SAS® GLOBAL FORUM 2019

USERS PROGRAM

APRIL 28 – MAY 1, 2019 | DALLAS, TX

Krzysztof Gajowniczek

- Ph.D. in Computer Science (data mining, machine learning) from Systems Research Institute, Polish Academy of Sciences.
- Postdoctoral researcher at 6 institutions such as University of Texas (Austin), University of California (Los Angeles) or Swinburne University of Technology (Melbourne).
- Author of 31 scientific articles.
- Principal investigator of 9 research projects.
- Reviewer of 120 scientific articles.

Smart Meters to Support Energy Efficiency on the Individual Household Level

Faculty of Applied Informatics and Mathematics

Warsaw University of Life Sciences

Presentation outline

Duration of each part

- Main presentation - approx. 45 min.
- QA session - approx. 15 min.

Presentation outline

Scope

- Background and motivation.
- Research questions.
- Data characteristics.
- Detecting household activity patterns.
- Load forecasting on the individual household level.
- Summary and conclusions.

Background and motivation

Electricity

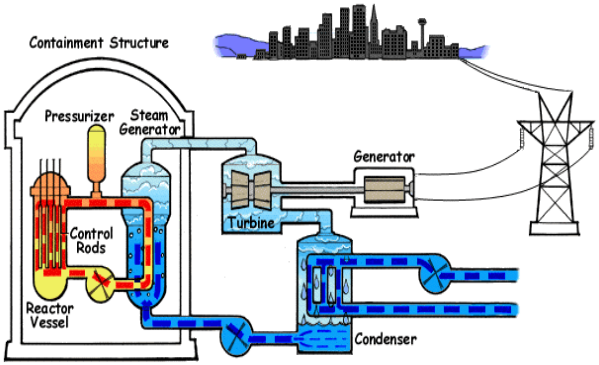
- Electric energy is created by the flow of electrons, often called "current," through a conductor, such as a wire.
- The amount of electric energy created depends on the number of electrons flowing and the speed of the flow.
- Energy can either be potential or kinetic.

Background and motivation

Nuclear Energy

PROS:

- Nuclear power plants provide a stable base load of energy.
- Low pollution that can contaminate the environment
- Energy released in a nuclear fussion reaction is ten million times greater than the amount released in burning a fossil fuel.



CONS:

- The initial construction costs of nuclear power plants are large.
- Accidents happen.
- Radioactive waste.

Background and motivation

Fossil Fuels (Coal, Oil or Petroleum, Natural Gas)

PROS:

- Available in Plenty - we have already relied on fossil fuels for many years now.
- Fossil fuels are extremely efficient.
- Fossil fuels are actually very easy to find.

CONS:

- Environmental degradation.
- Fossil fuels are a finite energy resource.
- Rising cost.
- Public health issues - fossil fuels are not at all environment friendly.

Background and motivation

Wind Power

PROS:

- No pollution that can contaminate the environment.
- Renewable source of energy.
- Wind farms can be built offshore.
- Farming and grazing can still take place on land occupied by wind turbines.

CONS:

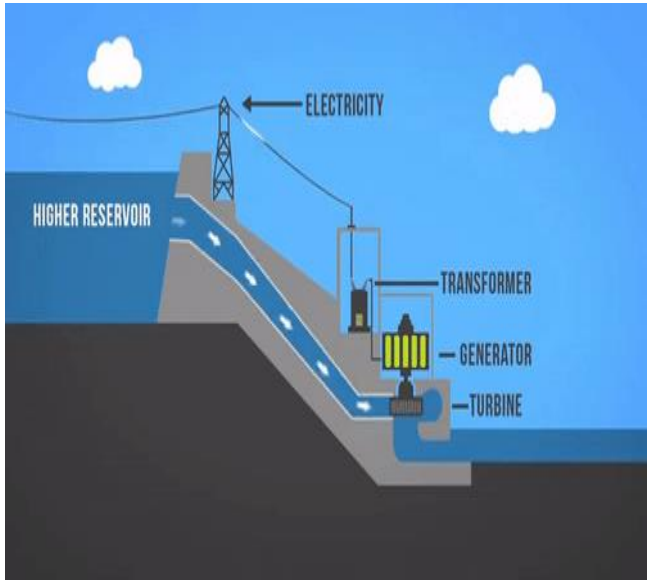
- Wind power is intermittent.
- Large wind farms can have a negative effect on the scenery.

Background and motivation

Hydroelectric Energy

PROS:

- No pollution that can contaminate the environment.
- Water can be accumulated above the dam and released to coincide with peaks in demand.
- Water used for hydropower can be reused.



CONS:

- Dams can be very expensive to build.
- There needs to be a sufficient and powerful enough supply of water.

Background and motivation

Solar Power

PROS:

- No pollution that can contaminate the environment.
- Solar power is a renewable resource.

CONS:

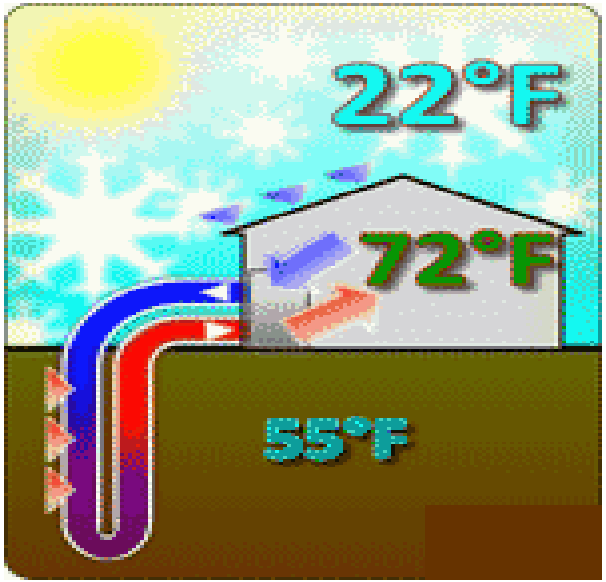
- Solar power does not produce energy if the sun is not shining.
- Solar power stations can be very expensive to build.

Background and motivation

Geothermal Energy

PROS:

- No pollution that can contaminate the environment – if done correctly.
- Once a geothermal plant is built, it is generally self-sufficient energy wise.

