International Conference on Advances in Materials, Computing and Communication Technologies
In Association withInternational Journal of Scientific Research in Science and Technology
Volume 9 | Issue 1 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Study and Grouping of Suburban Consumers Energy Behavioural Demand Using Smart Meter Information

Sathyanarayanan. P[1], Jayapiriya. A[2], Nanthini. K[2], Revathi[2], Gowri V[2]
[1]Assistant Professor, Department of EEE, Narasu's Sarathy Institute of Technology, Tamil Nadu, India
[2]Students, Department of EEE, Narasu's Sarathy Institute of Technology, Tamil Nadu, India

ABSTRACT

The main goal of this research is to discover the structure of home appliances usage patterns, hence providing more intelligence in smart metering systems by taking into account the usage of selected home appliances and the time of their usage. In particular, we present and apply a set of unsupervised machine learning techniques to reveal specific usage patterns observed at an individual household. The work delivers the solutions applicable in smart metring systems that might: (1) contribute to higher energy awareness; (2) support accurate usage forecasting; and (3) provide the input for demand response systems in homes with timely energy saving recommendations for users. The results provided in this paper show that determining household characteristics from smart meter data is feasible and allows for quickly grasping general trends in data.

Keywords—Data Mining; Users' Behaviors; Smart Metering; Smart Home; Energy Usage Patterns

I. INTRODUCTION

Smart metering systems are key components for creating environmental sustainability by managingenergy at homes. They are supposed to play an important role in reducing overall energy consumptionand increasing energy awareness of the users through being better informed about consumption patterns.Smart meter data is being analyzed through publicly availabledata sets such as the Irish smart meter trial andother research projects. Clustering customers basedon attributes derived from smart meter data creates betterunderstanding of the different types of energy behavioralgroups. The focus of this paper is to derive and clustersuitable attributes from smart meter data which can assista DNO with two main applications for better LV networkmodeling and management.

The primary application thatmotivates our approach is to help a DNO identify suitable customersfor energy management solutions such as demand-sideresponse (DSR) and through storage devices. These applicationshave already been considered recently. By choosing the correct attributes, a DNO can use the clusteringto identify suitable customer groups for demand reductionsolutions and hence, help to reduce network demand andvolatility. For example, customers with heavy but regulardemand in the evening time period could be ideal candidatesfor peak demand reduction through storage devices, whereasthose with irregular demand may be more suitable for DSR.

The secondary application is for identifying links betweenenergy behavioral usage and publically available information. Suchlinks can be used to improve modeling of residential customerdemand and reduce necessary monitoring on the LV network.However, besides these two, the methodology can also be usedfor further applications of smart meter-based.Choosing the correct attributes is potentially the mostimportant aspect of a successful clustering. This is particularlychallenging when considering extremely volatilehousehold-level demand which is much less regular thanhigher voltage demands [6]. The number of attributes shouldbe optimized to: ensure the data distribution over parameterspace is dense, reduce computational costs, and ensurethe results can be easily interpreted. In addition, if thenumber of attributes is not optimized then the clustering isless likely to be representative and fit for purpose. A maincontribution of this paper is a detailed analysis of a largeamount of domestic smart meter data, especially with regardto better understanding of the peak demands and sources ofvariability. This helps us to identify and minimize the importantattributes to be used in the clustering. In addition, wediscover four key time periods within which data should beanalyzed since they describe the most frequent largest demand. The possible means that can support energy and, in general, resource conservation at individual homeshave been the object of the studies in social sciences and in engineering sciences since the 1970s and arestill conducted nowadays [3]. It has been identified that the provision of feedback about energy usage isone of the most effective strategies for conservation [3,6].

