doi:10.3390/en6020579.
URL http://www.mdpi.com/1996-1073/6/2/579/

[35] S. Ramos, J. Duarte, J. Soares, Z. Vale, F. J. Duarte, Typical load profiles in the smart grid contextA clustering methods comparison, in: Proceedings of the Power and Energy Society General Meeting, 2012 IEEE, IEEE, San Diego, CA, USA, 2012, pp. 1–8.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6345565

[36] A. R. Florita, L. J. Brackney, T. P. Otanicar, J. Robertson, Classification of Commerical Building Electrical Demand Profiles for Energy Storage Applications, in: Proceedings of ASME 2012 6th International Conference on Energy Sustainability & 10th Fuel Cell Science, Engineering and Technology Conference (ESFuelCell2012), San Diego, CA, USA, 2012.
URL http://solarenergyengineering.asmedigitalcollection.asme.org/article.aspx?articleid=1697039

[37] D. De Silva, X. Yu, D. Alahakoon, G. Holmes, A Data Mining Framework for Electricity Consumption Analysis From Meter Data, IEEE Transactions on Industrial Informatics 7 (3) (2011) 399–407. doi:10.1109/TII.2011.2158844.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5893960

[38] S. M. Bidoki, N. Mahmoudi-Kohan, M. H. Sadreddini, M. Zolghadri Jahromi, M. P. Moghaddam, Evaluating different clustering techniques for electricity customer classification, in: Transmission and Distribution Conference and Exposition, 2010 IEEE PES, IEEE, New Orleans, LA, USA, 2010, pp. 1–5.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5484234

[39] F. Gullo, G. Ponti, A. Tagarelli, M. Ruffolo, D. Labate, Low-voltage electricity customer profiling based on load data clustering, in: Proceedings of the 2009 International Database Engineering & Applications Symposium (IDEAS 2009), ACM, Calabria, Italy, 2009, pp. 330–333.
URL http://dl.acm.org/citation.cfm?id=1620472

[40] T. Rsnen, M. Kolehmainen, Feature-based clustering for electricity use time series data, in: Adaptive and Natural Computing Algorithms: 9th International Conference, ICANNGA 2009, Springer, Kuopio, Finland, 2009, pp. 401–412.
URL http://link.springer.com/chapter/10.1007/978-3-642-04921-7_41

[41] G. Chicco, I. S. Ilie, Support vector clustering of electrical load pattern data, IEEE Transactions on Power Systems 24 (3) (2009) 1619–1628.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5072232

[42] Z. A. Vale, C. Ramos, S. Ramos, T. Pinto, Data mining applications in power systemsCase-studies and future trends, in: IEEE Transmission & Distribution Conference & Exposition: Asia and Pacific, 2009, IEEE, Seoul, South Korea, 2009, pp. 1–4.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5356830

[43] T. Rsnen, J. Ruuskanen, M. Kolehmainen, Reducing energy consumption by using self-organizing maps to create more personalized electricity use information, Applied Energy 85 (9) (2008) 830–840. doi:10.1016/j.apenergy.2007.10.012.
URL http://www.sciencedirect.com/science/article/pii/S030626190800007X

[44] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, F. J. G. Franco, Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps, IEEE Transactions on Power Systems 21 (4) (2006) 1672–1682. doi:10.1109/TPWRS.2006.881133.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1717570

[45] V. Figueiredo, F. Rodrigues, Z. Vale, J. B. Gouveia, An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques, IEEE Transactions on Power Systems 20 (2) (2005) 596–602. doi:10.1109/TPWRS.2005.846234.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1425550

[46] I. Yarbrough, Q. Sun, D. C. Reeves, K. Hackman, R. Bennett, D. S. Henshel, Visualizing building energy demand for building peak energy analysis, Energy and Buildings 91 (2015) 10–15. doi:10.1016/j.enbuild.2014.11.052.
URL http://www.sciencedirect.com/science/article/pii/S0378778814010081

[47] B. Diong, G. Zheng, M. Ginn, Establishing the foundation for energy management on university campuses via data analytics, in: Proceedings of the IEEE SoutheastCon 2015, Fort Lauderdale, FL, USA, 2015, pp. 1–7. doi:10.1109/SECON.2015.7132996.

