Database Establishment for Machine Learning in NILM

Po-Hsiang Lai, Mark Trayer, Sudhir Ramakrishna, and Ying Li
Samsung Telecommunications America
1301 E. Lookout Drive
Richardson, Texas 75082 USA
{s.lai, yli2, mtrayer, sramakri}@sta.samsung.com

ABSTRACT

Nonintrusive load monitoring (NILM) is a problem of identifying operating appliances and estimating their energy consumptions based on whole home electric signals. Machine learning concepts and methods have been gradually applied to tackle NILM. A key factor of enabling and advancing machine learning methods in any problem is the availability of proper databases. The Reference Energy Disaggregation Data Set (REDD) is one such initiative example for NILM. In this paper, we extend from this initiative to address three key properties: informative, diverse, and scalable for a database. These properties enable a broader range of application and research of machine learning methods in NILM. The importance of data sets consisting of single appliance, intermediate aggregate appliances, and whole home recordings in developing machine learning methods is also discussed.

Keywords

Nonintrusive load monitoring, load disaggregation, smart metering, machine learning.

1. INTRODUCTION

Nonintrusive load monitoring (NILM) is one of the fundamental problems in energy and sustainability research. NILM is to identify appliances being used and to estimate their power consumptions based on current and voltage measurements of the main circuit entering a house. NILM technology can be used to provide valuable information for optimizing appliance usages and reducing energy consumptions and costs. NILM technology also has the potential to detect malfunctioning or inefficient appliances. In recent years, a number of machine learning methods have been proposed for NILM with encouraging results [1,2,4,5,6,8,9,10,12]. As mentioned in [7], to enable more participation in NILM research, availability of data is of critical importance. Thus the Reference Energy Disaggregation Data Set (REDD) was established and shared publicly [7]. The REDD consists of high frequency, 15kHz, whole home current and voltage data, low frequency power data of individual circuits and plugs sampled at an order around 1 Hz. Focusing on the generalizing capability of supervised learning methods for disaggregation, REDD, at the publication date of this paper, includes data from 10 homes, and multiple appliances in the same category. Despite the low frequency data set of individual appliances in REDD spanning over days enabling methods based on long duration changes such as Hidden Markov Models (HMM) based methods, it is not as useful for methods based on high frequency short term appliance characteristics, such as microscopic methods [12]. In this paper, we address database establishment considerations for enabling application and research of a larger variety of machine learning concepts, methods, and evaluation criteria in NILM. A database is an organized collection of data sets, with additional functionalities such as sort, search, format conversion, and management of adding, editing, and accessing data by users. An example of a large machine learning database is the University of California Irvine Machine Learning Repository [11]. The properties discussed in the following apply to the database itself as well as data sets or individual recordings in it. As NILM data recording is generally costly and a diverse range of data is necessary, it is beneficial to have a collective effort that different institutions or individuals contribute data sets with different properties, such that these data sets emerge gradually as a comprehensive database.

2. MACHINE LEARNING OBJECTIVES IN NILM AND DATABASE PROPERTIES

The key NILM objectives are to classify operating appliances and to estimate individual appliance power consumption using whole home electricity signals. Classification and estimation are both supervised learning problems, where a labeled *training set* is used for tuning a machine learning method, while the tuned method has to perform well on the *test set* which is new to the method. Note that test sets should be used only for method evaluation, not for development or optimization purposes. In order to optimize a method while avoiding over fitting, it is useful to maintain another set of data called a *validation set* which is independent of both training sets and test sets [3]. The purpose of a validation set is to select methods which have the best predictive and generalizing capabilities in the sense that their performances on the training set and the validation set are consistent. In many classical supervised learning problems, the training set and the test set are assumed to have the same, or nearly the same, statistical properties. However, this is in general not the case in NILM problems, where a machine learning method has to be robust against unknown appliances, various operation modes, different manufactures, operation conditions, malfunctioning appliances, and different combinations of aggregate appliances. Thus, a suitable database for NILM has to enable researchers to easily partition or generate training, validation, and test sets under various scenarios for development and evaluation of robustness of a method. In addition, new appliances are continuously developed and adopted by consumers. New data has to be easily incorporated into the database for developing new methods and evaluating existing ones. On the other hand, the database should also enable the applications and innovations of various methods based on different features.

We introduce three database properties for further enabling machine learning in NILM, informative, diverse, and scalable.