The other strategies include the provision ofinformation about energy conservation, goal setting to induce energy efficiency and conservationactions, and the reward of savings in monetary terms.Recent years, with the development of smart metering solutions, have seen an increase in the numberof tools that individual users can employ to monitor their energy consumption. Most of the tools simplyprovide access to raw usage data including, for instance, readouts of the watt hours consumed by ahousehold or by a particular home appliance, and calculate the estimated costs. Previous researches haveshown that users seek solutions that can provide greater insight into energy usage and its impact as theydesire more real-time information to help them save money, and keep the homes comfortable andenvironmentally friendly [1-3].

In light of the cited works, the mechanisms that enable a user to linkthe activities and energy consumption by attaching contextual labels to energy events are a promisingstep to support energy conservation, however solutions for collecting annotations from users can be errorprone or too intrusive. Nevertheless, there are successful applications for collecting annotations as arepresentation of electricity consumption data, and therefore make sense of past energy usage,as reported in [2].

Many research projects are involved in development of user-friendly and convenient feedback toolsfor visualization of electricity consumption providing instantaneous consumption data, often throughsuggesting ambient displays aiming at emotional reactions. It can be found that both commerciallyand freely available resources use feedback tools including point of consumption devices such as Kill aWatt electricity usage monitors, information dashboards, analysis interfaces, and online profiling andvisualization tools such as Microsoft Hohm, PowerMeter by Google, ODEnergy, OPOWER or AlertMe.These tools offer precise quantitative measures of energy expenditures, historical and predictive chartingfacilities, cost breakdowns, and performance tracking. However, they require some effort to integratethem into home infrastructure, and they lack convenient feedback on real-time resource use. This might have an impact on the decision to discontinue development of some of them due to a lack of consumeruptake (PowerMeter service was ceased in 2011 and Hohm in 2012). Nevertheless, new tools areconstantly being developed to provide more detailed energy feedback. The itemized energy consumptionfrom different appliances can be achieved by individually monitoring each of them. However, thisstrategy is expensive due to the hardware costs and complex infrastructure that may be difficult todeploy. In this context, there is a significant number of researches focused on appliance recognition based on non-intrusive appliance load monitoring approach (NIALM). It involves the use of machine learning algorithms and optimization techniques to recognize energy signatures of home devices. The challenge in NIALM is that individual appliances have very different energy signatures that are hard to distinguish unless very sensitive and high resolution meters are used. Therefore, this isan area of research which is still being thoroughly explored [1].Based on NIALM, there have been research attempts devoted to load prediction on the individual household level [6–7]. They utilize smart meter data enriched with a set of household behavioral data(patterns of home appliances usage) and dwelling characteristics to benefit significant improvement interms of the accuracy of the forecasts generated at the household level. The proposed work fits into the research stream that looks at challenges associated with causal factorsthat impact energy usage of individual appliances observed at the household level. This is to provide customer feedback on usage patterns and derive significant underlying associations between severalcontextual factors including time of use and user activities. It shows a broad set of useful insights thatmay increase awareness and understanding of home energy consumption.Smart Meter DataElectricity measurements data were prepared using Mieo HA104 meter installed in one of thehouseholds in Warsaw, Poland for the purpose of SMEPI project (SMEPI—Smart Metering Poland, aHi-Tech project to develop smart metering solutions partially financed by National Centre for Researchand Development (NCBiR) and led by Vedia S.A in cooperation with GridPocket and Faculty of AppliedMathematics and Informatics at Warsaw University of Life Sciences). The household consisted of two adult people (in their mid-40s) and two pre-teen children. The adult members of the family we reemployed full time with standard office hours. The household was situated in a flat of about 140 m2 floor area and was equipped with various home appliances including a washing machine, refrigerator, dishwasher, iron, electric oven, two TV sets, audio set, coffee maker, desk lamps, computer, and a couple of light bulbs. The data were gathered during 60 days, starting from 29 August until 27 October 2012.However, for the analysis we extracted 44 days for which we gathered a set of user behavioral information such as devices' operational characteristics at the household. These data were produced bythe reference system which was constructed to collect binary data about the ON-OFF states of the devices.

