International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 6; Issue 12; December 2017; Page No. 8438-8442
DOI: http://dx.doi.org/10.24327/ijcar.2017.8442.1361

Research Article

IMPACTS OF TEMPORAL RESOLUTION EXPLOITATION HAND-PICKED BUNCH WAYS ON RESIDENTIAL ELECTRICITY LOAD PROFILES

Kavitha.P.M*., Usha S and R.Augusthiyar

Department of Computer Applications, SRM Institute of Science & Technology, Chennai-600 089 TN, India

ARTICLE INFO

Article History:
Received 6th September, 2017
Received in revised form 25th October, 2017
Accepted 4th November, 2017
Published online 28th December, 2017

Key words:

Final Clustering, Support Vector, Target

ABSTRACT

This paper presents an application of support vector clustering (SVC) to electrical load classification. The SVC approach includes Gaussian kernel and clustering algorithm to exploit the location of the bounded support vectors (BSVs) to define the outliers, identifying the clusters in function of the distance of the non-BSVs to the BSVs. Its implementation is comparatively less computational behavior and the single user defined threshold is used. Extended comparison to other clustering methods is included to show the effectiveness of the proposed approach in grouping multidimensional load pattern data into non-overlapping clusters. The successive task is to identify the outliers and clustering few details.

INTRODUCTION

At the present stage of evolution of the energy systems and markets, there is a specific and growing interest towards load pattern classification, especially from the point of view of the distribution suppliers. The emerging trend towards unbundling the distribution and supply services, as well as new degrees of freedom enabled by the legislation for setting up dedicated tariff offers for specific customer groups, are providing new opportunities to the supply operators.

Furthermore, the introduction of advanced metering technologies allows the supply operators for receiving several data from the field, to be analyzed through dedicated algorithms. The scopes themselves of load pattern classification have become manifold. Some key aspects include the customer classification for load profiling on the basis of representative load patterns for tariff purposes or for analyzing the customer interactions with the markets within demand response programs, the classification of the load pattern data of individual customers belonging to different periods of the year, and the detection of the outliers exhibiting anomalous behavior, such as the ones affected by various forms of nontechnical losses.

This paper addresses the classification procedures aimed at assisting the formation of load profiles for tariff purposes. A widely used approach for classifying electricity customers is based on analyzing the shape of the electrical consumption.

Corresponding author:* **Kavitha.P.M
Department of Computer Applications, SRM Institute of Science & Technology, Chennai-600 089 TN, India

The structure of a comprehensive approach to electricity customer classification has been illustrated.

The initial data are given by the representative load patterns (RLPs), constructed by averaging the load data monitored during a period of observation in a specific loading condition (e.g., a weekday or weekend day in a given season) after having detected and eliminated possible bad data. The approach consists of successive stages performing feature selection, customer classification by means of suitable clustering techniques, clustering validity assessment by using properly defined metrics and indicators, and finally load profile formation on the basis of the load patterns aggregated in the same customer class [1].

The core of the classification procedure is the clustering algorithm. Various methods have been proposed and tested on different types of load pattern data, among which the modified follow-the-leader, k-means, fuzzy k-means, hierarchical clustering, self organizing maps, and other statistical and neural methods. Extended comparisons have been run to identify the most suitable techniques, using various clustering validity measures to rank the methods applied to different numbers of clusters.

Robustness of the ranking has been tested by assessing the persistence of a specific method to provide the best outcomes for different numbers of clusters. Alternative classification methods can be validated and tested according to the same procedure. This paper illustrates and discusses the use of support vector clustering (SVC) for electrical load pattern classification. Besides being the first application of SVC to the electrical load pattern classification, the method presented in this paper is not a straightforward testing of an existing technique, but contains an original development that makes part of the procedure easy to be applied to a large number of multidimensional data.

SVC is an unsupervised learning method and a nonparametric clustering algorithm. The performance of the SVC method depends on the selection of a suitable kernel function. Given a data set, the SVC application consists of two steps, namely, the creation of the support vectors, followed by cluster labeling. The load pattern classification presented in this paper is a significant example of SVC application to a multidimensional data set. For this purpose, after the first step of creation of the support vectors, a specifically developed deterministic algorithm for cluster identification is introduced. This algorithm exploits the characteristics of the support vectors to provide a simple and effective procedure that completes the SVC calculations [2].