[48] S. P. Pieri, IoannisTzouvadakis, M. Santamouris, Identifying energy consumption patterns in the Attica hotel sector using cluster analysis techniques with the aim of reducing hotels CO2 footprint, Energy and Buildings 94 (2015) 252–262. doi:10.1016/j.enbuild.2015.02.017.
URL http://www.sciencedirect.com/science/article/pii/S0378778815001127

[49] J. Ploennigs, B. Chen, P. Palmes, R. Lloyd, e2-Diagnoser: A System for Monitoring, Forecasting and Diagnosing Energy Usage, in: Data Mining Workshop (ICDMW), 2014 IEEE International Conference on, IEEE, Shenzhen, China, 2014, pp. 1231–1234. doi:10.1109/ICDMW.2014.56.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7022741

[50] M. Georgescu, I. Mezic, Site-Level Energy Monitoring and Analysis using Koopman Operator Methods, in: Proceedings of the 2014 ASHRAE/IBPSA-USA Building Simulation Conference (SimBuild 2014), ASHRAE/IBPSA, Atlanta, GA, USA, 2014.
URL https://www.ashrae.net/File%20Library/docLib/Events/ASHRAE-IPBSA-USA/Presentations/32_Georgescu.pdf

[51] M. Heidarinejad, M. Dahlhausen, S. McMahon, C. Pyke, J. Srebric, Cluster analysis of simulated energy use for LEED certified U.S. office buildings, Energy and Buildings 85 (2014) 86–97. doi:10.1016/j.enbuild.2014.09.017.
URL http://linkinghub.elsevier.com/retrieve/pii/S0378778814007087

[52] A. Morn, J. J. Fuertes, M. Domnguez, M. A. Prada, S. Alonso, P. Barrientos, Analysis of electricity bills using visual continuous maps, Neural Computing and Applications 23 (3-4) (2013) 645–655. doi:10.1007/s00521-013-1409-8.
URL http://link.springer.com/10.1007/s00521-013-1409-8
[53] D. F. M. Cabrera, H. Zareipour, Data association mining for identifying lighting energy waste patterns in educational institutes, Energy and Buildings 62 (2013) 210–216. doi:10.1016/j.enbuild.2013.02.049.
URL http://dx.doi.org/10.1016/j.enbuild.2013.02.049
[54] T. G. Nikolaou, D. S. Kolokotsa, G. S. Stavrakakis, I. D. Skias, On the Application of Clustering Techniques for Office Buildings' Energy and Thermal Comfort Classification, IEEE Transactions on Smart Grid 3 (4) (2012) 2196–2210. doi:10.1109/TSG.2012.2215059.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6387350
[55] S. Petcharat, S. Chungpaibulpatana, P. Rakkwamsuk, Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings, Energy and Buildings 52 (2012) 145–152. doi:10.1016/j.enbuild.2012.06.006.
URL http://dx.doi.org/10.1016/j.enbuild.2012.06.006
[56] L. An, R. Horesh, Y. T. Chae, Y. M. Lee, Estimation of Thermal Parameters of Buildings Through Inverse Modeling and Clustering for a Portfolio of Buildings, in: Proceedings of the Fifth National Conference of IBPSA-USA (SimBuild 2012), Madison, WI, USA, 2012.
URL http://www.ibpsa.us/simbuild2012/Papers/SB12_TS05a_4_An.pdf
[57] G. Bellala, M. Marwah, M. Arlitt, G. Lyon, C. E. Bash, Towards an understanding of campus-scale power consumption, in: Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys 2011), ACM Request Permissions, Seattle, WA, USA, 2011, p. 73. doi:10.1145/2434020.2434043.
URL http://dl.acm.org/citation.cfm?doid=2434020.2434043
[58] D. Lehrer, J. Vasudev, Visualizing Energy Information in Commercial Buildings: A Study of Tools, Expert Users, and Building Occupants, Tech. rep., Center for the Built Environment - UC Berkeley (Nov. 2011).
URL https://escholarship.org/uc/item/6vp5m5m3.pdf
[59] J. Granderson, M. A. Piette, G. Ghatikar, Building energy information systems: user case studies, Energy Efficiency 4 (1) (2010) 17–30. doi:10.1007/s12053-010-9084-4.
URL http://link.springer.com/article/10.1007/s12053-010-9084-4
[60] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, P. Patargias, Using principal component and cluster analysis in the heating evaluation of the school building sector, Applied Energy 87 (6) (2010) 2079–2086. doi:10.1016/j.apenergy.2009.12.007.
URL http://www.sciencedirect.com/science/article/pii/S0306261909005303
[61] J. Sedano, J. R. Villar, L. Curiel, E. De La Cal, E. Corchado, Improving energy efficiency in buildings using machine intelligence, in: Intelligent Data Engineering and Automated Learning-IDEAL 2009, Springer, 2009, pp. 773–782.
URL http://link.springer.com/chapter/10.1007/978-3-642-04394-9_95
[62] D. Lehrer, Research Scoping Report: Visualizing Information in Commercial Buildings, Tech. rep., Center for the Built Environment - UC Berkeley (Sep. 2009).
[63] Y. Agarwal, T. Weng, R. K. Gupta, The energy dashboard: improving the visibility of energy consumption at a campus-wide scale, in: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys 2009), ACM Press, Berkeley, CA, USA, 2009, p. 55. doi:10.1145/1810279.1810292.
URL http://portal.acm.org/citation.cfm?doid=1810279.1810292
[64] J. C. Lam, K. K. W. Wan, K. L. Cheung, L. Yang, Principal component analysis of electricity use in office buildings, Energy and Buildings 40 (5) (2008) 828–836. doi:10.1016/j.enbuild.2007.06.001.
URL http://linkinghub.elsevier.com/retrieve/pii/S0378778807001764
[65] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou,