Detecting Patterns Using Sequential Association Rules: The problem of discovering sequential patterns is based on a database containing information about events that occurred within a specified period of time. The aim of the sequential association rules is tofind the relationship between the occurrences of certain events in the selected time period [7].

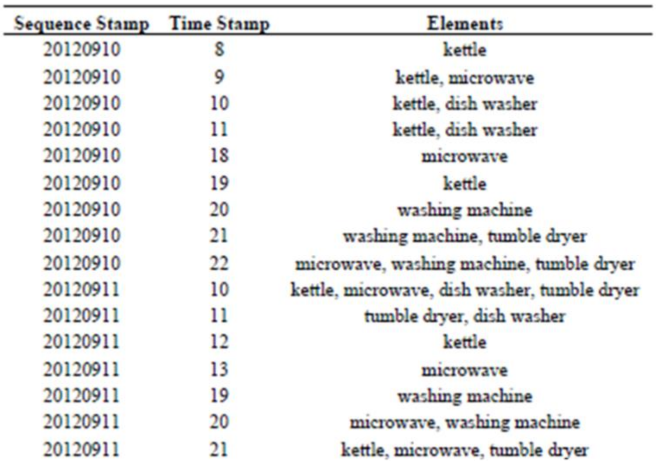
The problem of discovering frequent item sets is to find all item sets occurring in the database D withthe support higher or equal to minimum support threshold supplied by a user. An itemset with the supporthigher than minsup is called a frequent item set.The support of the rule X → Y is the ratio of the number of transactions that support both theantecedent and the consequent of the rule to the total number of transactions. The support of a ruledenotes its statistical significance. Rules with low support tend to describe relationships that are notcommon in the database. On the other hand, rules with high support are covered by many transactionsin the database and they describe common patterns.The confidence of the rule X → Y is the ratio of the number of transactions that support both theantecedent and the consequent of the rule to the number of transactions that support only the antecedentof the rule. The confidence of a rule denotes its statistical strength. High confidence indicates strongcorrelation between elements contained in the antecedent and the consequent of the rule. Low confidence denotes weak correlation between elements and may indicate purely coincidental co-occurrence of elements.Lift of the rule X → Y in the database D is called the measure of the rule correlation, indicating whatis the impact of an element X for occurrence of an element Y.

In other words, lift measures how manytimes more often X and Y occur together than expected if they where statistically independent. Lift is not downward closed and does not suffer from the rare item problem. Also, lift is susceptible to noise insmall databases. Rare item sets with low counts (low probability) which per chance occur a few times(or only once) together can produce enormous lift values.

A sequence is an ordered list of elements ▯ X▯, X▯, . . . , X▯▯ where X▯ is a set of items, ∀i X▯⊆ L.Each set X▯ is called a sequence element. The length of a sequence X is the number of sequence elements.

Each sequence element has a timestamp denoted as ts▯X▯▯. A sequence ▯ X▯, X▯, . . . , X▯▯ is containedin another sequence ▯ Y▯, Y▯, . . . , Y▯▯ if there exist integers i▯▯ i▯▯ i▯ in such thatX▯▯ Y▯▯, X▯▯ Y▯▯, … , X▯▯ Y▯▯. The sequence ▯ Y▯▯, Y▯▯, … , Y▯▯▯ is called an occurrence of X in Y.There are three main time constraints involved in sequential pattern discovery, namely, the minimumand the maximum time gap between consecutive occurrences of elements within a sequence element(called min-gap and max-gap respectively) and the size of the time window which allows for mergingitems into sequence elements, denoted as window-width [3].