With this technique, it has been possible to obtain a performance comparable or better than the one of the other methods analyzed, especially for relatively small number of clusters. This result is of particular relevance for the suppliers at the distribution side, interested in identifying the outliers and in defining a relatively small number of customer classes to be subject to different tariff offers.

Related Work

[1] T. Fu, "A review on time series data mining,": Time series is an important class of temporal data objects and it can be easily obtained from scientific and financial applications. A time series is a collection of observations made chronologically. The nature of time series data includes: large in data size, high dimensionality and necessary to update continuously. Moreover time series data, which is characterized by its numerical and continuous nature, is always considered as a whole instead of individual numerical field. The increasing use of time series data has initiated a great deal of research and development attempts in the field of data mining. The abundant research on time series data mining in the last decade could hamper the entry of interested researchers, due to its complexity. In this paper, a comprehensive revision on the existing time series data mining research is given. They are generally categorized into representation and indexing, similarity measure, segmentation, visualization and mining. Moreover state-of-the-art research issues are also highlighted. The primary objective of this paper is to serve as a glossary for interested researchers to have an overall picture on the current time series data mining development and identify their potential research direction to further investigation.[2] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer,": This paper deals with the classification of electricity customers on the basis of their electrical behavior. Starting from an extensive field measurement-based database of customer daily load diagrams, the authors searched for the most appropriate indices or sets of indices to be used for customer classification. They propose two original measures to quantify the degree of adequacy of each index. Using the indices as distinguishing features, they adopt an automatic clustering algorithm to form customer classes. Each customer class is then represented by its load profile. They use the load profiles to study the margins left to a distribution company for fixing dedicated tariffs to each customer class. They take into account new degrees of freedom available in the competitive electricity markets, which increase flexibility in the tariff definition under imposed revenue caps. Results of a case study performed on a set of customers of a large distribution company are presented.[3] C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L. Filipova-Neumann, "Cluster analysis of smart metering data-An implementation in practice,": Utilities and electricity retailers can benefit from the introduction of smart meter technology through process and service innovation. In order to offer customer specific services, smart meter mass data has to be analyzed. In the article we show how to integrate cluster analysis in a business Intelligence environment and apply cluster analysis to real smart meter data to identify detailed customer clusters. The broad roll-out of smart meters poses significant challenges to electricity utilities. On the one hand, large investments in the metering infrastructure are required and on the other hand traditional business processes need to be redesigned. The processing and utilization of customer consumption data from smart meters may enable electricity companies to achieve large efficiency gains. In our work we have presented an implementation and evaluation of a cluster analysis approach for smart meter data within a business intelligence environment. The identified clusters are plausible and yield a true information gain for the utility. Given a direct integration with existing IT systems (e.g., ERP or CRM), this approach may be relevant for metering service companies as well as utilities. Based on the cluster analysis, metering service companies can offer innovative service products like energy management planning or regional load profiles. Suppliers can profit from the possibility of designing segment-specific rates which allow a better integration of the demand side into the control of the electricity system. Our electricity load data analysis gives rise to subsequent research questions. We will need to validate the robustness of the identified clusters using a broader data set. Moreover, the integration of additional data sources, such as current rate information, industry code, or household properties such as demographic and socio-economic data can further support effective clustering. There still is an insufficient understanding how the willingness of households to adapt their load behavior corresponds with their socioeconomic properties. As this is a central question for future demand-centered control paradigms, research needs to focus on questions like assessing demand elasticity and load shifting potentials.[4] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification,": The recent evolution of the electricity business regulation has given new possibilities to the electricity providers for formulating dedicated tariff offers. A key aspect for building specific tariff structures is the identification of the consumption patterns of the customers, in order to form specific customer classes containing customers exhibiting similar patterns. This paper illustrates and compares the results obtained by using various unsupervised clustering algorithms (modified follow-the-leader, hierarchical clustering, K-means, fuzzy K-means) and the self-organizing maps to group together customers with similar electrical behavior. Furthermore, this paper discusses and compares various techniques-Sammon map, principal component analysis (PCA), and curvilinear component analysis (CCA)-able to reduce the size of the clustering input data set, in order to allow for storing a relatively small amount of data in the database of the distribution service provider for customer classification purposes. The effectiveness of the classifications

