Extracting Discriminative Features for Event-based Electricity Disaggregation

Om Patri, Anand Panangadan, Charalampos Chelmis, Viktor Prasanna

Ming-Hsieh Department of Electrical Engineering,
University of Southern California, Los Angeles, CA

USC Viterbi School of Engineering

IEEE SusTech 2014

University of Southern California

Outline

- Introduction
- Dataset
- Approach
- Classifier
- Evaluation

Introduction

- Smart Grids: millions of smart meters that collect electricity consumption data at fine granularities.

- Access to appliance-level data still not to full potential. Benefits:
- For users
  - Significant energy use reduction through personalized feedback
  - Helps appliance purchase decisions
- For utilities
  - Enhanced load prediction models
  - Incentives recommendation

Electricity Disaggregation

- Electricity disaggregation
  - break down aggregate electricity consumption into appliance level itemized measurements without any plug level sensors
- Nonintrusive load monitoring (NILM) approaches for disaggregation have been used in the past
  - NILM disaggregation problem is that of identifying events associated with individual appliance activity from a stream of measurements, typically active power and voltage, provided at a high temporal rate
- Every switching on/off of an appliance is an 'event'

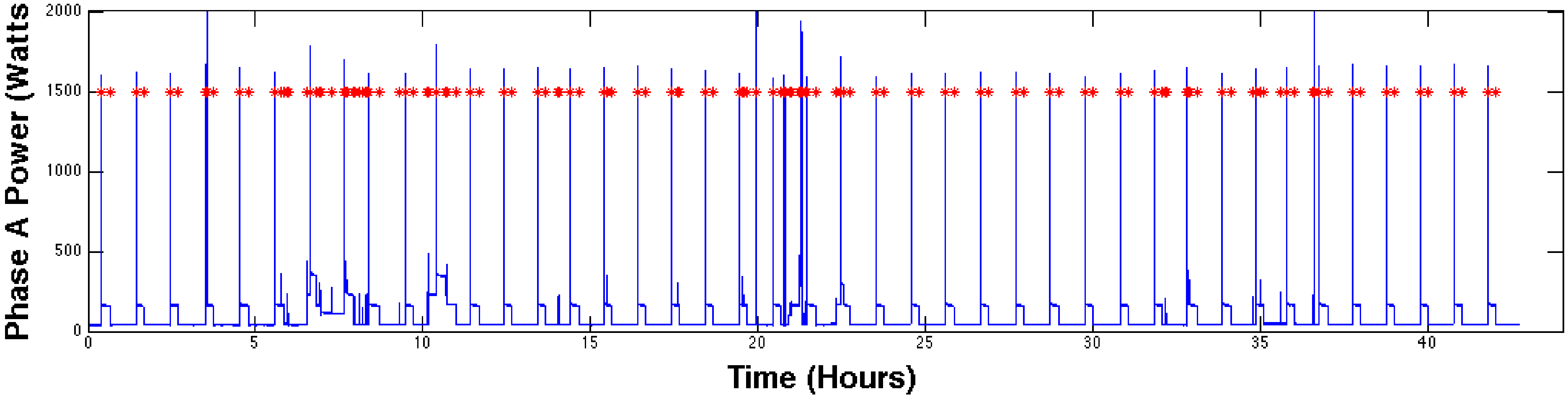
- We frame the NILM disaggregation problem in the context of time series classification and use a machine learning approach to solve it

Dataset

- We use the Building-Level fUlly-labeled dataset for Electricity Disaggregation (BLUED) dataset [Anderson et al., 2012] http://inferlab.org/publications/blued/

