Making energy disaggregation practical

Nipun Batra
IIIT Delhi
November 1, 2015
Buildings contribute significantly to overall energy consumption.
Buildings getting constructed at a rapid rate.

Dubai 1991

Dubai 2013
Buildings are an attractive target towards sustainability
Residential buildings can contribute up to 93% of building energy usage.
“If you cannot measure it, you cannot improve it” - Kelvin
Sensor deployments have several challenges*

1. Homes are not a power panacea
2. Homes have poor connectivity
3. Homes are hazardous
4. Limited user interaction
5. Aesthetics matter

2. Batra et al. “It’s different. Insights into home energy consumption in India”. Buildsys 2013
Non intrusive Load monitoring (NILM) or Energy disaggregation

Smart meter

Machine learning
Why NILM can work

Different appliances can have unique “signatures”
Making NILM practical

1. Comparable
2. Utility-driven
3. Scalable
Making NILM comparable

eEnergy 2014 and Buildsys 2014
What is the best NILM approach?

Despite 30+ years of NILM research, really hard question.

3 main problems
I: Hard to assess generality

- Previous contributions evaluated only on single dataset
- Non-trivial to set up similar experimental conditions for direct comparison
2: Lack of comparison against same benchmarks

• Newly proposed algorithms rarely compared against same benchmarks
• Lack of “open source” reference algorithms often lead to reimplementation
3: “Inconsistent” disaggregation performance metrics

- Different performance metrics proposed in the past
- Different formulae for same metric, eg. 4+ versions of “energy assigned”

\[
\text{Acc} = 1 - \frac{\sum_{t=1}^{T} \sum_{i=1}^{n} |\hat{y}_t^{(i)} - y_t^{(i)}|}{2 \sum_{t=1}^{T} \hat{y}_t}
\]

\[
\left(\sum_{t,i} \|\hat{y}_t^{(i)} - \bar{y}_t^{(i)}\|_2^2 / \left(\sum_{t,i} \|\bar{y}_t^{(i)}\|_2^2\right)\right)^{1/2}
\]

\[
\left|\sum_{i} x_t^{(n)} - \frac{\sum_{i} y_t^{(n)}}{\sum_{i} x_t^{(n)}}\right| / \sum_{i} x_t^{(n)}
\]

\[
MNE(n) = \frac{\sum_{t=1}^{T} |\theta_t^n - y_t^n|}{\sum_{t=1}^{T} \theta_t^n}
\]
And NILMTK was born

Open source **NILM toolkit** to enable **easy comparative analysis** of NILM algorithms **across data sets**
NILMTK pipeline

Data interface

- REDD
- BLUED
- UK-DALE
- NILMTK-DF
- Statistics
- Training
- Preprocessing
- Model
- Disaggregation
- Metrics
NILMTK-DF: Common data format
10 data sets released

Data interface

- REDD
- BLUED
- UK-DALE

Statistics → Preprocessing → Training → Model → Disaggregation → Metrics

NILMTK-DF
Statistical functions

Suite of commonly used statistical functions

Data interface

Ground truth quality

[Bar chart showing the quality of different datasets: REDD, SMART*, PECAN, AMPds, iAWE, UK_DALE]
Preprocessing

Data interface

- REDD
- BLUED
- UK-DALE

Statistics

NILMTK-DF

Preprocessing

Training

Model

Disaggregation

Metrics

Time (day/month/year)

Fridge
Washer dryer
Kitchen outlets
Mains 1
Mains 2

Dropout rate (%)
Train and Disaggregate

Data interface

REDD

BLUED

UK-DALE

Statistics

NILMTK-DF

Preprocessing

Training

Model

Disaggregation

Metrics
Train and Disaggregate

Hart’s event detection algorithm

Factorial Hidden Markov Model (FHMM)

Combinatorial Optimisation

<table>
<thead>
<tr>
<th>Appliance</th>
<th>Off power</th>
<th>On power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>Fridge</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
NILMTK impact

• 10+ papers using NILMTK (4 in Buildsys 2015)
• 2 user contributed NILM algorithms
• 3 user contributed NILM data sets
• Best demonstration award at Buildsys 2014
Making NILM utility-driven

Buildsys 2015 and Percom 2016 (under submission)
If you can measure, can you improve?
Does NILM really save energy?

Does telling you that HVAC takes 56% save you energy?

Lack of specific actionable insights

Misc.: 22%
Light: 10%
Fridge: 11%
HVAC: 56%
Exploring the value of Energy disaggregation

1. Can disaggregated traces provide actionable insights?

Submetered appliance traces → Appliance energy modelling → Appliance energy models → Identify homes needing feedback
Exploring the value of Energy disaggregation

2. Do existing NILM techniques provide traces with sufficient fidelity to support feedback?

Diagram:
- **Aggregate household power** → **NILM** → **Appliance energy modelling** → **Appliance energy models** → **Identify homes needing feedback**
- **Disaggregated appliance traces**
Feedback methods on Fridge and HVAC

- Both appliances common across homes
- Both appliances contribute heavily to overall energy consumption
Fridge energy modelling
We can break down fridge energy with less than 4% error.
13 out of 95 homes can be given feedback based on usage energy saving up to 23% fridge energy.
17 out of 95 homes can be given feedback on **excess defrost** saving up to 25% fridge energy.
HVAC modelling

Objective

1. Learn set point from weather and energy data
2. Optimising setpoint can save upto 20-30% HVAC
HVAC feedback

- 84% accuracy on giving feedback based on setpoint temperature
Exploring the value of Energy disaggregation

2. Do existing NILM techniques provide traces with sufficient fidelity to support feedback?

- Aggregate household power
- NILM
- Appliance energy modelling
- Appliance energy models
- Identify homes needing feedback
- Disaggregated appliance traces
Benchmark NILM algorithms on our data set give accuracy comparable to state-of-the-art