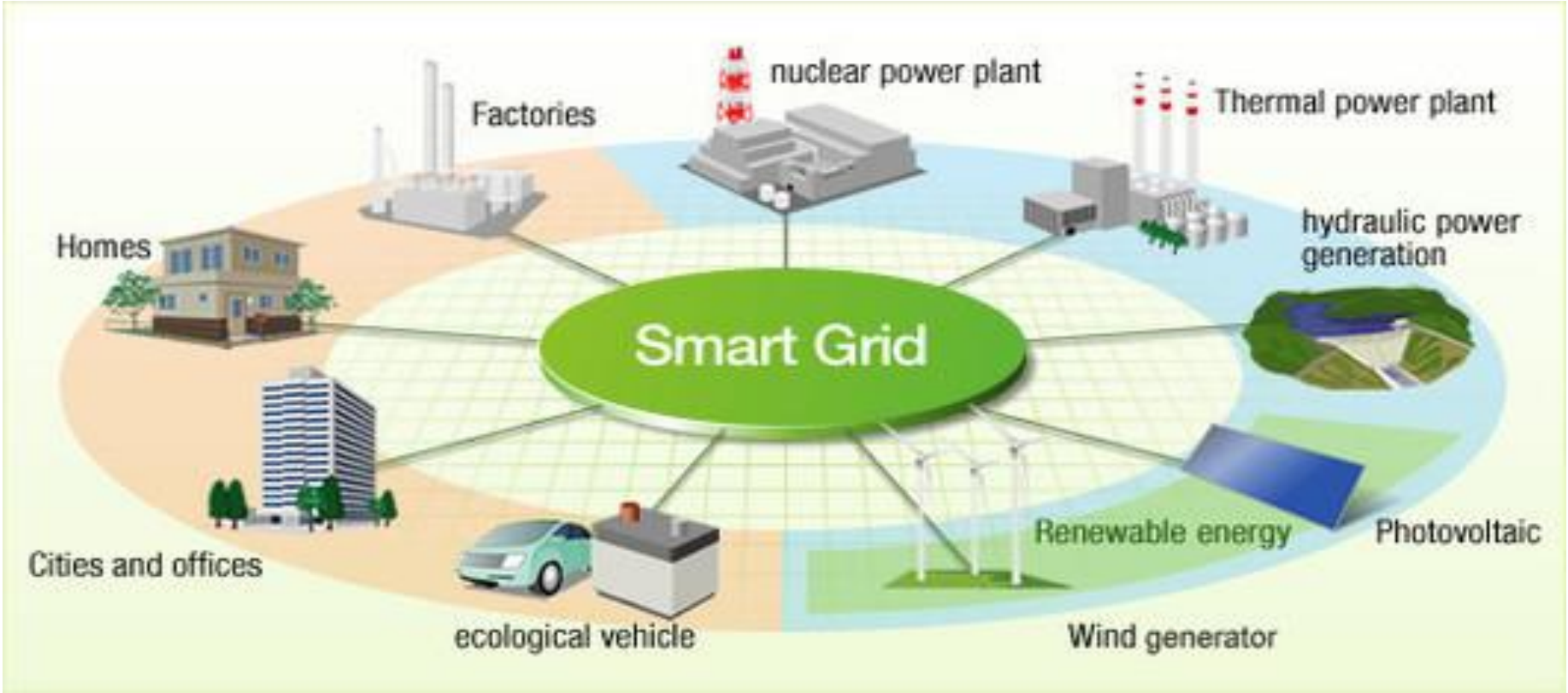


CONS:

- If done incorrectly, geothermal energy can produce pollutants.
- Improper drilling into the earth can release hazardous minerals and gases.

Background and motivation

Smart Grid



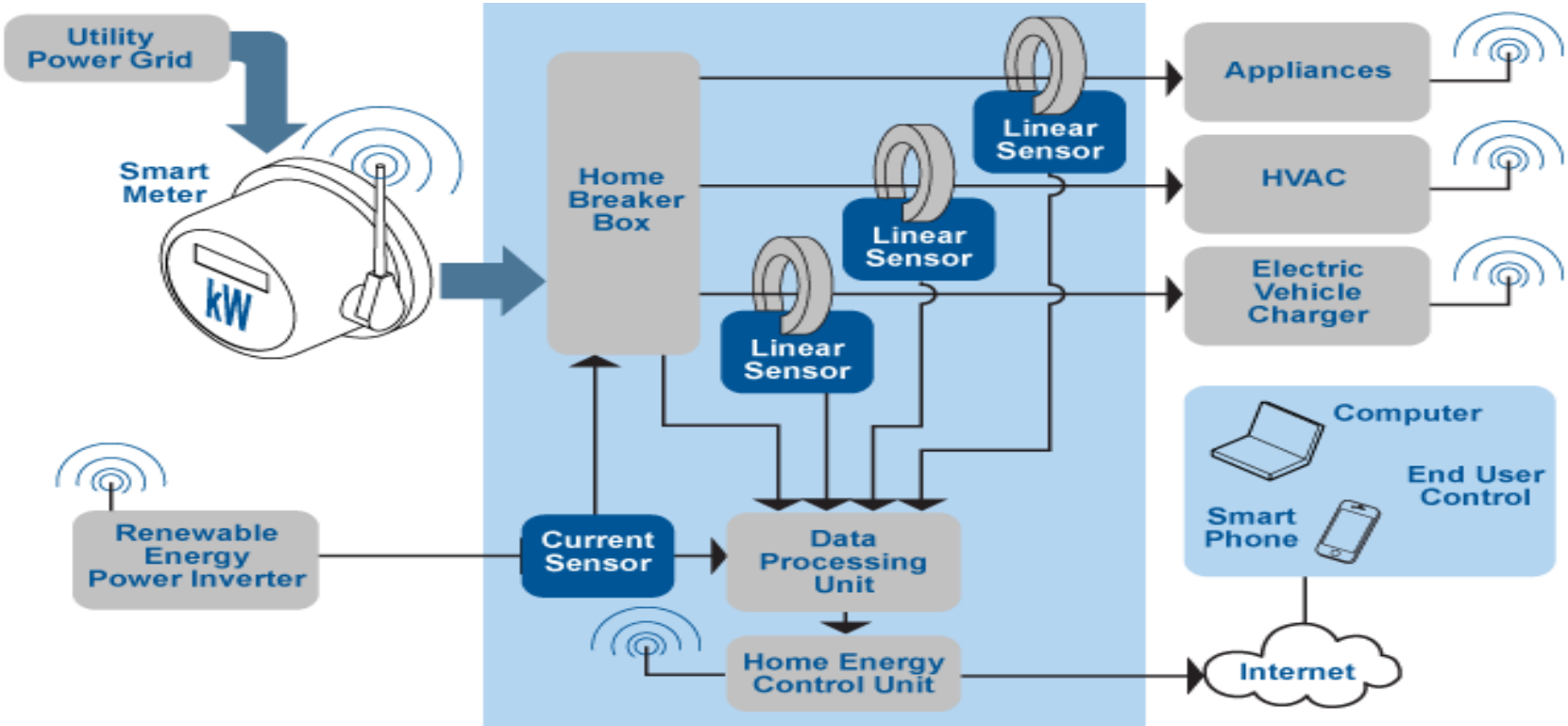
Background and motivation

Smart Grid

- Allow consumers to play a part in optimizing the operation of the system.
- Provide consumers with greater information and options for how they use their supply.
- Significantly reduce the environmental impact of the whole electricity supply system.
- Maintain or even improve the existing high levels of system reliability, quality and security of supply.