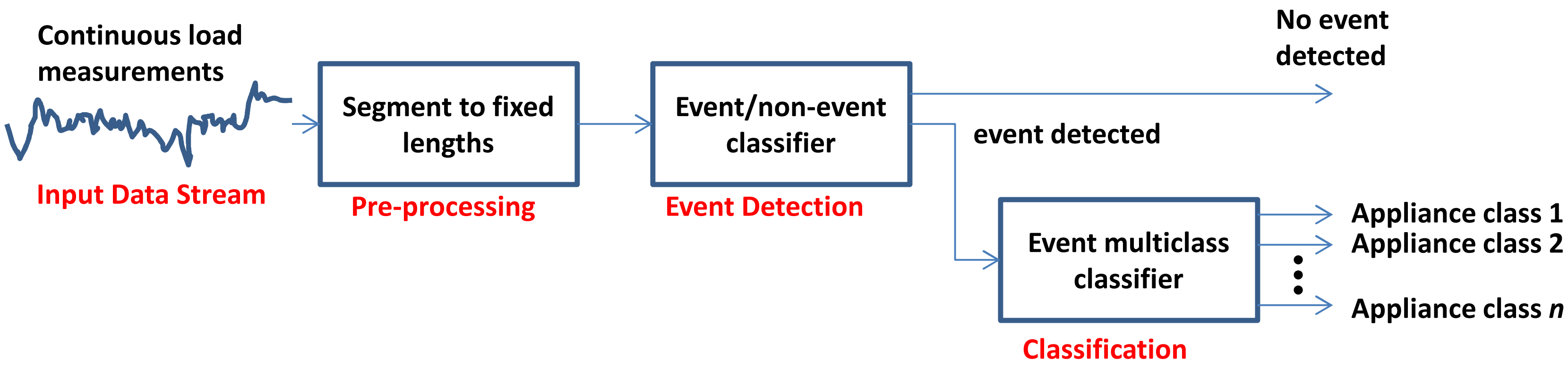
- Characteristics of dataset
  - Voltage, current, power measurements for a single house in the US for a week
  - 40+ appliances in the house
  - Every appliance event is labeled and timestamped to provide ground truth
  - Every switching on/off of an appliance is an '**event**'
  - Data from two phases : phase A and phase B (treated independently)
  - Used real power data at 60 Hz for both phases
  - About 37 million data points for each phase in the one week period
  - About 900 events in phase A and 1600 events in phase B

BLUED Dataset



- First quarter of power data for phase A shown in figure with events (red)
- This portion alone has about 9 million data points and 200 events
- The events are from various different appliances on phase A

Approach



- Pre-processing: Divide stream into several fixed length segments, needed for input to time series classification method for training and testing
- Event Detection: detect events v/s non-events
- Event Classification: for each detected event, identify which class of appliances was responsible for generating it
- Classifier: for both stages, we use Fast Shapelets [Rakthanmanon et al., SDM 2013] as the classifier

Approach

- Pre-processing
  - Converts long input data stream into multiple labeled time series
  - Consider a window around each event to extract multiple segments
  - Fixing length of window: we considered all data points 1 second before the event and 5 seconds after the event – thus, leading to a 6 second (360 data points) window

- Event Detection (first stage)
  - Supervised binary classifier
  - To train classifier, provide subsequences (of length 360) of both categories – with no event, and with a load change event
  - After training the classifier, a new subsequence can be presented for testing, and the predicted result will tell us if it contains an event or not