obtained with the algorithms tested is compared in terms of a set of clustering validity indicators. Results obtained on a set of nonresidential customers are presented.[5] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping,": In the current structure of the electricity business, distribution and supply services have been unbundled in many jurisdictions. As a consequence of unbundling, electricity supply to customers is now provided on a competitive basis. In this context, the electricity suppliers need to get accurate information on the actual behavior of their customers for setting up dedicated commercial offers. Customer grouping on the basis of consumption pattern similarity is likely to provide effective results. This paper provides an overview of the clustering techniques used to establish suitable customer grouping, included in a general scheme for analyzing electrical load pattern data. The characteristics of the various stages of the customer grouping procedure are illustrated and discussed, providing links to relevant literature references. The specific aspect of assessing the performance of the clustering algorithms for load pattern grouping is then addressed, showing how the parameters used to formulate different clustering methods impact on the clustering validity indicators. It emerges that the clustering methods able to isolate the outliers exhibit the best performance. The implications of this result on the use of the clustering methods for electrical load pattern grouping from the operator's point of view are discussed.[6] S. Verdu, M. Garcia, C. Senabre, A. Marin, and F. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps,": Different methodologies are available for clustering purposes. The objective of this paper is to review the capacity of some of them and specifically to test the ability of self-organizing maps (SOMs) to filter, classify, and extract patterns from distributor, commercialize, or customer electrical demand databases. These market participants can achieve an interesting benefit through the knowledge of these patterns, for example, to evaluate the potential for distributed generation, energy efficiency, and demand-side response policies (market analysis). For simplicity, customer classification techniques usually used the historic load curves of each user. The first step in the methodology presented in this paper is anomalous data filtering: holidays, maintenance, and wrong measurements must be removed from the database. Subsequently, two different treatments (frequency and time domain) of demand data were tested to feed SOM maps and evaluate the advantages of each approach. Finally, the ability of SOM to classify new customers in different clusters is also examined. Both steps have been performed through a well-known technique: SOM maps. The results clearly show the suitability of this approach to improve data management and to easily find coherent clusters between electrical users, accounting for relevant information about weekend demand patterns.

Proposed Work

Architecture

Third-party data is information that's collected by an entity that doesn't have a direct relationship with consumers. It's basically anything that isn't first-party data, as explained above. For example, a third-party data provider might pay publishers to let it collect information about their visitors, and use it to piece together detailed profiles about users' tastes and behaviors as they move around the Web. This information can then be sold to advertisers to help them target their ad buys.

First party data is loosely defined as information you yourself have collected about your audience. (The "first party" is you.) In the context of display advertising, first party data is most often cookie-based data, and it can include information gathered from website analytics platforms, CRM systems, and business analysis tools.In general, first party data is the most valuable data you can collect about your audience, and it becomes a powerful resource when tied to display ad campaign design. Because first party data provides specifics about your already-existing users and customers, it's the key component of Site retargeting, Face book retargeting, and retargeting. With first party data, you can target returning customers by leveraging information that you already have about their past purchases and product interests. This strategy is a major element in Amazon's astonishing success: their personalized recommendations of books and other products are a perfect first party data example. Overall, the more dynamic and personalized the ad, the better the chances for conversion. First party data is always the most useful and valuable, but eventually you're likely to find yourself in a position where you want to reach an audience that you don't have first-hand information about. This is where second party and third party data become useful.

Second party data is essentially somebody else's first party data. Second party data isn't usually commoditized, but you can often work out an arrangement with trusted partners who are willing to share their customer data with you (and vice versa). For instance, a high-end watch company might partner with a yacht blog to find new customers, based on demographic overlap. Second party also plays a large role in audience extension and audience targeting. The possibilities are endless, and the key is to seek out, form, and maintain mutually beneficial partnerships.

Companies such as BlueKai, Peer39, and eXelate sell third party data. These companies are also known as DMPs or Data Aggregators. They're the behemoths of the data world. The data these large companies provide is typically purchased on a large scale from publishers. The benefit of third party data is the sheer volume of user data you can access. However, this data is also widely accessible to competitors, so you aren't gaining unique audience intelligence when you tap into third party data resources. Third party data is great for demographic, behavioral, and contextual targeting, and can be used to remove boot traffic. It also plays a critical role in solutions like audience targeting and audience extension. Third party data providers charge a fee to use their segments, usually on a CPM basis. Depending on the data, the CPM can be anywhere from $0.50 to $5.00.