Background and motivation

Smart Metering



Background and motivation

Smart Metering - advantages for Electric Companies

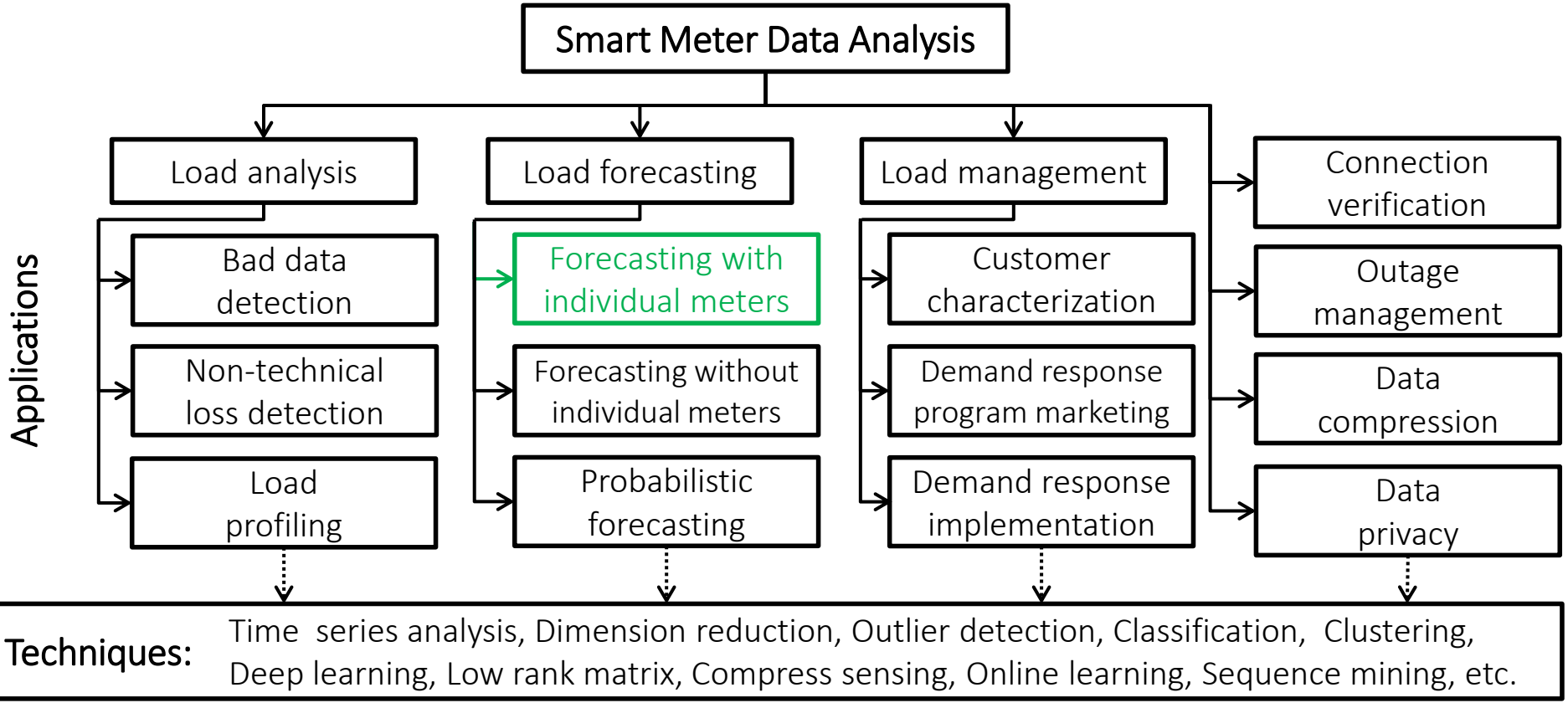
- Eliminates manual monthly meter readings.
- Monitors the electric system much more quickly.
- Makes it possible to use power resources more efficiently.
- Provides real-time data that is useful for balancing electric loads while reducing power outages (blackouts).
- Enables dynamic pricing, which raises or lowers the cost of electricity based on demand.
- Avoids the capital expense of building new power plants.
- Helps to optimize income with existing resources.

Background and motivation

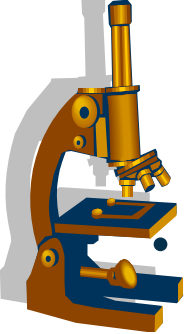
Smart Metering - advantages for Users of Electricity

- Far greater (and more detailed) feedback regarding energy use.
- Enable consumers to adjust their habits in order to lower electricity bills.
- Reduces the number of blackouts and system-wide electricity failures.

Background and motivation



Source: [1] Wang, Y., Chen, Q., Hong, T., & Kang, C. (2018). Review of smart meter data analytics: Applications, methodologies, and challenges. IEEE Transactions on Smart Grid.

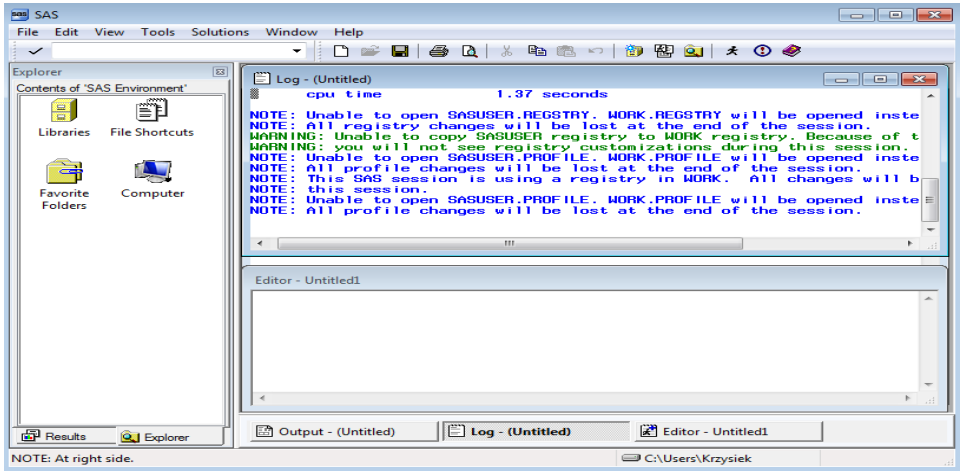


Research questions

- Is it possible to provide accurate load forecasting for 24 hours on the individual household level and to what extent?
- Are the clustering and sequence recognition algorithms good tools for identifying patterns of household behavior?
- Do the usage pattern variables of the household enhance the forecasting accuracy of individual consumer loads?
- What kind of forecasting methods and algorithms are appropriate to address high volatility data?

Software

- SAS Base 9.4:
  - ✓ 4 GL.
  - ✓ SQL.
  - ✓ MACRO.
- SAS Enterprise Miner Workstation 13.1.