Approach

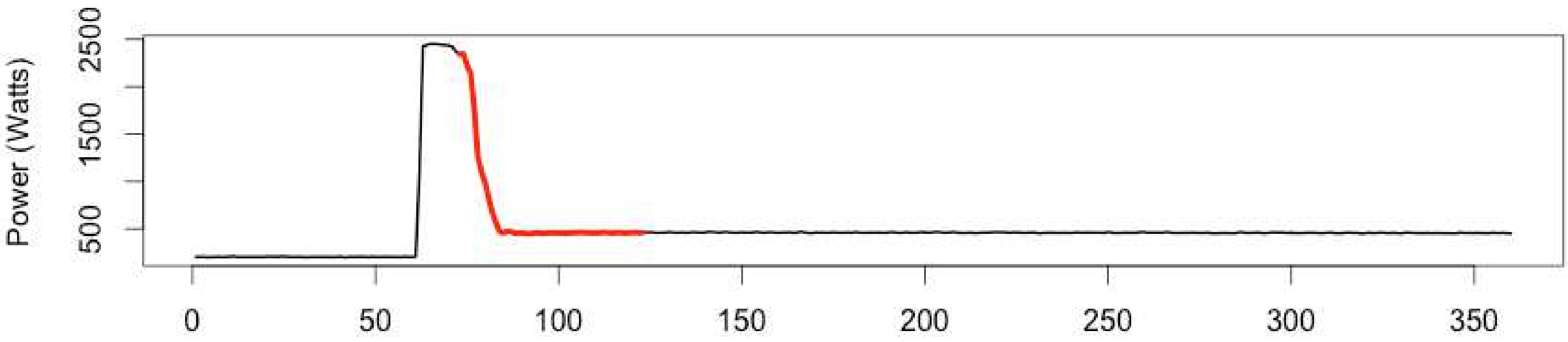
- Event Classification (second stage)
  - Actual disaggregation stage, supervised multiclass classifier
  - Training data contains fixed length subsequences of load measurements known to contain events from a specific class of appliances
  - Given a new segment, classifier outputs predicted class of appliances which produced the event
  - Appliance classes determined by us based on power usage and no. of events

Phase A appliance classes	
Class	Examples
1. Refrigerator	refrigerator
2. Lights	backyard lights, washroom light, bedroom light
3. High-Power (> 150W)	hair dryer, air compressor, kitchen chopper

Phase B appliance classes	
Class	Examples
1. Lights	desktop lamp, basement light, closet lights
2. High-Power (> 150W)	printer, iron, garage door
3. Low-Power (< 150W)	computer, LCD monitor, DVR/blu-ray player

Classifier: What are Shapelets ?

- We used the Fast Shapelets classifier, which is the fastest supervised time series shapelets classifier
- A ‘shapelet’ is a distinctive subsequence within a time series, that can discriminate between the two classes
- Algorithm for automatically finding shapelets first introduced by [Ye and Keogh, KDD 2009]
- Example of a shapelet (red) of length 51 found at position 73 (out of 360), from the event detection experiment on phase B power data



Classifier: How Shapelets Work ?

- Finding shapelets: of all possible candidate subsequences of all lengths, shapelet is the one with maximum information gain between Class 1 and Class 2

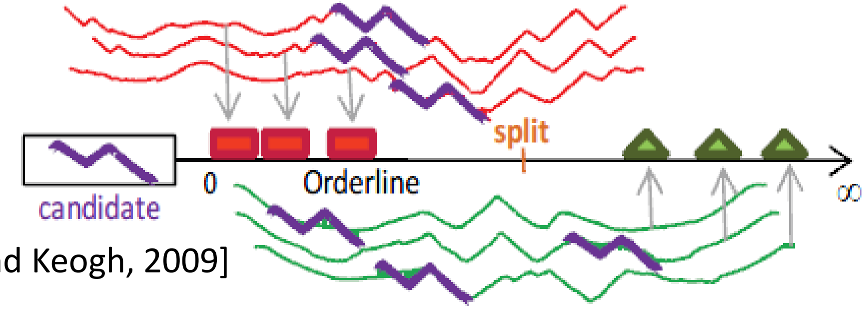
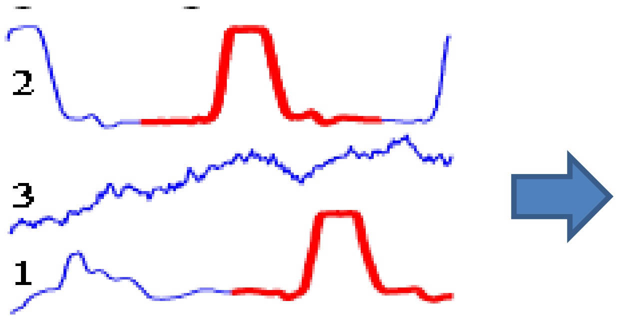
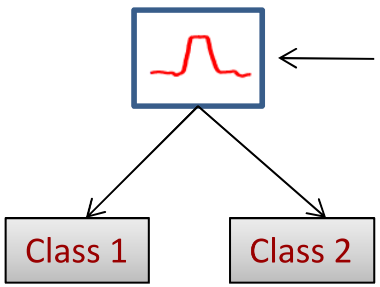
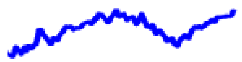


