Smart Meters and Real-time Electricity Consumption Monitoring Algorithms to Reduce Electricity Theft in Developing Countries

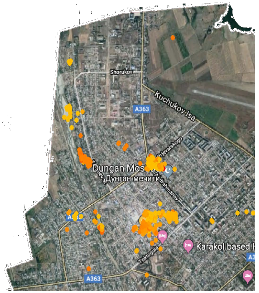
Team 23 - Jessie Ou[1], Jiwoo Song[2], Bernard Coles[3], Zhenxuan Wang[4], Dr. Robyn Meeks[5,6]

[1]Department of Computer Science, [2]Department of Mechancial Engineering and Materials Science, [3]Department of Sociology, [4]Nicholas School of the Environment, [5]Sanford School of Public Policy, [6]Energy Initiative

Introduction

Background:

- Utility companies worldwide lose more than $25 billion every year to electricity theft[1].
- In Kyrgyz Republic, the high losses have led to a 20% deficit in the energy sector[2], deteriorating the country's macroeconomic stability.



Karakol, Kyrgyzstan. Orange dots indicate smart meter locations.

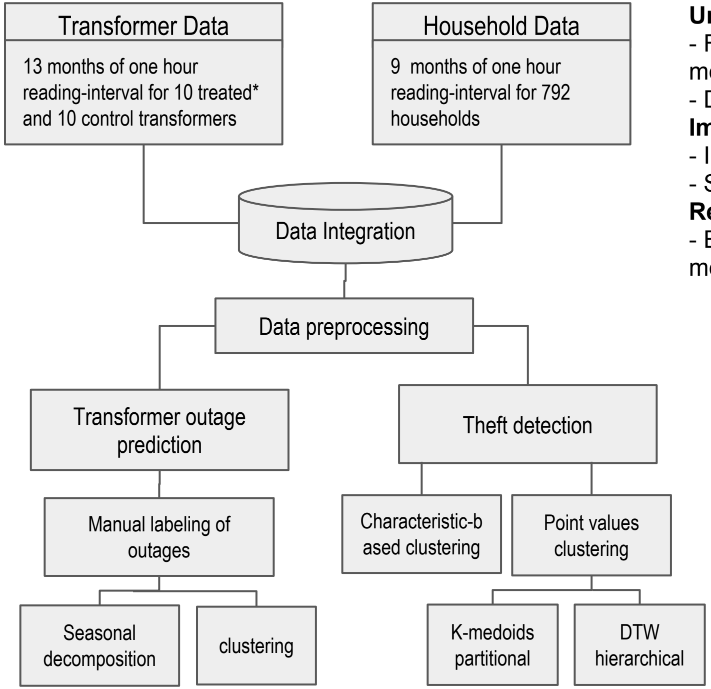
Gap in Knowledge:

- Detecting electricity theft has traditionally been addressed by physical checks of tamper-evident seals, making theft identification both labor-intensive and time-consuming. Labelled datasets are expensive and rare.
- This project aims to present novel unsupervised learning methods with potential applications for electricity utilities in developing country settings to automate fraud detection

Goals:

1) assess energy quality through predicting transformer outages with energy output data
2) classify user profiles and extract atypical and potentially malicious consumers by clustering household consumption trends

Pipeline



Transformer Data

13 months of one hour reading-interval for 10 treated* and 10 control transformers

Household Data

9 months of one hour reading-interval for 792 households

Data Integration

Data preprocessing

Transformer outage prediction

Manual labeling of outages

Seasonal decomposition

clustering

Theft detection

Characteristic-based clustering

Point values clustering

K-medoids partitional

DTW hierarchical

*Treatment group has smart meters installed at both the household and transformer level, whereas control group has smart meters at transformers but keeps analog meters at households

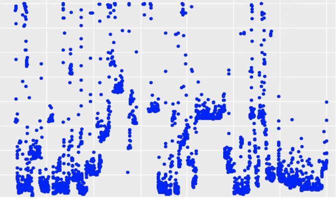
Preprocessing

Understanding missing data patterns
- Frequent malfunction during winter months
- Distant households miss more data

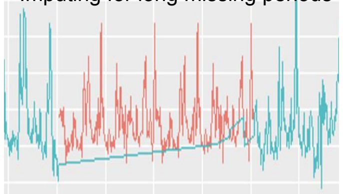
Imputing missing values
- Interpolation for 1-3 hour periods
- Simulating longer average consumption

Removing outlier values
- Ex: when engineers manually reset meters

Percentage missing each hour



Imputing for long missing periods

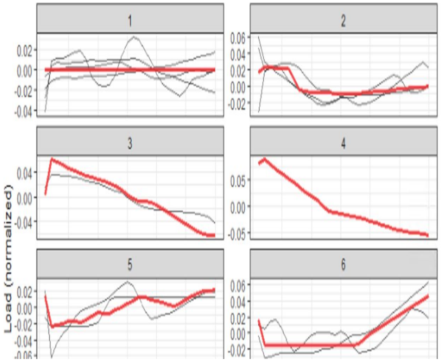


Clustering User Profiles

Direct Cluster on Point Values Using a Distance Metric

K-Medoids Partitional

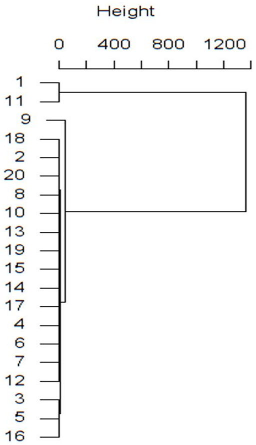
1. Extract trend from time series decomposition
2. Extract seasonal profiles (monthly) using mean seasonal profile model-based representation
3. Determine the optimal number of clusters using Davies_Bouldin index computation
4. Apply K-medoids partitional clustering



Panels 1, 2, 5 & 6 represent typical extracted profiles. Panels 3 & 4 are considered outliers because they contain the least number of customers. Red lines are medoids of clusters.