Informative: This property focuses on availability of information in the recordings of a database, such that a database is more informative if the data it contains can serve as training sets and test sets for more methods. Note that a data set is informative as a test set is not necessarily suitable as a training set, and *vice versa*. For example, the 15kHz whole home data in the REDD is suitable as a test set to evaluate high frequency methods. However, as the necessary high frequency information of each appliance is neither directly available nor is easily derived from it accurately, training microscopic methods [12] requires additional data sets. On the other hand, a data set of one minute high frequency recordings of individual appliances turning on is informative for training and validating transient based methods in single appliance cases. By synthesizing data to mimic aggregate appliance scenarios, it is also informative for validating methods under scenarios where multiple appliances are turned on within a short period of time. However it is clearly not informative for evaluating any methods or products realistically, which requires long duration whole home data sets.

Thus, data recorded at a high sampling rate and a long recording duration is essential to enable more methods based on different features across temporal resolutions and scales. In addition, detailed information about the appliances and recording environments are also important. Such information enables one to develop methods which are more robust or are able to predict how signatures and power consumptions change under new operation conditions with perhaps unknown appliances.

Diverse: The performance of a machine learning method is strongly influenced by within category differences and across category differences. A database has to be diverse enough to enable machine learning methods to capture these differences. At the single appliance recoding level, diversity of a database is characterized as the number of appliance categories included and variations within each category, such as manufactures, appliance models, operation modes, geolocations, environment parameters, and "age" or malfunction of appliances. Depending on how large the within category differences are, it may be beneficial to cluster appliances within the same category into subcategories for machine learning methods to learn specific parameters for each. Also as mentioned above, a diverse database enables one to build methods to predict changes in appliance signatures or changes in user behavior. For aggregation appliance cases, diversity is about the amount of different combinations of aggregate appliance data, including different operation modes and manufactures, being available in or derivable from the database. It is in general not possible to record all combinations, while what combinations are most practically relevant can be learnt from whole home data sets.

Scalable: New data has to be easily incorporated into the database. As the amount of data increases, the burden of using the database should be kept low, such that data processing, method modification, and comparison between methods before and after adding new data are easy. Detailed descriptions of data information mentioned previously, data format, and pointers are essential. In addition, important derivatives of the data can also largely increase the usability of the database. For example, locations of events and steady state regions may be provided along with each recording, such that training sets and test sets for related methods are easy to construct. One may argue that such information is rather subjective and method dependent. However, it is likely to be easier to modify than to derive these information from scratch, especially during the development of supervised learning methods where data labels are necessary. Related processing function files are useful to be included in the database.

3. IMPORTANCE OF DATA SET TYPES

Here we separate data sets into three types based on the number of simultaneously operating appliances. Starting from a dichotomy of single appliance data sets or whole home main circuit data sets, we suggest that intermediate aggregate data sets are also useful for developing machine learning methods for NILM.

Single appliance data sets: The single device data sets discussed here is "stand alone" recordings that no other appliance shares the same circuit during recording, or the impacts of other appliances shall be minimal. These data sets are essential that most existing NILM methods use single appliance data sets as a major part of their training sets. Also, validation using single appliance data sets directly reveals the strength and weakness of a method within and across appliance categories, and also provides a proxy upper bound of the accuracy when extending the method for whole home data sets. Thus, an informative and diverse single appliance data set is important for fine tuning methods and pushing the performance ceiling of real application scenarios. Single appliance data can also be used to synthesize aggregate appliance data sets.

Whole home main circuit data sets: These data sets are the best test sets, as they are recorded under the same conditions which NILM technology shall be applied. In addition, it also provides behavioral information about when and how appliances are used.

Intermediate aggregate appliance data sets: To improve accuracy of a method, one has to focus on when and why the method makes an error. It is costly to record additional whole home data for extracting similar scenarios. By establishing intermediate aggregate appliance data sets, one can specifically focus on scenarios related to the error, e.g. specific combinations of aggregate appliance categories. A collection of such data sets also enables one to characterize challenging cases and to develop new methods focusing on them. These methods can then be used with existing methods to form a *committee*, whose performance can be significantly better than individual methods [3].

Cases with the same appliance categories as the error scenarios but different manufactures or operation modes are also important. However recording all possible combinations is impractical. Specially designed aggregate appliance data sets can be used to guide the synthesis of aggregate appliance data sets from single appliance data sets. For example, when multiple appliances are plugged to the same power strip, the composite current and power are often not linear in individual current and power. Scaling factors and nonlinear effects of electric signals, and noise models can be learnt from cross comparison between aggregate and single appliance data sets.

4. CONCLUSION

Availability of a database which is informative, diverse, and scalable is critical for applying and developing machine learning methods for NILM. Intermediate aggregate appliance data sets focusing on challenging cases can facilitate the development of new methods; also along with single and whole home main circuit data sets, synthesized data sets can be more realistic. A comprehensive database can be established gradually by merging data sets from different sources.