Figure from [Ye and Keogh, 2009]

- Information gain used to determine distance threshold
- Decision tree classifier based on distance from shapelet(s)







New time series:
Compute distance to shapelet.
If ≤ threshold, then class 1,
otherwise class 2.

Classifier: Why Use Shapelets ?

- Very fast classification time, since only the shapelet subsequences, and not the full training data is needed for classification

- Does not make assumptions on the nature of the data (unlike traditional time series approaches such as autoregressive methods, ARIMA)

- Visually interpretable

- The extracted subsequences/patterns can be used for further analysis and feedback from domain experts

Evaluation

- 50%-50% training-testing split in all experiments, first half of week used for training, and second half for testing
- Refrigerator has most number of events (switches on/off automatically)
- Categorized appliances into classes based on type of appliance, and number of events from the appliance

Phase A appliance classes	
Class	Examples
1. Refrigerator	refrigerator
2. Lights	backyard lights, washroom light, bedroom light
3. High-Power (> 150W)	hair dryer, air compressor, kitchen chopper

Phase B appliance classes	
Class	Examples
1. Lights	desktop lamp, basement light, closet lights
2. High-Power (> 150W)	printer, iron, garage door
3. Low-Power (< 150W)	computer, LCD monitor, DVR/blu-ray player

Evaluation: Event Detection

- Baseline: assign test instance class label of majority class in training data
- Experiments on balanced (1:1) as well as imbalanced (1:4) data splits
- Shapelet classifier trained on balanced data
- Baseline accuracy on balanced 50%, and on imbalanced 75%

Balance	Accuracy	test instances	Confusion Matrix			
1:1	**98.42%**	886	real		event	non-event
			pred	event	432	3
			pred	non-event	11	440
1:4	**98.56%**	2215	real		event	non-event
			pred	event	432	21
			pred	non-event	11	1751

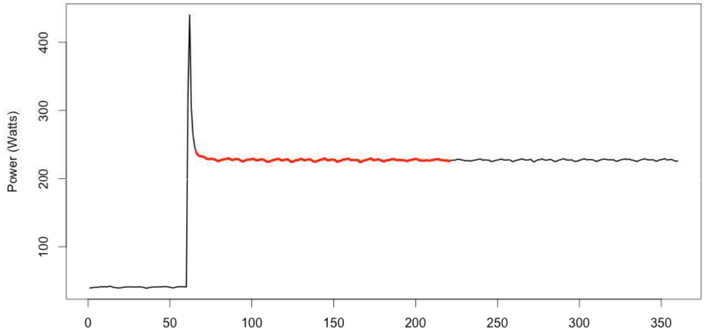
Phase A
Precision = 98%
Recall(1:1) = 99%
Recall(1:4) = 95%

