Adaptive Time Series Forecasting of Energy Consumption using Optimized Cluster Analysis

Peter Laurinec, Marek Lóderer, Petra Vrablecová, Mária Lucká, Viera Rozinajová, Anna Bou Ezzeddine

12.12.2016

Slovak University of Technology in Bratislava

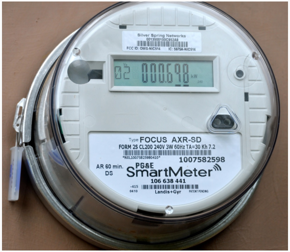
Motivation

More accurate forecast of electricity consumption is needed due to:

- Optimization of electricity consumption.
- Production of electricity. Overvoltage in grid.
- Distribution (utility) companies. Deregulation of the market. Purchase and sale of electricity.

Smart grids:

- Intelligent networks.
- Smart meters (every consumer has them).
- Usually 48 measurements per day.
- Advanced methods of forecast.



Forecast methods

Factors influencing electricity load:

- Seasonality (daily, weekly, ...)
- Weather (temperature, humidity, ...)
- Holidays
- Random effects

Methods:

- Time series analysis
- Regression
    - Linear model
    - AI methods
- Time series data mining + clustering

Clustering and forecast of load

Hypothesis:

Hypothesis:

- Clustering of consumers (creation of more predictable groups of consumers) improves forecast accuracy against simple aggregate forecast.

Clustering and forecast of load

Hypothesis:

- Clustering of consumers (creation of more predictable groups of consumers) improves forecast accuracy against simple aggregate forecast.

Target:

Clustering and forecast of load

Hypothesis:

- Clustering of consumers (creation of more predictable groups of consumers) improves forecast accuracy against simple aggregate forecast.

Target:

- Compare and find most suitable forecast methods that are suitable for the combination with clustering.

Data from smart meters

Available data come from smart meters from Ireland.

Data from smart meters

Available data come from smart meters from Ireland.

Ireland

Data from smart meters

Available data come from smart meters from Ireland.

Ireland

- 3639 consumers. Residences.

Data from smart meters

Available data come from smart meters from Ireland.

Ireland

- 3639 consumers. Residences.
- 48 measurements per day.

Data from smart meters

Available data come from smart meters from Ireland.

Ireland

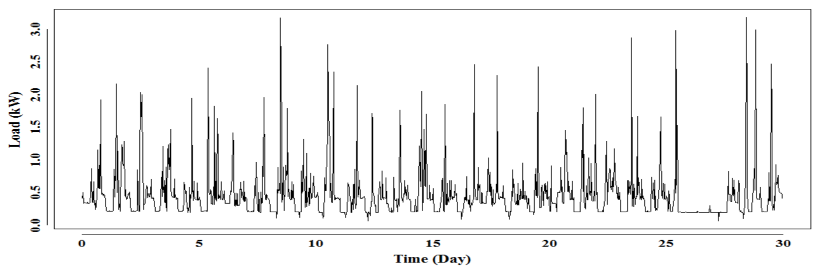
- 3639 consumers. Residences.
- 48 measurements per day.
- Test set from three months of year 2010 (February, May and September).

Data from smart meters

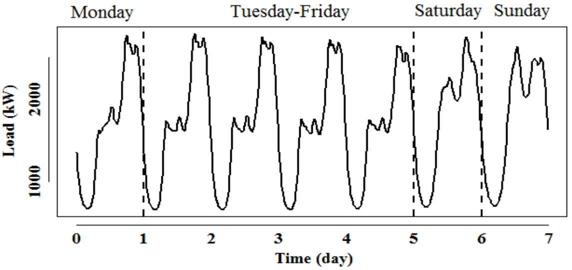
Available data come from smart meters from Ireland.