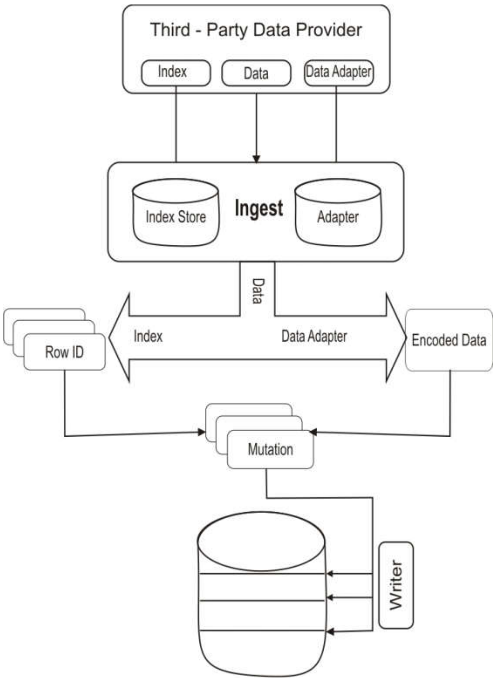

Fig 1 System Architecture

Details of Work

Data analysis requires a lot of computing and complex time for result and understanding. Big data contain many techniques for analysis consists of a lot of computation which is done using big data algorithms. There are some algorithms like cluster, map-reduce, data mining algorithms such as k-means, classification methods, support vector machine, apriori algorithm, EM, page rank, etc.. All these algorithms are effectively working according to their need. For large growth of data, the data analytics uses advanced analytic techniques like predictive analytics, data mining, statistical analysis, complex SQL, data virtualization, and artificial intelligence. ADV (advanced data virtualization) is the best fit for the growing big data analytics. BI supports real time dashboard and key performance indicators (KPI) and sometimes OLAP cube for which in-memory databases will move. Text mining and text analytics give the unstructured data more efficiently. There are two specified modules surpassingly used in this system to produce effective result.1) Determination of the support vectors 2) Formation of Final Cluster.

Determination of the Support Vectors

The data set is defined by (generally multidimensional) data points belonging to the original data space. Conceptually, these points are mapped from the data space to a high-dimensional feature space according to a nonlinear transformation defined on the basis of a user-specified kernel function, depending on some notion of distance between the input data. Then, an optimization procedure is set up to minimize the radius of the sphere enclosing the image of the data points mapped onto the space. The details of the mathematical development of the SVC method are illustrated and are not repeated here for the sake of brevity. Indications are provided in the sequel on the final formulation, defining the relevant variables and parameters. The optimization problem is turned into the maximization of the Wolfe dual form of the Lagrangian.

Formation of the Final Clusters

The results of the calculations carried out at the first stage of the SVC procedure are the values of the radius for and a number of SVs and BSVs that correspond to specific points in the original data space. In particular, the SVs can be interpreted as points located at the cluster boundaries, and the BSVs as outliers. This property can be interesting in specific applications in which it is significant to identify specific points delimiting the cluster regions, especially for non-overlapping groups of data.

However, this information is not sufficient to form the final clusters, since the calculated distances indicate only amplitudes but not directions concerning the data transformed in the space. In other terms, the SV formation provides no information concerning neither possible cluster identification nor the related centroids.

In practice, there can be many points with similar radius, but this fact does not imply that they belong to the same cluster, since the corresponding images in the transformed data space can be far from each other, even though their distances from the (unique) center of the sphere in the space are similar. Indeed, none of the algorithms used for computing the support vectors is able to directly provide the partitioning of the original data into clusters.

A post-processing phase is then needed to form the clusters on the basis of the support vectors and of the interpretation of their properties. The classical post-processing technique adopts a geometric approach based on the distances and operating in the transformed space to form an adjacency matrix between each pair of points in the original data space. This adjacency matrix contains only binary values, and the clusters are found by searching for the connected components on the graph induced by the adjacency matrix.