Data characteristics

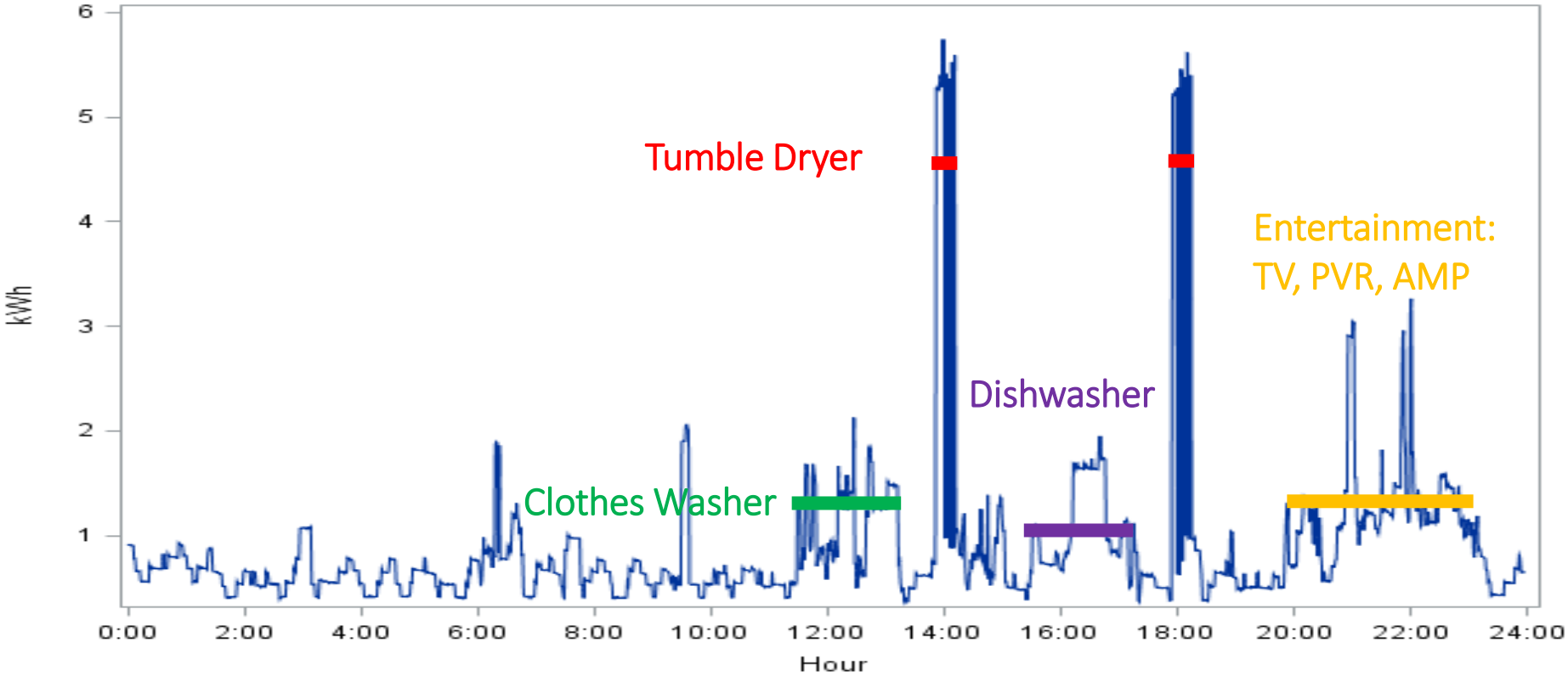
Households locations

- Vancouver, Canada (one household):
  - ✓Almanac of Minutely Power dataset (AMPds) [2].
  - ✓Recorded energy consumption data (at one minute intervals) using 21 sub-meters.
  - ✓Time span between April 1st, 2012, and March 31st, 2014.
- Austin, USA (46 households):
  - ✓WikiEnergy dataset, constructed by Pecan Street Inc. [3].
  - ✓Recorded energy consumption data (at one hour intervals) using 24 sub-meters.
  - ✓Time span between March 1st, 2013, and April 30th, 2014.

[2] Makonin, S., Popowich, F., Bartram, L., Gill, B., & Bajić, I. V. (2013). AMPds: A public dataset for load disaggregation and eco-feedback research. In 2013 IEEE Electrical Power & Energy Conference (pp. 1-6). IEEE.
[3] Pecanstreet, "Dataport," 2014, https://dataport.pecanstreet.org/.

Data characteristics

Daily load dissagregation



Detecting household activity patterns

Applinces scope

- The analysis was narrowed to the most energy-intensive household appliances.
- These were: **Tumble Dryer, Clothes Washer, Dishwasher, Kettle and Microwave.**
- The other appliances were not considered due to their **insignificant activities** (e.g., Basement Plugs and Lights), **continuous activity** (e.g., Security Equipment), or those **not showing any repetitive patterns** (e.g., Electronics Workbench).

Detecting household activity patterns

Activity segmentation – definitions and objective

- Hierarchical cluster analysis is an algorithmic approach to find discrete groups with varying degrees of similarity in a data set represented by a similarity matrix.
- These groups are hierarchically organized as the algorithms proceed and may be presented as a dendrogram.
- One of the most popular agglomerative clustering algorithm is Ward's method.
- The purpose of this analysis is to discover similar profiles or, in other words, appliances with similar switch ON probability distribution through the whole day.

Detecting household activity patterns

Activity segmentation – data preparation

```
libname a "D:\SAS_Global_2019";

proc import
    datafile = "D:\SAS_Global_2019\ampds.csv"
    dbms = csv
    out = a.ampds
    replace;
    delimiter = ",";
run;
```

VIEWTABLE: A.Ampds

	Hour	kWh	Kettle	Microwave	Washingmachine	Tumbledryer	Dishwasher
1	0	0.1357891108	0	0	0	0	0
2	1	0.1764955718	0	0	0	0	0
3	2	0.1326900715	0	0	0	0	0
4	3	0.1591732171	0	0	0	0	0
5	4	0.1603281302	0	0	0	0	0
6	5	0.1672249343	0	0	0	0	0
7	6	0.1375887459	0	0	0	0	0
8	7	0.1449635967	0	0	0	0	0
9	8	0.1646980525	1	0	0	0	0
10	9	0.3085970964	2	2	0	0	0
11	10	0.6670214077	1	0	0	0	1
12	11	0.927971697	2	0	0	0	1
13	12	0.4448989927	0	0	0	0	0
14	13	0.2374523244	0	0	0	0	0
15	14	0.1700440603	0	0	0	0	0
16	15	0.3033420571	0	0	0	0	0
17	16	0.3918140895	0	0	0	0	0
18	17	0.4070038825	0	0	0	0	0
19	18	0.5118122648	0	4	0	0	0
20	19	0.435109534	2	0	0	0	0
21	20	0.9542875942	0	0	1	0	0
22	21	0.4069515342	0	0	2	1	0
23	22	1.5124483225	0	2	1	1	0
24	23	0.3133371571	0	0	0	0	0
25	0	0.2425246064	0	0	0	0	0

Detecting household activity patterns

Activity segmentation – data preparation

- The starting point for the usage pattern detection was to prepare a matrix with switching on probabilities for each of the individual devices over a specified time period.
- The probabilities were estimated using the following formula:

$$P = \frac{Number\ of\ turn\ ON\ events\ in\ hour\ i}{Total\ number\ of\ turn\ ON\ events}.$$

