NILMTK: An open source toolkit for Non-intrusive load monitoring

Nipun Batra[1], Jack Kelly[2], Oliver Parson[3], Haimonti Dutta[4], William Knottenbelt[2],
Alex Rogers[3],Amarjeet Singh[1], Mani Srivastava[5]

1. ABSTRACT

Non-intrusive load monitoring, or energy disaggregation, aims to separate household energy consumption data collected from a single point of measurement into individual appliances. In recent years, the field has rapidly increased in size with the national rollouts of smart meters, and as a result many new disaggregation algorithms have been proposed. However, empirically comparing such algorithms is currently virtually impossible, as a result of the different data sets used, the lack of availability of algorithms, and the wealth of accuracy metrics which have been employed. To address this challenge, we present NILMTK; an open source toolkit designed specifically to enable the comparison of energy disaggregation algorithms. Our toolkit includes parsers to a range of existing data sets, a standard output format for disaggregation algorithms, a set of statistics for describing datasets, a reference benchmark disaggregation algorithm, and a suite of accuracy metrics. We demonstrate that our toolkit lowers the barrier to entry by simplifying the use of multiple data sets, while also encouraging direct comparisons to be made between algorithms through common output formats and accuracy metrics.

2. INTRODUCTION

Non-intrusive load monitoring (NILM), or energy disaggregation, aims to break down a household's aggregate electricity consumption into individual appliances (Hart, 1992). The motivations for such a process are threefold. First, informing a household's occupants of how much energy each appliance consumes empowers them to take steps towards reducing their energy consumption (Darby2006). Second, personalised feedback can be provided which quantifies the savings of certain appliance-specific advice, such as the financial savings when an old inefficient appliance is replaced by a new efficient appliance. Third, if the NIALM system is able to determine the time of use of each appliance, a recommender system would be able to inform the household's occupants of the potential savings through deferring appliance use to a time of day when electricity is either cheaper or has a lower carbon footprint.

Such benefits have drawn significant interest in the field since its foundation 25 years ago. In recent years, the combination of national smart meter meter rollouts (e.g. REF UK, USA) and also decreasing hardware costs of household electricity sensors has lead to a rapid expansion of the field. Such rapid growth over the past 5 years has been evidenced by the wealth of academic papers published, international meetings held (NLM 2012, EPRI NILM 2013), startup companies founded (e.g. Bidgely, Neurio) and data sets released.

However, three core barriers are currently preventing the direct comparison of existing approaches, and as a result are limiting progress within the field. First, each contribution to date has only been evaluated on a single data set as a result of the difficulty in moving from one data set to another, and consequently it is hard to assess the generalisability of approaches. Furthermore, many papers subsample existing data sets to select specific households, appliances and time periods, making experimental results harder to reproduce. Second, newly proposed approaches are rarely benchmarked against the same algorithms, further increasing the difficulty in empirical comparisons of performance between different papers. Moreover, the lack of availability of state-of-the-art algorithms often leads to the reimplementation of such approaches. Third, each paper targets a subtly different use case for NILM and therefore evaluates the accuracy of their proposed approach using a different set of performance metrics. As a result the numerical performance calculated by such metrics cannot be compared between two papers. These three barriers have led to successive extensions to state-of-the-art algorithms being proposed, while a direct comparison between such approaches has remained impossible.

Similar barriers also have arisen in other research fields and prompted the development of toolkits specifically designed to support a given area. For example, PhysioToolkit offers access to over 50 databases of physiological data and provides software to support the processing and analysis of such data for the biomedical research community [1]. Similarly, CRAWDAD collects

89 data sets of wireless network data in addition to supporting software to aid the analysis of such data by the wireless network community [2]. However, no such toolkit is available to the NILM community.

Against this background, we propose NILMTK; an open source toolkit designed specifically to enable the comparison of energy disaggregation algorithms. The primary aim of the toolkit is to provide a complete pipeline from data sets to accuracy metrics, therefore lowering the barrier to entry for researchers to plug in a new algorithm and compare its performance to the current state of the art. The toolkit has been released as open source software in an effort to encourage researchers to contribute data sets, benchmark algorithms and accuracy metrics as they are proposed, with the goal of enabling a greater level of collaboration within the community. In addition, the toolkit has been designed using a modular structure, therefore allowing researchers to reuse or replace individual components as required. Last, the toolkit has been written in python with flat file input and output formats, ensuring compatibility with existing algorithms written in any language and designed for any platform.

Our contributions are summarised as follows: We propose REDD+, the standard structure for energy disaggregation data sets used by NILMTK based on an extension of the format used for the REDD data set. Furthermore, we provide parsers from multiple existing data sets into our proposed REDD+ format. In addition, we propose a single output format for the disaggregated data produced by NILM algorithms. We provide an implementation of a benchmark disaggregation which uses combinatorial optimisation to disaggregate household energy data into individual appliances. We demonstrate the ease by which NILMTK allows this algorithm can be applied to a range of existing data sets, and present results of its performance accuracy. We present a suite of accuracy metrics which are able to evaluate the performance of any disaggregation algorithms which produce output compatible with NILMTK. This allows the performance of a disaggregation algorithm to be evaluated for a range of use cases.

The remainder of this paper is organised as follows. In Section X we provide an overview of related work from the field of NILM and also other similar fields of research. In Section X we present NILMTK, and give a detailed description of its components. In Section X we demonstrate the empirical evaluations which are enabled by NILMTK, and provide analysis of existing data sets and accuracy metrics. Finally, in Section X we conclude the paper and propose directions for future work.

3. NILMTK

We took the following three views/design goals while coming up with a design of NILM toolkit:

- **Analysis**: The toolkit should facilitate easy analysis of NILM datasets and expose the entire pipeline from data importing to disaggregation and finally results.

- **Ease of adding new algorithms**: The toolkit should provide consistent interfaces for new disaggregation algorithms.

- **Ease of deploying**: The toolkit should be built in such a way that the learnt model is easily deployable and can interface with online as well as offline data.

  With these three design goals in mind, we decided to use Python as the programming language for building NILMTK

4. REFERENCES

[1] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[2] D. Kotz and T. Henderson. Crawdad: A community resource for archiving wireless data at dartmouth. *Pervasive Computing, IEEE*, 4(4):12–14, 2005.