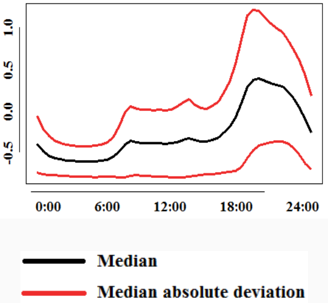
Ireland

- 3639 consumers. Residences.
- 48 measurements per day.
- Test set from three months of year 2010 (February, May and September).



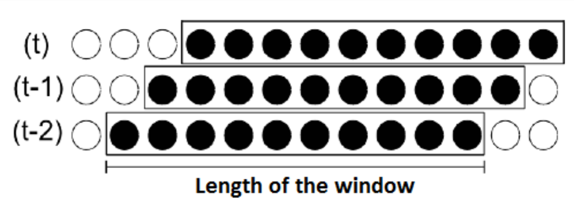
Median weekly and daily profile





What we want to do

Data come every day: batch processing (sliding window approach).



Batch of the length of two weeks (enough, fast).

Adaptability:

- Automatic selection of the number of clusters.

Proposed approach

Aggregation with clustering

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks

Proposed approach

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)
3. Computation of representations of time series (Compared 4 methods)

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)
3. Computation of representations of time series (Compared 4 methods)
4. Automatic determination of optimal number of clusters K (DB-index)

Proposed approach

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)
3. Computation of representations of time series (Compared 4 methods)
4. Automatic determination of optimal number of clusters K (DB-index)
5. Clustering of representations (K-means with centroids initialization by K-means++)

Proposed approach

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)
3. Computation of representations of time series (Compared 4 methods)
4. Automatic determination of optimal number of clusters K (DB-index)
5. Clustering of representations (K-means with centroids initialization by K-means++)
6. Summation of K time series by found clusters

Proposed approach

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)
3. Computation of representations of time series (Compared 4 methods)
4. Automatic determination of optimal number of clusters K (DB-index)
5. Clustering of representations (K-means with centroids initialization by K-means++)
6. Summation of K time series by found clusters
7. Training of K forecast models and the following forecast (Compared 10 methods)

Proposed approach

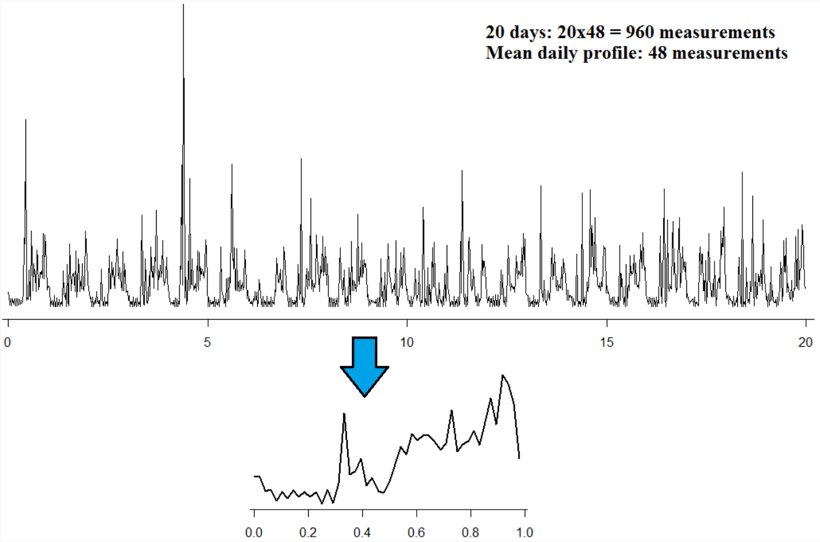
Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)
3. Computation of representations of time series (Compared 4 methods)
4. Automatic determination of optimal number of clusters K (DB-index)
5. Clustering of representations (K-means with centroids initialization by K-means++)
6. Summation of K time series by found clusters
7. Training of K forecast models and the following forecast (Compared 10 methods)
8. Summation of forecasts and evaluation