P. Doukas, E. Primikiri, V. Geros, M. N. Assimakopoulos, R. Mitoula, S. Zerefos, Using intelligent clustering techniques to classify the energy performance of school buildings, Energy and Buildings 39 (1) (2007) 45–51. doi:10.1016/j.enbuild.2006.04.018.
URL http://linkinghub.elsevier.com/retrieve/pii/S0378778806001502

[66] J. E. Seem, Pattern recognition algorithm for determining days of the week with similar energy consumption profiles, Energy and Buildings 37 (2) (2005) 127–139. doi:10.1016/j.enbuild.2004.04.004.
URL http://linkinghub.elsevier.com/retrieve/pii/S0378778804001434

[67] V. Mansur, P. Carreira, A. Arsenio, A Learning Approach for Energy Efficiency Optimization by Occupancy Detection, in: Internet of Things. User-Centric IoT, Vol. 150, Springer International Publishing, Cham, 2015, pp. 9–15.
URL http://link.springer.com/10.1007/978-3-319-19656-5_2

[68] A. A. Adamopoulou, A. M. Tryferidis, D. K. Tzovaras, A context-aware method for building occupancy prediction, Energy and Buildingsdoi:10.1016/j.enbuild.2015.10.003.
URL http://www.sciencedirect.com/science/article/pii/S0378778815303212

[69] C. Fan, F. Xiao, H. Madsen, D. Wang, Temporal Knowledge Discovery in Big BAS Data for Building Energy Management, Energy and Buildingsdoi:10.1016/j.enbuild.2015.09.060.
URL http://www.sciencedirect.com/science/article/pii/S0378778815302991

[70] N. Gayeski, S. Kleindienst, J. Gagne, B. Werntz, R. Cruz, S. Samouhos, Data and Interfaces for Advanced Building Operations and Maintenance - RP 1633 Final Report, Tech. rep., ASHRAE (Jun. 2015).

[71] S. DOca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy and Buildings 88 (2015) 395–408. doi:10.1016/j.enbuild.2014.11.065.
URL http://www.sciencedirect.com/science/article/pii/S0378778814010329

[72] U. Habib, G. Zucker, Finding the Different Patterns in Buildings Data Using Bag of Words Representation with Clustering, in: 2015 13th International Conference on Frontiers of Information Technology (FIT), 2015, pp. 303–308. doi:10.1109/FIT.2015.60.