The starting point for the usage patterns detection, based on the sequential association rules, was todetermine the transaction matrix. Each transaction has a time stamp indicating the occurrence of theelements in the specified sequence. In this case, we assume that a single sequence is the whole day,therefore, the tag sequence is the particular date. The time stamp is the hour at which specific devices were turned ON. Created transaction table takes into account only the binaryinformation (the appliance was turned ON or not), but does not include the number of switch ON statesin a given hour. In the analyzed period, there are theoretically 24 × 44 = 1056 transactions (the numberof hours multiplied by the number of days), whereas the used SPADE algorithm (Sequential PatternDiscovery using Equivalence classes [3]) does not include empty transaction (hours, in which none ofthe tested devices was turned ON); therefore, the final transaction table contains only 319 transactions.Given the rules with the support of more than 0.1, the minimum time difference between successiveelements in the sequence of 1 and a maximum time difference between successive elements in thesequence of 1, the following behavior patterns can be observed:with the support equal to 0.1 and with the confidence of 100%, if in a certain hour the washing machine operated, in the next hour the tumble dryer and kettle operated;with the support equal to 0.1 and with the confidence of 100%, if in a certain hour the washingmachine operated, in the next hour the washing machine and kettle operated, and in the next hourthe washing machine also operated, so did the tumble dryer and kettle;rule No. 4 with the support equal to 0.15, and with the confidence of 75% shows that theoccurrence in a sequence of such devices as kettle, dish washer and washing machine influencesthe occurrence in a sequence of such appliances as tumble dryer and kettle.With the support equal to 0.1 and with the confidence of 66%, if in a certain hour the kettleoperated, in the next hour the washing machine was turned ON, then in the next hour the washingmachine and microwave were in operation.All these observed sequential rules have lift greater than one, which means that the occurrence of the elements in the left side of the rules influence the occurrence of the elements contained on the right sideof the sequential rule.



Sequence Stamp	Time Stamp	Elements
20120910	8	kettle
20120910	9	kettle, microwave
20120910	10	kettle, dish washer
20120910	11	kettle, dish washer
20120910	18	microwave
20120910	19	kettle
20120910	20	washing machine
20120910	21	washing machine, tumble dryer
20120910	22	microwave, washing machine, tumble dryer
20120911	10	kettle, microwave, dish washer, tumble dryer
20120911	11	tumble dryer, dish washer
20120911	12	kettle
20120911	13	microwave
20120911	19	washing machine
20120911	20	microwave, washing machine
20120911	21	kettle, microwave, tumble dryer

Detecting Patterns Using Hierarchical Clustering: Hierarchical cluster analysis is an algorithmic approach to find discrete groups with varying degreesof similarity in a data set represented by a similarity matrix. These groups are hierarchically organizedas the algorithms proceed and may be presented as a dendrogram. Many of these algorithms are greedy(*i.e.*, the optimal local solution is always taken in the hope of finding an optimal global solution) andheuristic, requiring the results of cluster analysis to be evaluated for stability.Hierarchical clustering methods can be divided into agglomerative and divisive approach.Agglomerative clustering is a widespread approach to cluster analysis. Agglomerative algorithmssuccessively merge individual entities and clusters that have the highest similarity values computedusing for instance Euclidean distance.One of the most popular agglomerative clustering algorithm is Ward's method [24]. This is analternative approach for performing cluster analysis. Basically, it looks at cluster analysis as an analysisof variance problem, instead of using distance metrics or measures of association. It will start out at theleaves and work its way to the trunk, so to speak. It looks for groups of leaves that it forms into branches,the branches into limbs and eventually into the trunk. Ward's method starts out with clusters of size 1and continues until all the observations are included in one cluster.

In general, Ward's method can be defined and implemented recursively by a Lance–Williamsalgorithm. The Lance–Williams [5] algorithms are an infinite family of agglomerative hierarchicalclustering algorithms which are represented by a recursive formula for updating cluster distances interms of squared similarities at each step (each time a pair of clusters is merged).The recurrence formula allows, at each new level of the hierarchical clustering, the dissimilaritybetween the newly formed group and the rest of the groups to be computed from the dissimilarities ofthe current grouping. This approach can result in a large computational savings compared withre-computing at each step in the hierarchy from the observation-level data.The purpose of this analysis is to discover similar profiles or, in other words, appliances with similarswitch ON probability distribution through the whole day or the whole week. As a result of groupingusing Ward's method with the Euclidean distance measure, the following dendrogram was obtained as presented in Figure 2.