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Dataset</th>
<th>#Homes</th>
<th>Algorithm</th>
<th>Fridge</th>
<th>HVAC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE (W)</td>
<td>Error</td>
</tr>
<tr>
<td>Kolter [15]</td>
<td>2012</td>
<td>REDD</td>
<td>6</td>
<td>Additive</td>
<td>-</td>
<td>62.5 △</td>
</tr>
<tr>
<td>Parson [18]</td>
<td>2012</td>
<td>REDD</td>
<td>6</td>
<td>Difference</td>
<td>83</td>
<td>55</td>
</tr>
<tr>
<td>Parson [19]</td>
<td>2014</td>
<td>Coldenψ</td>
<td>117</td>
<td>Bayesian</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Batra [5]</td>
<td>2014</td>
<td>iAWE</td>
<td>1</td>
<td>FHMM</td>
<td>-</td>
<td>50</td>
</tr>
<tr>
<td>Current work</td>
<td>Data port</td>
<td>240</td>
<td>CO*</td>
<td>85</td>
<td>19</td>
<td>0.65</td>
</tr>
<tr>
<td>Current work</td>
<td>Data port</td>
<td>240</td>
<td>FHMM*</td>
<td>95</td>
<td>20</td>
<td>0.63</td>
</tr>
<tr>
<td>Current work</td>
<td>Data port</td>
<td>240</td>
<td>Hart</td>
<td>82</td>
<td>21</td>
<td>0.72</td>
</tr>
</tbody>
</table>
Exploring the value of Energy disaggregation

2. Do existing NILM techniques provide traces with sufficient fidelity to support feedback?
NILM algorithms show poor accuracy in identifying homes which can be given feedback based on **usage energy**
NILM algorithms don’t identify the defrost state and thus prevent feedback based on defrost energy.

Defrost state is hard to detect!
NILM algorithms show poor accuracy in identifying homes needing HVAC setpoint feedback.
Take aways

1. Appliance level data **does** enable actionable energy saving feedback
2. Feedback accuracy can be **low** despite **good** disaggregation accuracy
3. We, the disaggregation community, need to **revisit the metrics** by which we measure progress
Making NILM scalable

IPSN 2016 (under submission)
3 fundamental problems

1. Lights (and other low power appliances) show poor disaggregation accuracy. Light are third highest overall in terms of loads.

2. Current NILM algorithms are often supervised and need careful tuning and model specification.

3. Most techniques assume 1 min. or less sampling interval. Existing smart meters sample once every 15 mins.
Can we leverage **big data** to make NILM more scalable?
Is **big data** more valuable than **precise** data?
precise data

- Smart meter
- 1 min sampling
- Fine tune model per home
big data

- Large number of homes
big data

- Large number of homes
- Submeter small subset of homes
- Use single reading per month
Similar homes have similar per-appliance energy consumption.
Approach: Neighbourhood NILM

I: Feature extraction and normalisation

II: Find K nearest neighbours for test home on extracted features

III: Predict energy usage of an appliance for test home as average of that appliance across K neighbours
Features

• Energy consumption:
  • Past 12 months household aggregate
  • Ratios (Min. energy/Max. energy)
• Static household properties: #occupants, Area, #rooms
Neighbourhood NILM comparable or better than best reported NILM accuracy
Neighbourhood NILM significantly accurate in Washing machine, dish washer, dryer- all pain points for traditional NILM.
High accuracy of “Oracle” suggests that our approach is promising.
Take away

**Big data** more valuable than **precise data** for the problem of energy disaggregation
Conclusions

Making NILM practical in 3 ways:

1. **Comparable** - NILMTK
2. **Utility-driven** - Energy saving feedback, Inferring household characteristics
3. **Scalable** - Neighbourhood NILM
Future work
Neighbourhood NILM with 15 minute meter data

1. Can we reduce the number of neighbours needed when we use 15 minute meter data
2. 15 minute data will present daily patterns, in addition to monthly patterns in current implementation
3. Metrics and utilities on 15 minute resolution
Homes “changing” behaviour pose an interesting challenge to Neighbourhood NILM

1. Balance between “historical” data and recent trends?
2. Continue having same neighbours?
   1. When to “change” the neighbours of a home
Scaling NILM to “similar” commercial buildings/different appliance types

1. Class of commercial buildings have exact same electrical infrastructure
2. Deployment across 10 dairy booths in New Delhi
“The cost of impact is a bug report/feature request a day on Github” :)

NILMTK / nilmtk

Issues

117 Open  306 Closed

- Scoring FHMM prediction
  #449 opened 5 days ago by knoxm

- Metadata handling tutorial needed
  #448 opened 15 days ago by gjwo

- Interpretation of Hart85 training
  #447 opened 16 days ago by gjwo

- Surfacing data from clustering to plot coloured scatter charts in Hart85
  #446 opened 19 days ago by gjwo

- convert_greend gives error
  #443 opened 23 days ago by cklemenj

- Interpretation of Hart_85 disaggregation results
  #442 opened 25 days ago by gjwo

- Advice needed on Hart output
  #438 opened 27 days ago by gjwo

- Missing graphs and timestamp comparison error
  #437 opened 27 days ago by gjwo
Conclusions

Making NILM practical in 3 ways:

1. **Comparable**- NILMTK
2. **Utility-driven**- Energy saving feedback
   Inferring household characteristics
3. **Scalable**- Neighbourhood NILM
Other work

1. Insights into home energy consumption in India [Buildsys 2013]
2. Inferring household characteristics from NILM [under submission Percom 2016]
3. Improving NILM performance using additional data [ICMLA 2013]