Several methods can be applied to build the adjacency matrix. However, in the presence of several high-dimensional and partially overlapped data (as in the case of the electrical load patterns), the creation of the adjacency matrix becomes computationally intensive, and the interpretation of its induced graph could become quite complicated.

Another method based on the application of fuzzy techniques has been proposed in which the parameters of the method are set up by using heuristic methods specifically developed for assisting the parameter learning process. A novel and computationally simpler method dedicated to the cluster formation after having obtained the support vectors.

Experimental Setup

The results of clustering depend on both the algorithm and the resolution of the data. The main aim of this paper is not to compare the performance of the algorithms, but to compare the clustering results when the data resolution is varied. Thus, three popular types of algorithm were selected: A partitioning algorithm, k-means is one of the most common methods. From an initial partitioning, a converging process in which data elements are moved from one group to another is carried out until stable partitions are achieved. The convergence of the algorithm depends on the initial partitioning. Therefore such algorithms must be run several times with different initializations. Agglomerative hierarchical algorithms. These bottom-up algorithms create a new cluster for each one of the data elements then successively merge the closest subgroups until the specified number of clusters is achieved. There are different variations depending on the criterion used to compute the distance to merge cluster.

1. Subgroup Description Language: A subgroup is expressed as a conjunction of conditions and can be thought of as an induced rule. A condition is a test on a variable (or in data mining terms, an attribute), for example "" or".'' The number of conditions, indicating the complexity of a subgroup, is 1 in these examples. The conjunction "AND" has two conditions joined by AND. A selector is a test on a variable. Two features "Location" and "Age" are tested here and the complexity of the subgroups so defined is two.
2. Target: This is a variable whose patterns interest users and on which a quality measure assesses subgroups.
3. Search Strategy: The simple brute-force search method exhaustively tests all combinations of selectors. It is practically prohibitive because when the number of selectors increases, the search space increases exponentially. On the other hand, the beam search method, in any iteration, only expands a fixed number of the most promising selectors or conjunctions of selectors found so far, by appending additional selectors. This fixed number is called beam width.
4. Quality measure: A quality measure evaluates the quality of candidate subgroups. Most quality measures are designed to obtain a balance between the two objective values

CONCLUSION

In this project, the work analyses the impact of the temporal resolution when clustering electricity load profiles. Several algorithms have been systematically tested by changing the resolution of the input data (of real household consumption). The results are evaluated with internal and external validity measures, and the efficiency was computed using a large-scale synthetic data set.

References

1. T. Fu, "A review on time series data mining," *Eng. Appl. Artif. Intell.,* vol. 24, no. 1, pp. 164-181, 2011.
2. G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," IEEE Trans. *Power Syst.,* vol. 18, no. 1, pp. 381-387, Feb. 2003.
3. C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L. Filipova-Neumann, "Cluster analysis of smart metering data-An implementation in practice," *Business Inf. Syst. Eng.,* vol. 4, no. 1, pp. 31-39, 2012.
4. G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," IEEE Trans. *Power Syst.,* vol. 21, no. 2, pp. 933-940, May 2006.
5. G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68-80, 2012.
6. S. Verdu, M. Garcia, C. Senabre, A. Marin, and F. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Trans. Power Syst*., vol. 21, no. 4, pp. 1672-1682, Nov. 2006.
7. Noriaki Nishio, Atsuko Mutoh, and Nobuhiro Inuzuka "On Computing Minimal Generators in Multi-Relational Data Mining with respect to θ-Subsumption ", IEEE 2012.
8. Albert Bifet Yahoo! Research Barcelona Avinguda Diagonal 177, 8th floor Barcelona, 08018, Catalonia "Mining Big Data in Real Time", IEEE 2012
9. Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam " A Study of Data Mining Tools in Knowledge Discovery Process", *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307, Volume-2, Issue-3, July 2012

How to cite this article:

Kavitha.P.M., Usha S and Augsthiyar (2017) 'Impacts of Temporal Resolution Exploitation Hand-Picked Bunch Ways on Residential Electricity Load Profiles ', *International Journal of Current Advanced Research,* 06(12), pp. 8438-8442.
DOI: http://dx.doi.org/10.24327/ijcar.2017.8442.1361

****