Applied Energy 211 (2018) 1106–1122

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network

Chao Liu[a], Adedotun Akintayo[a], Zhanhong Jiang[a], Gregor P. Henze[b,c], Soumik Sarkar[a,*]

[a] Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA
[b] Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO 80309, USA
[c] National Renewable Energy Laboratory, Golden, CO 80401, USA

HIGHLIGHTS

- Building energy disaggregation using spatiotemporal pattern network (STPN).
- Novel STPN framework to capture multivariate time series features.
- Selecting most informative variables using information theoretic metric on STPN.
- Building similarity metric formulation using STPN for achieving scalability.
- STPN framework outperforms state-of-the-art techniques in various NILM test cases.

ARTICLE INFO

Keywords:
Non-intrusive load monitoring (NILM)
Spatiotemporal pattern network (STPN)
Multivariate time-series

ABSTRACT

Non-intrusive load monitoring (NILM) of electrical demand for the purpose of identifying load components has thus far mostly been studied using univariate data, e.g., using only whole building electricity consumption time series to identify a certain type of end-use such as lighting load. However, using additional variables in the form of multivariate time series data may provide more information in terms of extracting distinguishable features in the context of energy disaggregation. In this work, a novel probabilistic graphical modeling approach, namely the spatiotemporal pattern network (STPN) is proposed for energy disaggregation using multivariate time-series data. The STPN framework is shown to be capable of handling diverse types of multivariate time-series to improve the energy disaggregation performance. The technique outperforms the state of the art factorial hidden Markov models (FHMM) and combinatorial optimization (CO) techniques in multiple real-life test cases. Furthermore, based on two homes' aggregate electric consumption data, a similarity metric is defined for the energy disaggregation of one home using a trained model based on the other home (i.e., out-of-sample case). The proposed similarity metric allows us to enhance scalability via learning supervised models for a few homes and deploying such models to many other similar but unmodeled homes with significantly high disaggregation accuracy.

1. Introduction

Non-intrusive load monitoring is a well-established problem that involves disaggregating the total electrical energy consumption of a building into its constituent electric load components without the necessity for extensive metering installations on individual end-uses. Such problems are relevant and challenging from the perspective of software-based detection for control of end-use patterns as well as the future internet of things (IoT). Recently, such problems have been gaining widespread attention due to the potential benefits of energy disaggregation for the purposes of energy efficiency and the development of smart grid systems [1–3], especially in the presence of increasing penetration of renewable energy resource [4]. An overview of load disaggregation concepts was provided in [5], beginning from the pioneering disaggregating technique in [6] that detected sharp changes in signals to optimization of error terms for pattern detection using genetic algorithms [7]. Fourier transforms of end-use signals also provided a useful categorization of their patterns [8]. Authors in [8] evaluated the performance of factorial hidden Markov models (FHMM)[9] for the energy disaggregation problem.

Recent approaches have either attempted determining the hidden states of the FHMM using modified tractable sparse Viterbi algorithm

* Corresponding author.
E-mail address: soumiks@iastate.edu (S. Sarkar).

https://doi.org/10.1016/j.apenergy.2017.12.026
Received 5 August 2017; Received in revised form 25 October 2017; Accepted 1 December 2017

[10], but the methods cannot scale properly for multivariate analysis without conditioning. With conditional Factorial Semi-Hidden Markov Model (CFSHMM) [11], the internal states are conditioned on the additional appliances features, in addition to a suitably chosen prior shape distribution representing the appliances relationships. CFHSMM's maximization step however assumes independence and cannot handle the relational dependencies that we consider here.

In (Dinesh et al. 2016), spectral and energy level information of appliances are determined from the subspace components derived from the active power, after which mean shift is used to classify each from short windows. However, using aided Linear Integer Programming ALIP [12] mainly included extra constraints, some appliance based knowledge and median filtering to improve the computational efficiency of the ALIP on several appliances. The usage of such ALIP is still limited to univariate disaggregation. Indeed, many past work on non-intrusive load monitoring for energy disaggregation has been primarily based on univariate real power measurements at different sampling intervals [3,13–20].

With the importance of NILM, there was the need for standardized datasets for benchmarking the large variety of published algorithms. In that light, authors of [21] were motivated to build the Reference Energy Disaggregation Data set (REDD), facilitated by hardware and software systems designed to collect real and reactive power of appliances from multiple homes in Boston, MA in 2011. An open source toolkit, NILMTK was subsequently developed [22] as a common platform to enhance the reproducibility of the algorithms' results on available datasets such as REDD [21], BLUED [23], Smart∗ [24], and several others. For the purpose of benchmarking the results of the algorithms, several performance evaluation metrics had been proposed [25–27].