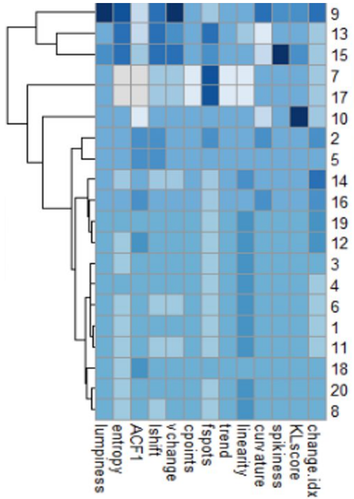
DTW Hierarchical

1. Compute distance between time series using Dynamic Time Warping as the dissimilarity measure
2. Apply Ward method, a minimum variance technique that minimizes within-cluster variance



Cluster on Global Features Extracted from Time Series

1. Extract 13 global features that capture the underlying structural characteristics of individual time series
2. Apply complete-linkage hierarchical clustering using Euclidean distance
-> Reduce the dimensionality of the time series and is much less sensitive to noisy data



*All 3 clustering examples were performed on meters 1-20 (each meter represents a household) on April 2019's data.

Comparing Results from Different Clustering

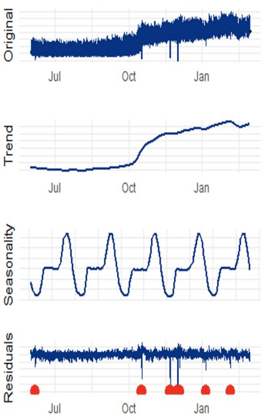
Algorithms	K-medoids	DTW Hierarchical	Features
# of clusters	6	3	NA
Outliers (meterno)	5, 7, 8, 9, 17, 18	1, 9, 11	9, 13, 15

- Meter 9 was consistently identified as an outlier by all 3 clustering algorithms - indicating a high probability of fraud
- Meters 5, 7, 8 showed consumption in the first half of the month but near 0 consumption in the later half - a trend linked to potential theft
- Meters 13, 15 had long periods of missing data
- By comparing and interpreting results obtained from different clustering algorithms, we were able to identify abnormal consumption trends and possible fraud.
- However, clustering results for theft detection varied depending on the algorithm applied, and since our dataset was unlabelled, we had no means of fine-tuning and evaluating our clustering algorithms.

Predicting Transformer Outages

Seasonal decomposition

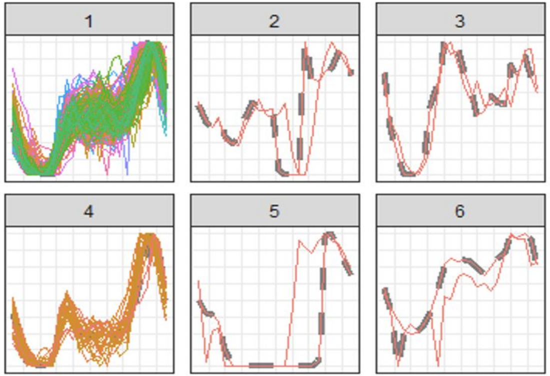
Seasonal decomposition was used to detect energy changes that deviated from the typical pattern. After subtracting the overall trend and daily seasonality, an output of residuals may be used to identify outliers. Three algorithms were used to this effect and their efficacies are shown below.



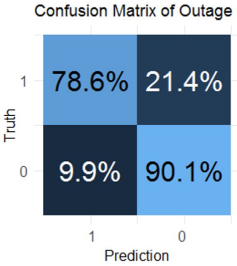
	Loess	Holt	ARIMA	Random
Mean PR-AUC	33.63	33.63	33.60	4
Std.Dev	1.69	1.69	1.70	

Applying seasonal decomposition on transformer energy data yielded PR-AUC values of 33.6% when compared to manually labeled data, an improvement over random choice (4%).

Clustering



Daily transformer output patterns may be clustered to detect anomalous outputs. Using a hierarchical with dynamic time warping (DTW) algorithm, outlier clusters were labeled as outages. The result yielded a sensitivity (percentage of outages correctly labeled) of 78.6%.



Total Truth:
1: 140
0: 3156
= 4.2%
Anomaly

Conclusions

- Our work serves to analyze energy quality and evaluate energy theft in Kyrgyzstan, Central Asia.
- Our work also serves as a proof of concept for applying unsupervised clustering techniques on time series data for market segmentation and anomaly detection.
- Statistical methods (e.g. seasonal decomposition) and machine learning methods (e.g. unsupervised clustering) can track transformer outages with high accuracy

Future Work

- To validate the model, the algorithms may be tested on a labeled dataset using ground-truth outage reports.
- Ideally, utility companies can use the suspect list generated by our clustering algorithms and inspect suspicious customers. The inspection records can then provide labels for our dataset, enabling supervised learning approaches.

References

[1] S. Depuru et al., Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft (2010) Energy Policy 39(2):1007-1015

[2] World Bank Group, *Analysis of the Kyrgyz Republic's Energy Sector*, May 2017.