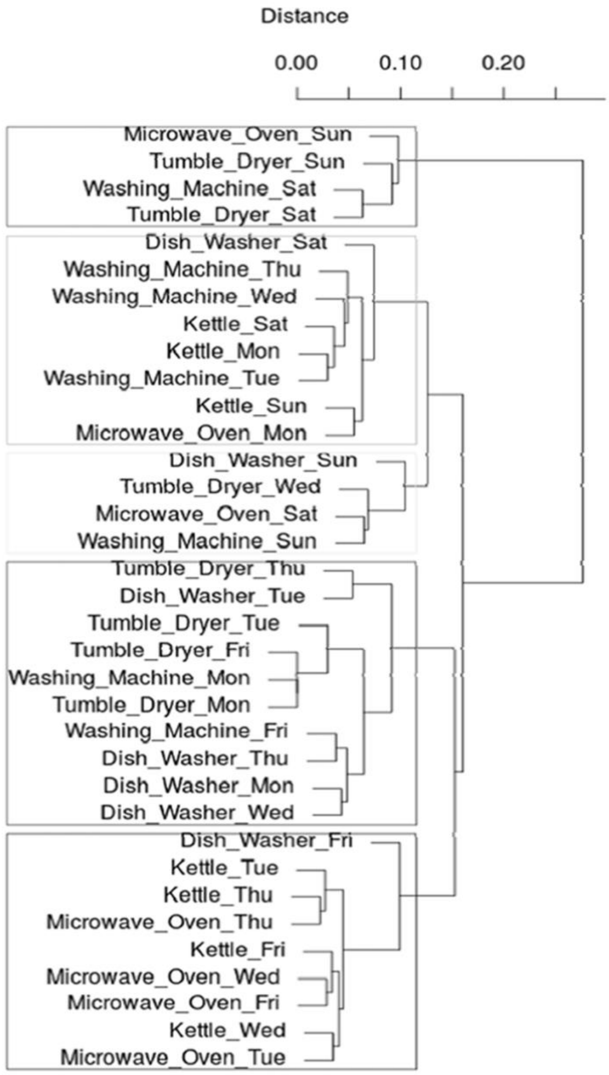
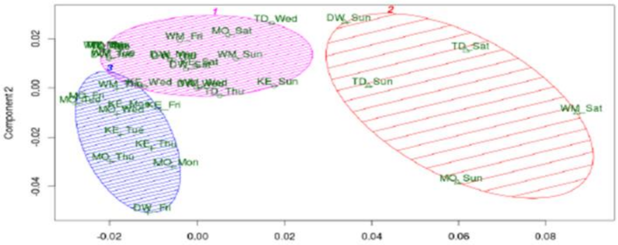


Figure 2.Dendrogram for grouping the electrical appliances throughout the whole week

Detecting patterns Using C-Means Clustering and Multidimensional Scaling: The simplest unsupervised learning algorithms that solve the well-knownclustering problem. The procedure follows a simple and easy way to classify a given data set through acertain number of clusters (assume □ clusters) fixed a priori. The main idea is to define □ centroids,one for each cluster.Clustering is the process of partitioning a group of data points into a small number of clusters.In general, we have □ data points□□, □□ 1. . . □that have to be partitioned in □ clusters. The goal is toassign a cluster to each data point. □ -means is a clustering method that aims to find the positions □□, □□ 1. . . □of the clusters that miimize the distance from the data points to the cluster. □-means clustering solves:

$$\arg\min_{s} \sum_{k=1}^{c} \sum_{x\in S_k} d(\boldsymbol{x},\mu_k) = \arg\min_{s} \sum_{k=1}^{c} \sum_{x\in S_k} \|\boldsymbol{x}-\mu_k\|_2^2$$