Detecting household activity patterns

Activity segmentation – data preparation

```
proc sql;
create table a.ampds as
select
   hour,
   coalesce( sum(Kettle) / (select sum(Kettle) from a.ampds), 0) as Kettle,
   coalesce( sum(Microwave ) / (select sum(Microwave ) from a.ampds), 0) as Microwave,
   coalesce( sum(Washingmachine ) / (select sum(Washingmachine ) from a.ampds), 0) as Washingmachine,
   coalesce( sum(Tumbledryer) / (select sum(Tumbledryer) from a.ampds), 0) as Tumbledryer,
   coalesce( sum(Dishwasher) / (select sum(Dishwasher) from a.ampds), 0) as Dishwasher
from
   a.ampds
group by
   hour;
quit;
```

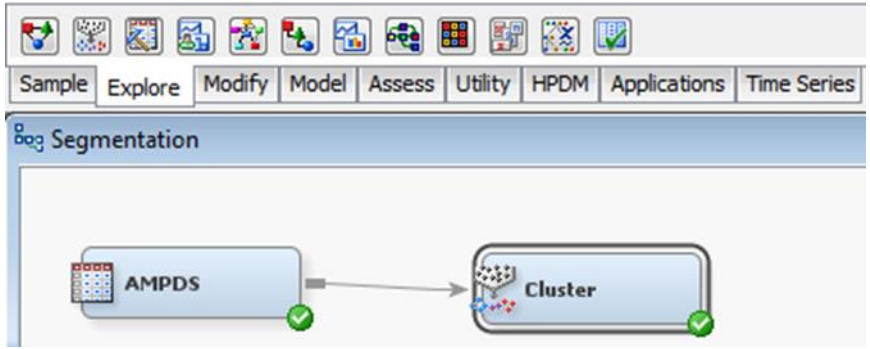
Detecting household activity patterns

VIEWTABLE: A.Ampds_

	hour	Kettle	Microwave	Washingmachine	Tumbledryer	Dishwasher
1	0	0.00	0.00	0.02	0.06	0.00
2	1	0.00	0.00	0.00	0.02	0.00
3	2	0.00	0.00	0.00	0.00	0.00
4	3	0.00	0.00	0.02	0.00	0.00
5	4	0.00	0.00	0.00	0.00	0.00
6	5	0.01	0.00	0.00	0.00	0.00
7	6	0.03	0.03	0.00	0.00	0.00
8	7	0.13	0.12	0.02	0.00	0.08
9	8	0.07	0.07	0.06	0.00	0.07
10	9	0.10	0.08	0.03	0.00	0.07
11	10	0.06	0.06	0.06	0.08	0.08
12	11	0.06	0.03	0.06	0.06	0.13
13	12	0.05	0.01	0.08	0.04	0.05
14	13	0.05	0.02	0.05	0.06	0.05
15	14	0.05	0.02	0.05	0.04	0.07
16	15	0.02	0.01	0.06	0.08	0.07
17	16	0.04	0.03	0.06	0.08	0.07
18	17	0.04	0.04	0.03	0.08	0.03
19	18	0.04	0.12	0.03	0.06	0.05
20	19	0.07	0.04	0.09	0.02	0.03
21	20	0.10	0.17	0.08	0.08	0.02
22	21	0.05	0.10	0.09	0.10	0.08
23	22	0.03	0.06	0.08	0.12	0.03
24	23	0.01	0.00	0.05	0.06	0.03

Detecting household activity patterns

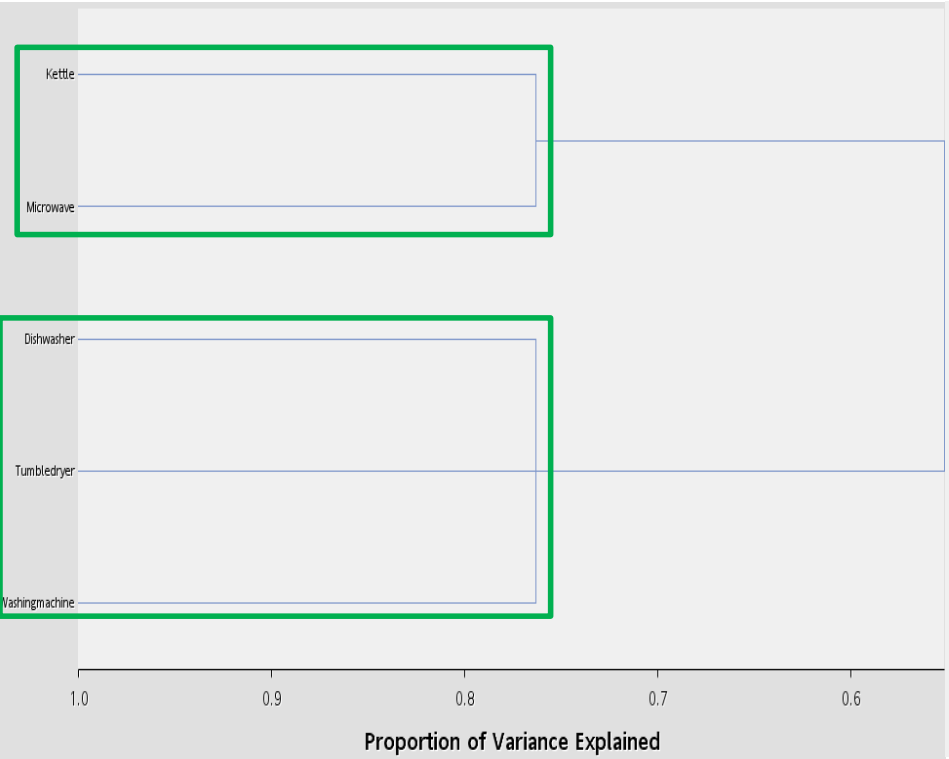
Activity segmentation – analysis settings