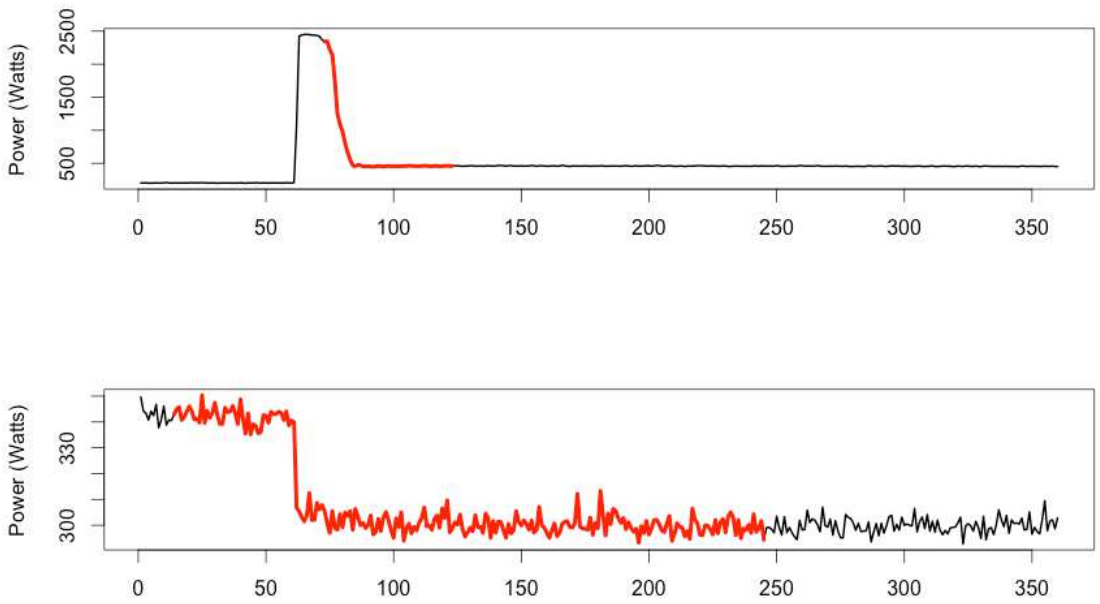
Balance	Accuracy	test instances	Confusion Matrix			
1:1	**98.32%**	1072	real		event	non-event
			pred	event	527	9
			pred	non-event	9	527
1:4	**97.91%**	2680	real		event	non-event
			pred	event	527	47
			pred	non-event	9	2097

Phase B
Precision = 98%
Recall(1:1) = 98%
Recall(1:4) = 92%

Evaluation: Event Detection

- Shapelets found (1 from phase A experiment on left, 2 from phase B experiment on right)





Evaluation: Event Classification

- Train-test split is 50%-50% as earlier
- Baseline accuracy on phase A is 77.25%, on phase B is 41.32%
- Results on phase A

Accuracy	test instances	Confusion Matrix				
		real	class label	1	2	3
83.75%	400	pred	1	295	36	9
		pred	2	9	26	5
		pred	3	5	1	14