Aggregation with clustering

1. Set of time series of electricity consumption of the length of two weeks
2. Normalization (z-score)
3. Computation of representations of time series (Compared 4 methods)
4. Automatic determination of optimal number of clusters K (DB-index)
5. Clustering of representations (K-means with centroids initialization by K-means++)
6. Summation of K time series by found clusters
7. Training of K forecast models and the following forecast (Compared 10 methods)
8. Summation of forecasts and evaluation
9. Remove first day and add new one to the training window (sliding window approach), go to step 1

Representations of time series



Representations of time series

Let T be a time series of length n, representation of T is a model $\hat{T}$ of reduced dimensionality d ($d \ll n$) such that $\hat{T}$ closely approximates T.

Why time series representations?

Representations of time series

Let T be a time series of length n, representation of T is a model $\hat{T}$ of reduced dimensionality d ($d \ll n$) such that $\hat{T}$ closely approximates T.

Why time series representations?

1. Reduce memory load.

Representations of time series

Let T be a time series of length n, representation of T is a model $\hat{T}$ of reduced dimensionality d ($d \ll n$) such that $\hat{T}$ closely approximates T.

Why time series representations?

1. Reduce memory load.
2. Accelerate subsequent machine learning algorithms.

Representations of time series

Let T be a time series of length n, representation of T is a model $\hat{T}$ of reduced dimensionality d ($d \ll n$) such that $\hat{T}$ closely approximates T.

Why time series representations?

1. Reduce memory load.
2. Accelerate subsequent machine learning algorithms.
3. Implicitly remove noise from the data.

Representations of time series

Let T be a time series of length n, representation of T is a model $\hat{T}$ of reduced dimensionality d ($d \ll n$) such that $\hat{T}$ closely approximates T.

Why time series representations?

1. Reduce memory load.
2. Accelerate subsequent machine learning algorithms.
3. Implicitly remove noise from the data.
4. Emphasize the essential characteristics of the data.

Representations of time series

Let T be a time series of length n, representation of T is a model $\hat{T}$ of reduced dimensionality d ($d \ll n$) such that $\hat{T}$ closely approximates T.

Why time series representations?

1. Reduce memory load.
2. Accelerate subsequent machine learning algorithms.
3. Implicitly remove noise from the data.
4. Emphasize the essential characteristics of the data.

Model based representations:

Representations of time series

Let T be a time series of length n, representation of T is a model $\hat{T}$ of reduced dimensionality d ($d \ll n$) such that $\hat{T}$ closely approximates T.

Why time series representations?

1. Reduce memory load.
2. Accelerate subsequent machine learning algorithms.
3. Implicitly remove noise from the data.
4. Emphasize the essential characteristics of the data.

Model based representations:

- Suitable for seasonal time series

Model based methods

- Representations based on statistical model.
- Extraction of regression coefficients $\Rightarrow$ creation of daily profiles.
- Creation of representation which is long as frequency of time series (48).

$$x_i = \beta_1 u_{i1} + \beta_2 u_{i2} + \cdots + \beta_{seas} u_{iseas} + \varepsilon_i, \; where \;\; i = 1, \ldots, n$$

New representation: $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_{seas})$.

Applied methods:
Robust Linear Model. Generalized Additive Model (smoothing function).

- Triple Holt-Winters Exponential Smoothing. Last seasonal coefficients as representation.