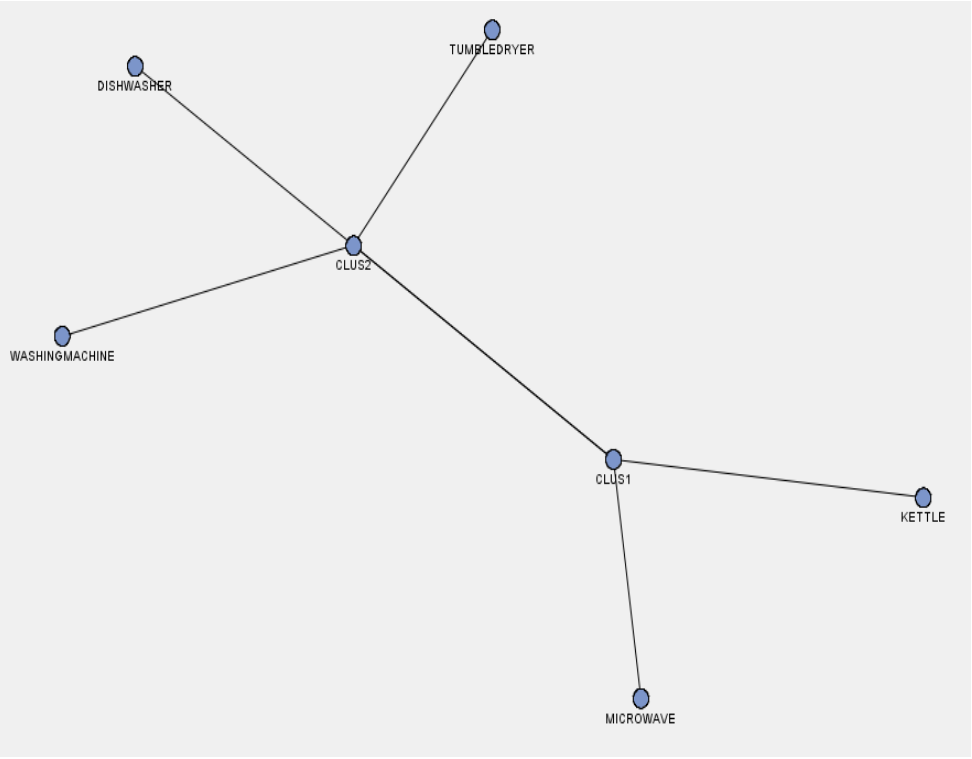


Train	
Variables	
Cluster Variable Role	Segment
Internal Standardization	Standardization
Number of Clusters	
Specification Method	Automatic
Maximum Number of Cluster	10
Selection Criterion	
Clustering Method	Ward
Preliminary Maximum	50
Minimum	2
Final Maximum	20
CCC Cutoff	3
Encoding of Class Variables	
Ordinal Encoding	Rank
Nominal Encoding	GLM
Initial Cluster Seeds	
Seed Initialization Method	Default
Minimum Radius	0.0
Drift During Training	No
Training Options	
Use Defaults	Yes
Settings	
Missing Values	
Interval Variables	Default
Nominal Variables	Default
Ordinal Variables	Default
Scoring Imputation Method	None
Score	
Cluster Variable Role	Segment
Hide Original Variables	Yes
Cluster Label Editor	

Detecting household activity patterns

Activity segmentation – results





Detecting household activity patterns

Activity segmentation – results

- Switch ON probability of the kettle and the microwave at certain times are very similar. In particular, it can be observed between 7 am and 9 am, which is usually associated with the users' activity related with breakfast preparation.
- A similar correlation in periods of joint work, can be seen in the case of washing machine and tumble dryer. In the investigated households, there is a logical relationship taking washing first and then drying the washed clothes.

Detecting household activity patterns

Activity sequence mining

- Sequence mining is an exploration technique that focuses on discovering statistically relevant patterns in the form of a sequence for a given data set.
- The resulting rules (patterns) adopt the following form of conditional statements: **if appliance A was used, then appliance B will be used next**.
- This kind of analysis gives insight that can help understand how power consumption is influenced by certain activities and their sequences and how those activities are related to each other.

Detecting household activity patterns

Activity sequence mining – evaluation measures

- Support is defined as the proportion of days containing a specific itemset of the appliances to all of the days:

$$supp(appliance\ A \rightarrow appliance\ B) = \frac{|appliance\ A \rightarrow appliance\ B|}{|all\ days|}$$

Detecting household activity patterns

Activity sequence mining – evaluation measures

- Confidence is defined as the proportion of the observed support of the specific rule to the support of the left side item (corresponds to the conditional probability denoting if the left side occurred then also with some probability the right side of rule will occur):

$$conf(appliance\ A \rightarrow appliance\ B) = \frac{supp(appliance\ A \rightarrow appliance\ B)}{supp(appliance\ A)}$$

Detecting household activity patterns

Activity sequence mining – evaluation measures

- Lift is defined as the proportion of the observed support of the rule to the product of the supports of both sides of the rule, and it shows, in business terms, how many times more likely the appearance of appliance B with appliance A is than that with any other randomly chosen appliance:

$$lift(appliance\ A \rightarrow appliance\ B) = \frac{supp(appliance\ A \rightarrow appliance\ B)}{supp(appliance\ A) \cdot supp(appliance\ B)}$$

Detecting household activity patterns

Activity sequence mining – data preparation

```
proc transpose
  data = a.sequence
  out = a.sequence_EM
  name = Appliance;
  by Date Hour;
  var Kettle
      Microwave
      Washingmachine
      Tumbledryer
      Dishwasher;
run;
```

VIEWTABLE: A.Sequence_em

	Date	Hour	Appliance	COL1
30	1	5	Dishwasher	0
31	1	6	Kettle	0
32	1	6	Microwave	0
33	1	6	Washingmachine	0
34	1	6	Tumbledryer	0
35	1	6	Dishwasher	0
36	1	7	Kettle	0
37	1	7	Microwave	0
38	1	7	Washingmachine	0
39	1	7	Tumbledryer	0
40	1	7	Dishwasher	0
41	1	8	Kettle	1
42	1	8	Microwave	0
43	1	8	Washingmachine	0
44	1	8	Tumbledryer	0
45	1	8	Dishwasher	0
46	1	9	Kettle	2
47	1	9	Microwave	2
48	1	9	Washingmachine	0
49	1	9	Tumbledryer	0
50	1	9	Dishwasher	0
51	1	10	Kettle	1
52	1	10	Microwave	0
53	1	10	Washingmachine	0
54	1	10	Tumbledryer	0
55	1	10	Dishwasher	1
56	1	11	Kettle	2
57	1	11	Microwave	0
58	1	11	Washingmachine	0

Detecting household activity patterns

Activity sequence mining – data preparation

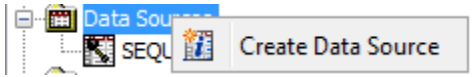
```
data a.sequence_EM;
   set a.sequence_EM;
   where COL1 > 0;
   drop COL1;
run;
```