The motivation of this present work is the vision of a smart grid [28–31], in which grid operators know what electric load components dominate the demand at each point of service and any point in time. Such capability enables the operators to call on flexible demand side resources for demand response and demand shaping that is perhaps due to large fluctuations of renewable energy sources. This vision of the smart grid involves estimation of the electric load breakdown, not only at the level of a single home but at different levels of aggregations such as residential transformers serving a few homes or distribution feeders serving many hundreds of homes [32]. In this context, a univariate measurement is possibly insufficient for the task of identifying the dominant load contributions such as air-conditioning related demand. Instead, at any level of aggregation, we hypothesize that a multivariate measurement of several concurrent electrical properties may improve the inference on what load components are contributing to the transformer or feeder demand. This may also work for other scenarios such as thermal load monitoring [33], energy intensity of domestic activities [34], and specific saving of appliances [35].

In electric power systems, magnitude and phase angle of the various wave forms can be presented by a complex number referred to as a phase vector, or phasor, and it can be measured by a newly developed phasor measurement unit (PMU) [36]. The wealth of measured multivariate time series data allows for the previously unavailable assessment of the incremental value of measuring not only the real power but several other electric system variables for the purpose of identifying usage patterns and load components. In addition, there are other resources (e.g., weather data, building property, occupancy and usage logs) that can provide multivariate data for the NILM. With this kind of multi-source information perspective, the performance of the energy disaggregation might be improved. The authors in [11] considered the disaggregation problem by deriving multi-appliance relationship. The authors in [28] proposed a knowledge-based classification of appliances signals based on measured variables into triangular or rectangular sections to perform disaggregation. Here however, data-driven NILM using multivariate relationship to individual end use using information theoretic measure is introduced. A related question that we aim to answer is: If multivariate time series is more useful than univariate time series for energy disaggregation and if so, what are the most valuable measurements to use?

In this context, we apply the recently developed framework of spatiotemporal pattern networks (STPN) that is built on the concept of symbolic dynamics [37]. STPN is proposed to model multivariate time series, via learning atomic patterns (APs, Markov models for individual variables) and relational patterns (RPs, Markov models to model causal interactions between variables) [38,39]. Patterns of multivariate time series data are formed based on these features (APs and RPs) and then used to study the characteristics of electricity usage.

The state-of-the-art NILM approaches are mostly supervised learning schemes, where measuring the usage of the end-uses is needed [16]. The goal of NILM is thus influenced significantly by the cost of sensor and metering installations. An unsupervised learning method which only employs unlabeled data is presented in [40] via building a set of probabilistic end-use models, yet how to determine the number of labeled end use models is still an on-going problem. However, even in a supervised context, if the number of the sub-metered homes can be reduced while preserving the disaggregation performance on the total number of buildings analyzed, the cost of NILM application will be greatly reduced. Therefore, this work proposes a similarity metric using the STPN concept for out-of-sample disaggregation (a model trained in one home with sub-metered end-uses is applied for disaggregation of another home without end-uses sub-metered). With the similarity metric, one can evaluate the number of homes that need to be sub-metered, and hence the installment cost and disaggregation performance can be estimated.

Contributions: The primary contributions of this paper include: (i) proposing an STPN framework for energy disaggregation using multivariate time-series data, (ii) applying a mutual information based metric to explore the energy consumption patterns and select the most valuable variables for disaggregation, (iii) validation of the STPN scheme in diverse test cases showing that it outperforms the state-of-the-art techniques in NILM, and (iv) proposing a similarity metric for homes/buildings based on STPN for out-of-sample disaggregation, which will enable us to predict end uses power for many homes without necessarily retraining them or expensive submetering in all of them.

Outline: The remaining paper is organized as follows. Section 2 provides a brief background and preliminaries including the definition of STPN and metrics in causality interpretation with STPN. Section 3 presents the approach to explore patterns of multivariate time series to distinguish types of end-uses, the framework of energy disaggregation using multivariate time-series data via STPN, and the similarity metric for out-of-sample disaggregation. Section 4 describes results for the discovered patterns in different kinds of appliances (solar yield, A/C usage, using data sets on Distribution phasor Measurement Units–DMUs), and disaggregation using ECO and RBSAM data sets. Finally, the paper is summarized and concluded with directions for future work in Section 5.

2. Background on Spatiotemporal Pattern Network (STPN)

In the view of multivariate time series data in NILM, features among different signals can be distinguished and used for disaggregation. As the STPN is suitable for extracting spatial and temporal features from time-series data, it has been successfully applied in classification and detection areas [41]. Different from the current NILM approaches that solely use the total power consumption for disaggregation, the proposed setup of STPN that can leverage multiple information sources which can improve the disaggregation performance.

Before introducing STPN, the concept of probabilistic finite state automaton (PFSA) is first defined here as a basis. PFSA is defined in the symbolic space that is generated via time series abstraction (preprocessing and discretization/partitioning [38]). As shown in Fig. 1, the time-series data is discretized into symbol sequences and then state sequences, PFSA is formed using *D*-Markov machine. More formally,