[73] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, Energy and Buildings 75 (2014) 109–118. doi:10.1016/j.enbuild.2014.02.005.
URL http://dx.doi.org/10.1016/j.enbuild.2014.02.005

[74] A. Domahidi, F. Ullmann, M. Morari, C. N. Jones, Learning decision rules for energy efficient building control, Journal of Process Control 24 (6) (2014) 763–772. doi:10.1016/j.jprocont.2014.01.006.
URL http://www.sciencedirect.com/science/article/pii/S0959152414000304

[75] B. Lange, N. Rodriguez, W. Puech, X. Vasques, Discovering unexpected information using a building energy visualization tool, in: A. M. Baskurt, R. Sitnik (Eds.), Proceedings of the IS&T/SPIE 2013 Conference on Electronic Imaging, Vol. 8650, SPIE, San Francisco, CA, USA, 2013, p. 86500Q. doi:10.1117/12.2000273.
URL http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2000273

[76] A. C. Bogen, M. Rashid, E. W. East, J. Ross, Evaluating a data clustering approach for life-cycle facility control, Journal of Information Technology in Construction Vol. 18 (2013) 99–118.
URL http://www.itcon.org/2013/6

[77] Z. J. Yu, B. C. M. Fung, F. Haghighat, Extracting knowledge from building-related data A data mining framework, Building Simulation 6 (2) (2013) 207–222. doi:10.1007/s12273-013-0117-8.
URL http://link.springer.com/10.1007/s12273-013-0117-8

[78] P. T. May-Ostendorp, G. P. Henze, B. Rajagopalan, C. D. Corbin, Extraction of supervisory building control rules from model predictive control of windows in a mixed mode building, Journal of Building Performance Simulation 6 (3) (2013) 199–219. doi:10.1080/19401493.2012.665481.
URL http://www.tandfonline.com/doi/abs/10.1080/19401493.2012.665481

[79] C. Fan, F. Xiao, W. Shengwei, Prediction of Chiller Power Consumption Using Time Series Analysis and Artificial Neural Networks, in: Proceedings of the 8th International Conference on Indoor Air

Quality, Ventilation and Energy Conservation in Buildings (CLIMA 2013), Prague, Czech Republic, 2013.

[80] D. Hong, J. Ortiz, K. Whitehouse, D. Culler, Towards Automatic Spatial Verification of Sensor Placement in Buildings, in: Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings (BuildSys 2013), Roma, Italy, 2013.
URL `http://www.cs.berkeley.edu/~jortiz/papers/buildsys13.pdf`

[81] B. Lange, N. Rodriguez, W. Puech, Energy Consumption Improvement Through a Visualization Software, in: Energy Efficiency - A Bridge to Low Carbon Economy, Intech, 2012, pp. 161–184.

[82] L. V. Thanayankizil, S. K. Ghai, D. Chakraborty, D. P. Seetharam, Softgreen: Towards energy management of green office buildings with soft sensors, in: Proceedings of the Fourth International Conference on Communication Systems and Networks (COMSNETS 2012), IEEE, Bangalore, India, 2012, pp. 1–6. `doi:10.1109/COMSNETS.2012.6151374`.
URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6151374`

[83] P. May-Ostendorp, G. P. Henze, C. D. Corbin, B. Rajagopalan, C. Felsmann, Model-predictive control of mixed-mode buildings with rule extraction, Building and Environment 46 (2) (2011) 428–437. `doi:10.1016/j.buildenv.2010.08.004`.
URL `http://linkinghub.elsevier.com/retrieve/pii/S0360132310002477`

[84] C. Duarte, B. Acker, R. Grosshans, M. Manic, K. Van Den Wymelenberg, C. Rieger, Prioritizing and Visualizing Energy Management and Control System Data to Provide Actionable Information For Building Operators, in: Proceedings from 2011 Western Energy Policy Research Conference, Boise, Idaho, USA, 2011.
URL `http://mhrg.if.uidaho.edu/papers/2011/WEPC11_DuarteAckerKevin_EMSDataVisualisation.pdf`

[85] A. Augello, M. Ortolani, G. L. Re, S. Gaglio, Sensor mining for user behavior profiling in intelligent environments, Advances in Distributed Agent-Based Retrieval Tools (2011) 143–158.
URL `http://link.springer.com/chapter/10.1007/978-3-642-21384-7_10`

[86] B. Dong, K. P. Lam, Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network, Journal of Building Performance Simulation 4 (4) (2011) 359–369. `doi:10.1080/19401493.2011.577810`.
URL `http://www.tandfonline.com/doi/abs/10.1080/19401493.2011.577810`

[87] M. Hao, M. Marwah, H. Janetzko, R. Sharma, D. A. Keim, U. Dayal, D. Patnaik, N. Ramakrishnan, Visualizing frequent patterns in large multivariate time series, in: Proceedings of SPIE Conference on Visualization and Data Analysis 2011, Vol. 7868, SPIE, San Francisco, CA, USA, 2011, pp. 78680J–78680J–10. `doi:10.1117/12.872169`.
URL `http://dx.doi.org/10.1117/12.872169`