5. ACKNOWLEDGMENTS

The authors would like to thank Mario Bergés for his invitation.

6. REFERENCES

[1] Berges, M., Goldman, E., Matthews, H. S., Soibelman L. Learning systems for electric consumption of buildings. In Proceedings of the 2009 ASCE International Workshop on Computing in Civil Engineering, Austin, Texas.

[2] Berges, M. E., Goldman, E., Matthews, H. S., Soibelman L. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. Journal of industrial ecology, Vol. 14, No. 5, pp. 844-858, 2010.

[3] Bishop, C.M. Pattern Recognition and Machine Learning. Anderson. Springer Science + Business Media, LLC. 2006

[4] Du Y., Du. L., Lu B., Harley R., and Harbetler T. A review of identification and monitoring methods for electric loads in commercial and residential buildings. In Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE), Atlanta, GA, USA, 2010.

[5] Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., Patel, S. Disaggregated end-use energy sensing for the smart grid, IEEE pervasive computing, Vol. 10, No. 1, pp. 28-39, Jan-Mar 2011.

[6] Gupta S., Reynolds M., Patel S. N., ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home, In the Proceedings of Ubicomp, Copenhagen, Denmark, 2010.

[7] Kolter J.Z. and Johnson M.J. REDD: A public data set for energy disaggregation research. In Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, San Diego, CA, USA, 2011.

[8] Laughman, C., Kwangduk L., Cox R., Shaw S., Leeb S., Norford L. and Armstrong P. Power signature analysis, IEEE power and energy magazine, pp. 56 -63, March/April 2003.

[9] Patel, S.N., Robertson, T., Kientz, J.A., Reynolds, M.S., Abowd, G.D. At the flick of a switch: detecting and classifying unique electrical events on the residential power line. In the Proceedings of Ubicomp, Innsbuck, Austria. 2007.

[10] Rowe, A., Berges, M., Rajkumar, R. Contactless sensing of appliance state transitions through variations in electromagnetic fields. BuildSys 2010, pp. 19-24, Nov., 2010.

[11] UCI Machine Learning Repository. Home page link http://archive.ics.uci.edu/ml/index.html

[12] Zeifman M. and Roth K. Nonintrusive load monitoring: review and outlook. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), Cambridge, MA, USA, 2011.

About the authors:

Po-Hsiang (Sean D.) Lai received the B.S. Applied Science degree and the D.Sc. degree both in electrical engineering from Washington University in Saint Louis, in 2006 and 2012. His research interests are machine learning and information theory in biometrics, imaging, and nonintrusive load monitoring. He is currently a senior engineer at Samsung Dallas Technology Lab.

Mark Trayer graduated from the University of Bradford (UK) in Electronic, Communication and Computer Engineering in 1989. He has been part of the Samsung research and development team in Richardson, TX since 2005, working across a diverse set of technologies, from Core Network Voice over IP systems through to Smart Energy and is currently a Principal Engineer in the Technology Enabling Group. His research interests are in the arena of the Smart Home; focusing on energy analytics, micro-grids, energy management algorithms and architectures. Prior to joining Samsung, Mark spent 16 years working for the (company formerly known as) Nortel in both the UK and the US.

Sudhir Ramakrishna received his B-Tech degree in 1992 from the Indian Institute of Technology, Mumbai, India, and the M.S & Ph.D. degrees in 1994 & 1998, from Rutgers University, New Brunswick, New Jersey. All his degrees are in Electrical Engineering.

He is a Principal Engineer in the Standards & RF Lab at the Samsung R&D Center in Richardson, TX, where he has been since 2008. In addition to smartgrid energy management algorithms, his research interests include the area of cellular communications- MIMO systems, problems in resource allocation and link- & system-level performance optimization algorithms.

Ying Li received the B.E. degree (with honors) and the M.E. degree (with honors) in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1997 and 2000 respectively, and the M.A. and Ph.D. degrees in electrical engineering at Princeton University, Princeton, NJ, in 2005 and 2008, respectively.

She has been with Samsung Telecommunications America, Dallas, TX, since October 2008, where she is involved in research on smart energy networks and wireless communications. She worked as a faculty member at Xi'an Jiaotong University from 2000 to 2003 and as a visiting scholar in Fuji Xerox Co. Ltd., Japan, from 2000 to 2001. Her research areas include optimization, smart energy networks, energy monitoring and management, communication networks, next generation wireless communications, heterogeneous networks, cross-layer design, content distribution, information theory, and signal processing.