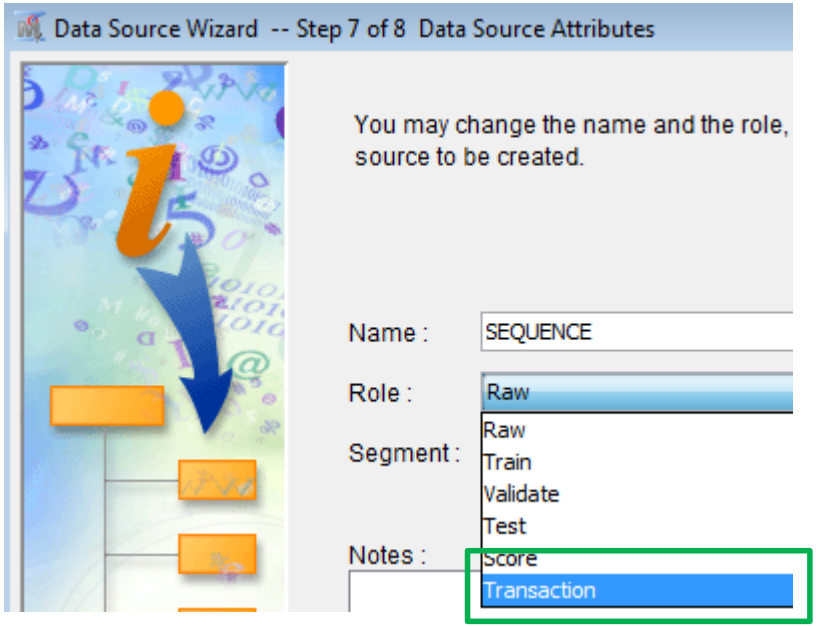
VIEWTABLE: A.Sequence_em

	Date	Hour	Appliance
1	1	8	Kettle
2	1	9	Kettle
3	1	9	Microwave
4	1	10	Kettle
5	1	10	Dishwasher
6	1	11	Kettle
7	1	11	Dishwasher
8	1	18	Microwave
9	1	19	Kettle
10	1	20	Washingmachine
11	1	21	Washingmachine
12	1	21	Tumbledryer
13	1	22	Microwave
14	1	22	Washingmachine
15	1	22	Tumbledryer
16	2	10	Kettle
17	2	10	Microwave
18	2	10	Tumbledryer
19	2	10	Dishwasher
20	2	11	Tumbledryer
21	2	11	Dishwasher
22	2	12	Kettle
23	2	13	Microwave
24	2	19	Washingmachine
25	2	20	Microwave
26	2	20	Washingmachine
27	2	21	Kettle
28	2	21	Microwave
29	2	21	Tumbledryer
30	2	22	Kettle
31	2	23	Tumbledryer
32	3	7	Kettle
33	3	8	Kettle
34	3	9	Kettle

Detecting household activity patterns

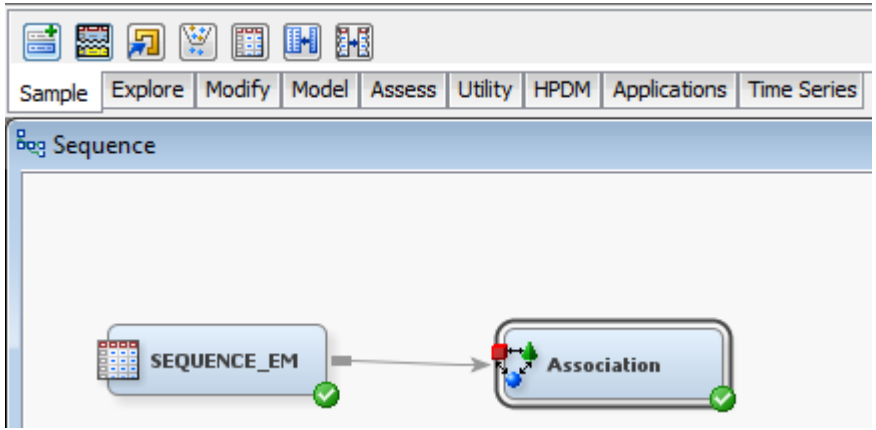
Activity sequence mining – analysis settings

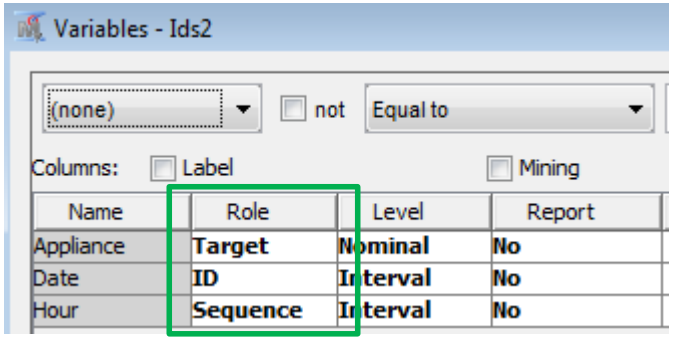




Detecting household activity patterns

Activity sequence mining – analysis settings





General	
Node ID	Assoc
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Maximum Number of Items	100000
Rules	
Association	
Maximum Items	4
Minimum Confidence Level	10
Support Type	Percent
Support Count	.
Support Percentage	5.0
Sequence	
Chain Count	6
Consolidate Time	0.0
Maximum Transaction Dura	0.0
Support Type	Percent
Support Count	.
Support Percentage	1.0
Rules	
Number to Keep	10000
Sort Criterion	Default
Number to Transpose	1000
Export Rule by ID	No
Recommendation	No

Detecting household activity patterns

Activity sequence mining – results

EMWS1.Assoc_RULES

	Support(%)	Confidence(%)	PseudoLift	Rule
1	5.0	66.66666666666666	8.888888888888888	Kettle & Washingmachine ==> Kettle ==> Washingmachine & Tumbledryer ==> Kettle & Washingmachine & Tumbledryer
2	5.0	66.66666666666666	8.888888888888888	Dishwasher ==> Kettle & Dishwasher ==> Washingmachine & Tumbledryer ==> Microwave & Washingmachine & Tumbledryer
3	5.0	66.66666666666666	8.888888888888888	Dishwasher ==> Washingmachine ==> Kettle & Washingmachine ==> Kettle & Washingmachine & Tumbledryer
4	5.0	66.66666666666666	8.888888888888888	Kettle & Washingmachine ==> Kettle ==> Washingmachine ==> Kettle & Washingmachine & Tumbledryer
5	5.0	66.66666666666666	8.888888888888888	Kettle & Dishwasher ==> Washingmachine ==> Kettle & Washingmachine ==> Kettle & Washingmachine & Tumbledryer
6	7.5	60.0	8.0	Kettle ==> Kettle ==> Kettle ==> Kettle ==> Washingmachine & Tumbledryer ==> Microwave & Washingmachine & Tumbledryer
7	7.5	60.0	8.0	Kettle ==> Kettle ==> Kettle ==> Washingmachine ==> Washingmachine & Tumbledryer ==> Microwave & Washingmachine & Tumbledryer
8	5.0	40.0	8.0	Kettle & Dishwasher ==> Dishwasher ==> Washingmachine ==> Kettle & Microwave & Tumbledryer
9	5.0	40.0	8.0	Kettle & Microwave ==> Tumbledryer ==> Washingmachine ==> Kettle & Microwave & Tumbledryer
10	5.0	40.0	8.0	Kettle & Dishwasher ==> Tumbledryer ==> Kettle ==> Kettle & Microwave & Tumbledryer
11	5.0	50.0	6.666666666666666	Microwave & Washingmachine ==> Washingmachine ==> Kettle & Washingmachine & Tumbledryer
12	5.0	50.0	6.666666666666666	Dishwasher ==> Kettle ==> Kettle ==> Washingmachine & Tumbledryer ==> Microwave & Washingmachine & Tumbledryer
13	5.0	50.0	6.666666666666666	Kettle & Dishwasher ==> Washingmachine ==> Washingmachine & Tumbledryer ==> Microwave & Washingmachine & Tumbledryer
14	5.0	50.0	6.666666666666666	Kettle ==> Kettle & Dishwasher ==> Kettle ==> Washingmachine & Tumbledryer ==> Microwave & Washingmachine & Tumbledryer