[88] D. Patnaik, M. Marwah, R. K. Sharma, N. Ramakrishnan, Data Mining for Modeling Chiller Systems in Data Centers, in: Advances in Intelligent Data Analysis IX, Vol. 6065, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 125–136.
URL `http://www.springerlink.com/index/10.1007/978-3-642-13062-5_13`

[89] D. Patnaik, M. Marwah, R. Sharma, N. Ramakrishnan, Sustainable Operation and Management of Data Center Chillers Using Temporal Data Mining, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009), KDD '09, ACM, New York, NY, USA, 2009, pp. 1305–1314. `doi:10.1145/1557019.1557159`.
URL `http://doi.acm.org/10.1145/1557019.1557159`

[90] A. Kusiak, Z. Song, Clustering-Based Performance Optimization of the Boiler Turbine System, IEEE Transactions on Energy Conversion 23 (2) (2008) 651–658. `doi:10.1109/TEC.2007.914183`.
URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4456516`

[91] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Automation in Construction 50 (2015) 81–90. `doi:10.1016/j.autcon.2014.12.006`.
URL `http://www.sciencedirect.com/science/article/pii/S0926580514002507`

[92] A. Capozzoli, F. Lauro, I. Khan, Fault detection analysis using data mining techniques for a cluster of smart office buildings, Expert Systems with Applications 42 (9) (2015) 4324–4338. doi:10.1016/j.eswa.2015.01.010.
URL http://www.sciencedirect.com/science/article/pii/S0957417415000251

[93] Y. Sun, T.-Y. Wu, G. Zhao, M. Guizani, Efficient Rule Engine for Smart Building Systems, IEEE Transactions on Computers 64 (6) (2015) 1658–1669. doi:10.1109/TC.2014.2345385.

[94] S. Li, J. Wen, A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform, Energy and Buildings 68 (2014) 63–71. doi:10.1016/j.enbuild.2013.08.044.
URL http://dx.doi.org/10.1016/j.enbuild.2013.08.044

[95] B. Chen, M. Sinn, J. Ploennigs, A. Schumann, Statistical Anomaly Detection in Mean and Variation of Energy Consumption, in: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR), IEEE, Stockholm, Sweden, 2014, pp. 3570–3575. doi:10.1109/ICPR.2014.614.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6977326

[96] J.-S. Chou, A. S. Telaga, Real-time detection of anomalous power consumption, Renewable and Sustainable Energy Reviews 33 (C) (2014) 400–411. doi:10.1016/j.rser.2014.01.088.
URL http://dx.doi.org/10.1016/j.rser.2014.01.088

[97] D. Wijayasekara, O. Linda, M. Manic, C. Rieger, Mining Building Energy Management System Data Using Fuzzy Anomaly Detection and Linguistic Descriptions, IEEE Transactions on Industrial Informatics 10 (3) (2014) 1829–1840. doi:10.1109/TII.2014.2328291.

[98] M. Le Cam, R. Zmeureanu, A. Daoud, Application of Data Mining techniques for energy modeling of HVAC sub-systems, in: Proceedings of eSim 2014 - IBPSA-Canada Conference, IBPSA, Ottawa, Canada, 2014.
URL http://www.ibpsa.org/proceedings/eSimPapers/2014/7A.3.pdf

[99] J. Ploennigs, B. Chen, A. Schumann, N. Brady, Exploiting Generalized Additive Models for Diagnosing Abnormal Energy Use in Buildings, in: Proceedings of the 5th ACM Workshop On Embedded Systems For Energy-Efficient Buildings (Buildsys '13), Roma, Italy, 2013.
URL http://dl.acm.org/citation.cfm?id=2528291

[100] H. Janetzko, F. Stoffel, S. Mittelstdt, D. A. Keim, Anomaly detection for visual analytics of power consumption data, Computers & Graphics 38 (2013) 27–37. doi:10.1016/j.cag.2013.10.006.
URL http://dx.doi.org/10.1016/j.cag.2013.10.006