    1. Smoothing factors were set automatically.

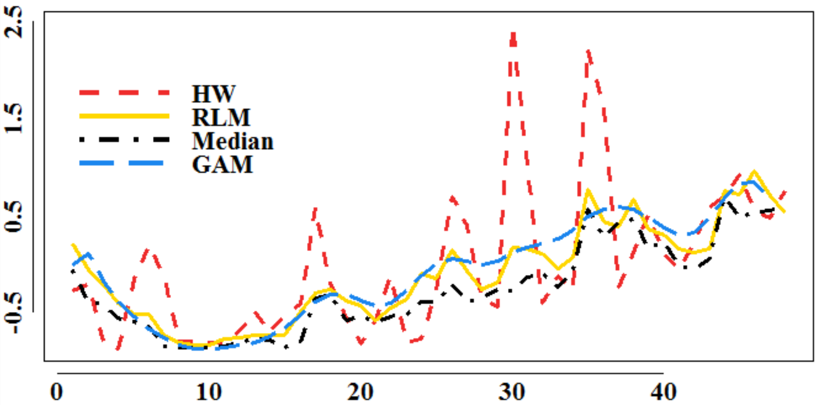
Model based methods

- Triple Holt-Winters Exponential Smoothing. Last seasonal coefficients as representation.

    1. Smoothing factors were set automatically.

- Median daily profile.

Comparison of model based representations



HW - Holt-Winters, RLM - Robust Linear Model, Median - Median daily profile, GAM - Generalized Additive Model.

Forecast methods

Ten methods:

- Double Seasonal Holt-Winters Exponential Smoothing
- STL decomposition + Exponential Smoothing
- STL decomposition + ARIMA
- Support Vector Regression (SVR)
- Random Forest
- Extreme Gradient Boosting (xgboost)
- Extremely Randomized Trees
- Bagging
- Multilayer Perceptron
- Deep Learning

Evaluation of forecast

The accuracy of the forecast of electricity consumption was measured by MAPE (Mean Absolute Percentage Error).

$$\mathrm{MAPE} = 100 \times \frac{1}{n} \sum_{t=1}^{n} \frac{|x_t - \bar{x}_t|}{x_t},$$

where x_t is a real consumption, $\bar{x}_t$ is a forecasted load and n is a length of the time series.

Experiments

Setup:

- Sliding window of 14 days.
- Number of clusters from 8 to 18.
- One day ahead forecast.
- One model for all days during the week and separate models for workdays and weekdays.
- Various lengths of training windows for forecast methods.
- Various features (inputs) to forecast models. Dummy time variables (SVR), time of the day and day of the week, sinus and cosinus form of time, lagged load (denoised).

- Mean values of MAPEs $\rightarrow$

Results

	Sum		Median		GAM	
Method Name	MAPE	window	MAPE	window	MAPE	window
DSHW	4.577	14	**4.304**	14	**4.279**	14
STL + ARIMA	7.310	9	**7.008**	8	**6.968**	8
STL + ETS	11.014	4	**7.767**	8	**7.672**	8
SVR	4.206	14	**4.046**	14	**4.114**	14
RF	4.26	9	**4.193**	9	**4.19**	9
XGB	4.015	14	4.004	14	4.007	14
ExRT	3.758	14	**3.754**	14	**3.754**	14
Bagg	3.960	14	**3.818**	14	**3.795**	14
MLPnet	5.776	14	**5.227**	9	**5.229**	9
DLnet	4.473	14	4.395	14	4.433	14
Mean	5.054		4.869		4.879	

Results - with separation of days

	Sum		Median		GAM	
Method Name	MAPE	window	MAPE	window	MAPE	window
STL + ARIMA	4.317	4	**3.982**	6	**4.001**	6
STL + ETS	6.024	4	**4.813**	4	**4.804**	3
SVR	3.828	10	3.886	10	3.819	10
RF	3.855	5	**3.817**	5	**3.805**	5
XGB	4.081	10	**4.023**	10	**4.030**	10
ExRT	3.800	10	**3.796**	10	3.797	10
Bagg	3.822	10	**3.685**	10	3.680	10
Mean	4.003		3.924		3.913	

Conclusion

- Clustering of consumers can improve forecast accuracy.
- Significant improvement: DSHW, STL + ARIMA, STL + ETS, RandomForest, Bagging and MLPnet.
- Same results as simple aggregation most of times: SVR, ExtraTrees, xgboost and DLnet.
- We have shown that the best representations are RLM, Median and GAM.
- Average forecast error of forecasting methods with separated workdays and weekend models was lowered by 0.852% in comparison with methods without model separation.