Unfortunately, there is no general theoretical solution to find the optimal number of clusters for anygiven data set. Although it can be proved that the procedure will always terminate and the □-meansalgorithm does not necessarily find the most optimal configuration, corresponding to the global objectivefunction minimum. A simple approach is to compare the results of multiple runs with different □ classes and choose the best one according to a given criterion, but we need to be careful because increasing □results in smaller error function values by definition, but also an increasing risk of overfitting. Thealgorithm is also significantly sensitive to the initial randomly selected cluster centers.Multidimensional scaling (MDS) [11] is a term that is applied to a class of techniques that analyses amatrix of distances or dissimilarities in order to produce a representation of the data points in areduced-dimension space. Most of the data reduction methods have analyzed the □□□ data matrix □orthe sample covariance or correlation matrix. Thus, MDS differs in the form of the data matrix on whichit operates—it is an individual-directed method. Of course given a data matrix, a dissimilarity matrixcould be constructed and then proceed with an analysis using MD techniques. However, data often arise already in the form of dissimilarities and so there is no recourse to the other techniques. Also, in other methods, the data-reducing transformation is linear. Some forms of multidimensional scaling permit anonlinear data-reducing transformation. There are many types of MDS, but all address the same basic problem: Given an □□□ matrix of dissimilarities and a distance measure find a configuration of □ points x□, …, x□ in the reduced dimension space □□□□□□□ so that the distance between a pair of points is close in some sense to thedissimilarity between the points. All methods must find the coordinates of the points and the dimensionof the space, □. Two basic types of MDS are metric and nonmetric MDS. Metric MDS assumes that thedata are quantitative and metric MDS procedures assume a functional relationship between the interpointdistances and the given dissimilarities. Nonmetric MDS assumes that the data are qualitative, havingperhaps ordinal significance and nonmetric MDS procedures produce configurations that attempt tomaintain the rank order of the dissimilarities. In our study we used one form of metric MDS, namelyclassical scaling.In general, given a set of n points in p-dimensional space, x□, … , x□, it is straightforward to calculatethe distance between each pair of points. Classical scaling (or principal coordinates analysis) is concerned with the converse problem to determine the coordinates of a set of points in a dimension □[8].Classical scaling is one particular form of metric MDS in which an objective function measuring thediscrepancy between the given dissimilarities, δ□□, and the derived distances in □□, d□□, is optimized. Thederived distances depend possible to calculate on the coordinates of the samples that we wish to find.There are many forms that the objective function may take. To find the minimum of the stress function,most implementations of MDS algorithms use standard gradient methods [2].The purpose of these computational experiment is to discover similar profile, in the same way as inthe previous case. As it was mentioned, the partitioning method divides the data into C disjoint clusters, so that objects of the same cluster are close to each other and objects of different clusters are dissimilar.The output of a partitioning method is simply a list of clusters and their objects, which may be hard tointerpret. Therefore, it would be useful to have a graphical display which describes the objects with theirinterrelations, and showing, at the same time, the clusters. Such a display was constructed using so-calledCLUSPLOT [3].For this purpose we have used the □-means algorithm, but of course also other clustering methodscan be applied. For higher-dimensional data sets a dimension reduction technique before constructingthe plot was applied. The MDS method yields components such that the firstcomponent explains as much variability as possible, the second component explains as much of the remaining variability as possible. The percentage of point variability explained by these two components (relative to all components) is listed below the plot. Then, CLUSPLOT uses the resulting partition, as well as the original data, to produce Figure 4.The ellipses are based on the average and the covariance matrix of each cluster, and their size is suchthat they contain all the points of their cluster. This explains why there is always an object on the boundary of each ellipse.