[101] R. Fontugne, N. Tremblay, P. Borgnat, P. Flandrin, H. Esaki, Mining Anomalous Electricity Consumption Using Ensemble Empirical Mode Decomposition, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, Vancouver, BC, Canada, 2013, pp. 5238–5242.
URL http://hiroshi1.hongo.wide.ad.jp/hiroshi/papers/2013/romain_ICASSP2013.pdf

[102] R. Fontugne, J. Ortiz, N. Tremblay, P. Borgnat, P. Flandrin, K. Fukuda, D. Culler, H. Esaki, Strip, bind, and search: a method for identifying abnormal energy consumption in buildings, in: Proceedings of the 12th ACM/IEEE Int Conference on Information Processing in Sensor Network (ISPN 2013), ACM, Philadelphia, PA, USA, 2013, pp. 129–140.
URL http://dl.acm.org/citation.cfm?id=2461399

[103] O. Linda, D. Wijayasekara, M. Manic, C. Rieger, Computational intelligence based anomaly detection for Building Energy Management Systems, in: Proceedings of the 5th International Symposium on Resilient Control Systems (ISRCS 2012), Salt Lake City, UT, USA, 2012, pp. 77–82. doi:10.1109/ISRCS.2012.6309297.
URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6309297&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Title%3A%22Computational+Intelligence+based+Anomaly+Detection+for+Building+Energy+Management+Systems%22%29

[104] Y. Zhu, X. Jin, Z. Du, Fault diagnosis for sensors in air handling unit based on neural network preprocessed by wavelet and fractal, Energy and Buildings 44 (2012) 7–16. doi:10.1016/j.enbuild.2011.09.043.
URL http://dx.doi.org/10.1016/j.enbuild.2011.09.043

[105] M. Wrinch, T. H. El-Fouly, S. Wong, Anomaly detection of building systems using energy demand frequency domain analysis, in: Power and Energy Society General Meeting, 2012 IEEE, IEEE, San Diego, CA, USA, 2012, pp. 1–6.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6344790

[106] Z. J. Yu, F. Haghighat, B. C. M. Fung, L. Zhou, A novel methodology for knowledge discovery through mining associations between building operational data, Energy and Buildings 47 (2012) 430–440. doi:10.1016/j.enbuild.2011.12.018.
URL http://dx.doi.org/10.1016/j.enbuild.2011.12.018

[107] D. Liu, Q. Chen, K. Mori, Y. Kida, A method for detecting abnormal electricity energy consumption in buildings, Journal Of Computational Information Systems 14 (2010) 4887–4895.
URL http://www.jofcis.com/publishedpapers/2010_6_14_4887_4895.pdf

[108] D. Jacob, S. Dietz, S. Komhard, C. Neumann, S. Herkel, Black-box models for fault detection and performance monitoring of buildings, Journal of Building Performance Simulation 3 (1) (2010) 53–62. doi:10.1080/19401490903414454.
URL http://www.tandfonline.com/doi/abs/10.1080/19401490903414454

[109] S. Wang, Q. Zhou, F. Xiao, A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults, Energy and Buildings 42 (4) (2010) 477–490. doi:10.1016/j.enbuild.2009.10.017.
URL http://dx.doi.org/10.1016/j.enbuild.2009.10.017

[110] C. Forlines, K. Wittenburg, Wakame: Sense Making of Multi-dimensional Spatial-temporal Data, in: Proceedings of the International Conference on Advanced Visual Interfaces (AVI 2010), AVI '10, ACM, Roma, Italy, 2010, pp. 33–40. doi:10.1145/1842993.1843000.
URL http://doi.acm.org/10.1145/1842993.1843000

[111] K. Yoshida, M. Inui, T. Yairi, K. Machida, M. Shioya, Y. Masukawa, Identification of Causal Variables for Building Energy Fault Detection by Semi-supervised LDA and Decision Boundary Analysis, in: Data Mining Workshops, 2008. ICDMW '08. IEEE International Conference on, Pisa, Italy, 2008, pp. 164–173. doi:10.1109/ICDM.W.o2r0k0s8h.o4p4s.2008.46.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4733934

[112] J. E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energy and Buildings 39 (1) (2006) 52–58. doi:10.1016/j.enbuild.2006.03.033.
URL http://linkinghub.elsevier.com/retrieve/pii/S0378778806001514

[113] S. Wang, J. Cui, Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method, Applied Energy.
URL http://www.sciencedirect.com/science/article/pii/S0306261904001953