Class	Precision (%)	Recall (%)
1	95.5	86.8
2	41.3	65.0
3	50.0	70.0

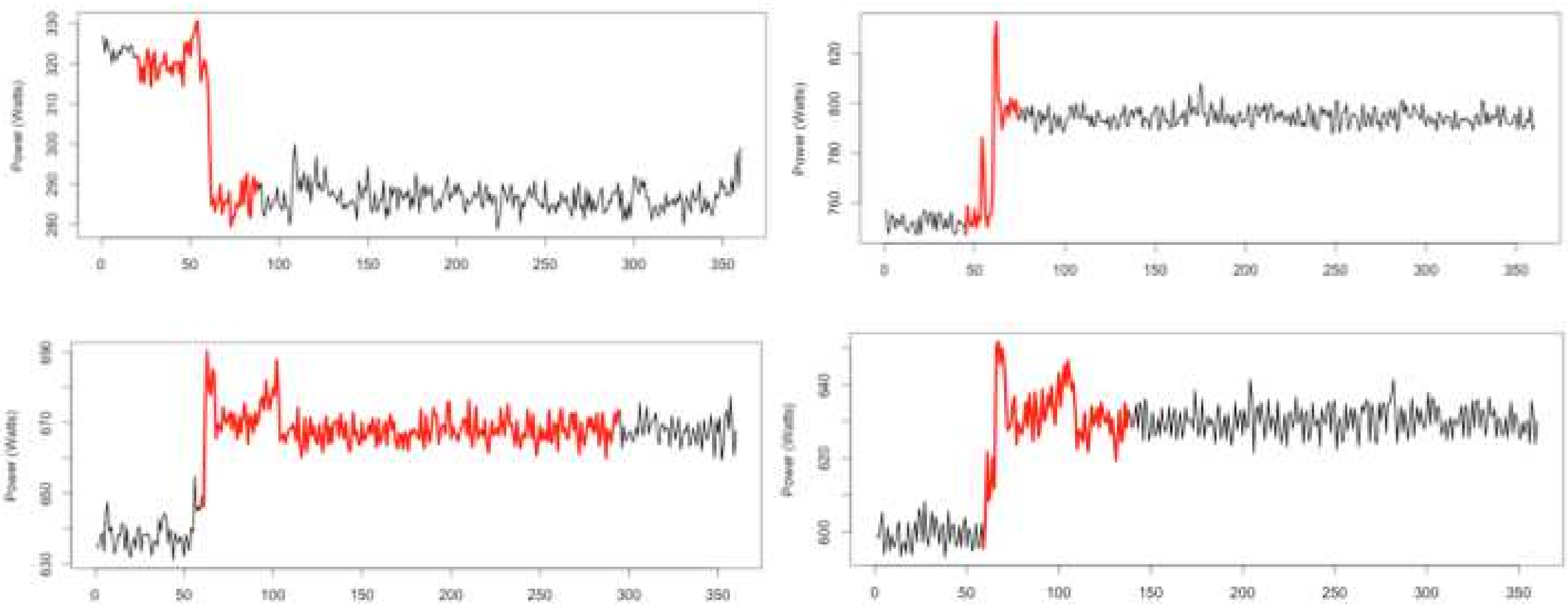
- Results on phase B

Accuracy	test instances	Confusion Matrix				
		real	class label	1	2	3
77.92%	317	pred	1	95	6	12
		pred	2	5	50	0
		pred	3	31	16	102

Class	Precision (%)	Recall (%)
1	72.5	84.1
2	69.4	90.9
3	89.5	68.5

Evaluation: Event Classification (Phase B)

- 7 shapelets from phase A expt, 14 from phase B expt (selective shown)



Observations

- Close to 100% accuracy in detecting events from non-events
  - It is easy for the shapelet classifier to pick up event signatures
- For appliance class identification task, the accuracy is significantly better than the baseline through our approach
  - This task is much harder than the first, because there are multiple appliance classes, with subtle differences between their signatures
  - Phase A: 77.25% -> 83.75%
  - Phase B: 41.32% -> 77.92%
  - Precision and Recall vary with the class of the appliance (trade-off)

- A variety of shapelet lengths and patterns are extracted throughout. Sometimes, the shapelet may not capture the exact change point for an event, but still provides good performance, because of relative comparison to differing instances of other class (also justifies our pre-processing method)

Other Disaggregation Approaches

- Examine identifying signature of consumer appliances [Gupta et al., 2010]
  - Hardware development and installation costs for special sensors
  - Still hard to identify signature due to number of appliances and diverse usage

- Different approaches for event detection in NILM [Anderson et al., 2012]
  - Expert heuristics
  - Probabilistic models
  - Matched filters

- We take a supervised time series classification approach, with appropriate pre-processing to transform the data into our required format

Summary

- Use of time series shapelets for electricity disaggregation
- Evaluated on publicly available large-scale BLUED dataset
- Our approach is able to achieve high accuracy in
  - Detecting events from non-events
  - Identifying which appliance class is associated with an event
- The classifier we used does not require the full training data for classification, so is fast, and doesn't make any assumptions about the data
- Results enable analysis and interpretation of appliance-class specific results to provide feedback to consumers and utilities, for energy efficiency
- Future Work:
  - Differentiate between appliance switching on and switching off events

Thank You !

Questions?

Contact:

Om Patri (patri@usc.edu)

Anand Panangadan (anandvp@usc.edu)

Viktor Prasanna (prasanna@usc.edu)

Webpage:

http://ganges.usc.edu/wiki/BDIA

Acknowledgement: This work is supported by Chevron U.S.A. Inc. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

Evaluation: Event Classification Phase A

- 7 shapelets from phase A expt, 14 from phase B expt (selective shown)