II. CONCLUSION

The worldwide adoption of smart metering systems supported by data analysis techniques leads tothe realization of dynamic tariffs, energy usage visualization, and efficient meter-to-cash billingprocesses. Nevertheless, there is a need to deliver simple and reliable tools that are intended to help consumers understand their energy usage and support their efforts in energy conservation.The simulations presented here can support development of tools that allow customers to gainimportant insights on energy consumption. For the policy makers and distribution entities, it can indicatethe direction towards provision of personalized and scalable energy efficiency programs and present aview of how the smart metering infrastructure can be enhanced in the near future. From this perspective, the results are interesting and constitute a promising step to support energy conservation. The set of clustering and association techniques helped to examine the interdependence between theusage patterns of home appliances and derive significant underlying associations between several contextual factors including time of the use and user activities. The proposed set of diversified data mining algorithms provides, in our opinion, the best way to illustrate individual patterns of energy consumption. We show that revealing household characteristics from smart meter data is feasible and offers appealing visualization of general patterns in data. We need to keep in mind that those particular techniques represent slightly different approaches to data analysis,thus they cannot be compared directly, since their evaluation may be, to a large extent, subjective andmay depend on user preferences.For future research, we see the following direction. Since the results are promising and visuallyappealing, we plan to design a larger experiment for a dozen or more households. Additionally, we aimto explore algorithmic approaches for mining usage patterns and utilize them for the purpose of energyconsumption forecasting and the development of unique, individualized energy management strategies. Additionally, since the electricity consumption of households varies over time based on the actions of individual electrical appliances operated by the members, we would like to propose the optimal structureof the data set, which takes into account the variability associated with the switch ON states of individual devices to support their accurate recognition. In future studies, special attention will also be focused onthe design of algorithms that in real time will be able to identify working states of the electrical appliances in the household. In the end, it is worth mentioning there are high expectations for combination of research onforecasting systems utilizing non-intrusive appliance recognition and user pattern behavior withmulti-agent systems

III. REFERENCES

[1]. E. Bitar, R. Rajagopal, P. Khargonekar, K. Poolla, and P. Varaiya. Bringing wind energy to market.IEEE Transactionson Power Systems, 2011.

[2]. T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning: data mining, inference, and prediction.Springer series in statistics.Springer, 2009.

[3]. S. Houde, A. Todd, A. Sudarshan, J. Flora, and K. C. Armel.Real-time feedback and electricity consumption: a field experimentassessing the potential for savings and persistence. Submitted,July 2011.

[4]. A. Krioukov, C. Goebel, S. Alspaugh, Y. Chen, D. E. Culler,and R. H.Katz. Integrating renewable energy using data analyticssystems: Challenges and opportunities. IEEE Data Eng.Bull., 34(1):3–11, 2011.

[5]. M. Pedersen. Segmenting residential customers: energy andconservation behaviors.Number 7, pages 229–241, 2008.

[6]. J. Yang and J. Leskovec.Patterns of temporal variation in onlinemedia. In Proceedings of the fourth ACM internationalconference on Web search and data mining, WSDM ’11, pages177–186, New York, NY, USA, 2011. ACM.

[7]. Rollins, S.; Banerjee, N.; Choudhury, L.; Lachut, D. A system for collecting activity annotations for home energy management. Pervasive Mob.Comput.2014, 15, 153–165.

[8]. Gustafsson, A.; Gyllensward, M. The power-aware cord: energy awareness through ambient information display. In Proceedings of the CHI EA ’05, Portland, OR, USA, 2–7 April 2005;pp. 1423–1426.

[9]. Rodgers, J.; Bartram, L. Exploring ambient and artistic visualization for residential energy use feedback. IEEE Trans. Vis. Comput. Graph.2011, 17, 2489–2497.

[10]. Firth, S.; Lomas, K.; Wright, A.; Wall, R. Identifying trends in the use of domestic appliances from household electricity consumption measurements. Energy Build. 2008, 40, 926–936.