Detecting household activity patterns

Activity sequence mining – results

Sequence	Support	Confidence	Lift
{washing machine} => {kettle, tumble dryer}	0.10	1.00	4.44
{kettle} => {kettle, tumble dryer}	0.10	1.00	4.44
{washing machine},{kettle, washing machine},{washing machine} => {kettle, tumble dryer}	0.10	1.00	4.44
{kettle},{dish washer},{kettle},{washing machine},{washing machine} => {kettle, tumble dryer}	0.15	0.75	3.33
{washing machine},{kettle},{washing machine} => {washing machine, tumble dryer}	0.10	0.66	2.96
{kettle},{washing machine} => {microwave, washing machine}	0.10	0.66	2.96

Detecting household activity patterns

Activity sequence mining – results

- With the support equal to 0.1 and with the confidence of 100%, if in a certain hour the washing machine operated, in the next hour the tumble dryer and kettle operated.
- With the support equal to 0.1 and with the confidence of 100%, if in a certain hour the washing machine operated, in the next hour the washing machine and kettle operated, and in the next hour the washing machine also operated, so did the tumble dryer and kettle.

Detecting household activity patterns

Activity sequence mining – results

- Rule No. 4 with the support equal to 0.15, and with the confidence of 75% shows that the occurrence in a sequence of such devices as kettle, dish washer and washing machine influences the occurrence in a sequence of such appliances as tumble dryer and kettle.
- With the support equal to 0.1 and with the confidence of 66%, if in a certain hour the kettle operated, in the next hour the washing machine was turned ON, then in the next hour the washing machine and microwave were in operation.

Detecting household activity patterns

- What can we do with such information?
- How to include them in load forecasting?

- Lets move to the next part.

Load forecasting on the individual household level

Motivations

- Load forecasting on the individual household level is a challenging task due to the extreme system volatility as a result of dynamic processes composed of many individual components.
- The forecasting performance at the individual level shows errors ranging from 20% to 100% (and even higher), and it depends on dwelling lifestyle and regularity of appliance usage.

Load forecasting on the individual household level

Motivations

- Typical home loads are between 1 and 3 kWh and can be influenced by a number of factors, such as:
  - ✓the operational characteristics of devices,
  - ✓the behaviors of the users,
  - ✓economic factors,
  - ✓time of the day,
  - ✓day of the week, holidays,
  - ✓weather conditions,
  - ✓geographic patterns
  - ✓other random effects.

Load forecasting on the individual household level

Motivations

- Aggregation reduces the inherent variability in electricity consumption resulting in increasingly smooth load shapes, and as a result, the relative forecasting errors typically seen at the level of substations and power systems has been quite low in terms of MAPE (1% – 2%).

Load forecasting on the individual household level

Accuracy measures

- Precision is defined as the measure of how close the model is able to forecast the actual load. To measure precision, the mean squared error (MSE) was used:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(W_i - P_i)^2$$

- where W_i is the observed load in hour i and P_i is the forecasted load in hour i.

Load forecasting on the individual household level

Accuracy measures

- Mean Absolute Percentage Error (MAPE) was the second measure that was used:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{W_i - P_i}{W_i}\right|$$

- where W_i is the observed load in hour i and P_i is the forecasted load in hour i.

Load forecasting on the individual household level

Accuracy measures

- Finally, an accuracy measure was used, which identifies how many correct forecasts the model makes, where the term correctness is defined by the user. This can be done by defining correct forecasts as values within a percentage range of the actual load. However for low loads, a percentage range may become insignificant.
- For a load of 0.1 kWh, a 15% range would be 0.085–0.115, and a forecast of 0.2 kWh will be considered wrong, but in practice, such a forecast would be acceptable. To address this false loss of accuracy, we set two scales to measure the accuracy.

Load forecasting on the individual household level

Accuracy measures

- In this study, we set a 15% range of error for accuracy, but if the load was smaller than 1 kWh, then we considered the range of ±0.15 kWh as the range of acceptable forecasts:

$$Accuracy = \sum 1\{W_{hi} > 1 \& |W_{hi} - P_{hi}| < P_{hi} * 0.15\} + \sum 1\{W_{hi} \leq 1 \& |W_{hi} - P_{hi}| < 0.15\}.$$

- where W_i is the observed load in hour i and P_i is the forecasted load in hour i.

Load forecasting on the individual household level

Predictors

- In this research, we focused on forecasting the electricity usage of a particular household for 24 hours ahead. The attributes were constructed based on time series with hourly electricity demand.
- Electricity demand varies over time depending on the time of **day** (daily cycles), **day of the week** (weekly cycles), **day of the month** (monthly cycles), **season** (seasonal cycles) and **occurrence of holidays**.
- Therefore, we enriched the analysis with an additional 76 dummy variables that described the hour (1-24), 31 variables associated with the day of the month, 7 variables associated with the day of the week, 12 variables associated with the month, one variable indicating a holiday and one variable indicating the sunset in a particular hour.

Load forecasting on the individual household level

Predictors

Attribute No.	Description	Formula
1–24	Hour indicator (dummy variable)	$G_i, i = 1, \dots, 24$
25–55	Day of the month indicator (dummy variable)	$D_i, i = 1, \dots, 31$
56–62	Day of the week indicator (dummy variable)	$T_i, i = 1, \dots, 7$
63–74	Month indicator (dummy variable)	$M_i, i = 1, \dots, 12$
75	Holiday indicator (dummy variable)	S
76	Sunset indicator (dummy variable)	N
77–100	Load of previous 24 hours	$Z_{g-i}, i = 1, \dots, 24$
101–104	Minimum load of previous 3, 6, 12, 24 hours	$min\{Z_{g-1}, \dots, Z_{g-i}\}. i = 3,6,12,24$
105–108	Maximum load of previous 3, 6, 12, 24 hours	$max\{Z_{g-1}, \dots, Z_{g-i}\}. i = 3,6,12,24$
109–114	Load in the same hour of the previous week (6 days)	$Z_{g,d-i}, i = 2, \dots, 7$

Load forecasting on the individual household level

Predictors

Attribute No.	Description	Formula
115–118	Load in the same hour of the same day in previous weeks	$Z_{g,d-i}, i = 14,21,28,35$
119–122	Average temperature observed over previous hourly periods	$avg\{T_{g-i}, \dots, T_{g-i[+1]}\}\ i = 1,3,6,12,24$
123–128	Average temperature observed in the same hour over the previous week	$T_{g,d-i}, i = 2, \dots, 7$
129–132	Average weekly temperature observed in previous i-day periods	$avg\{T_{g,d-i}, \dots, T_{g,d-i[+1]}\}.\ i = 7,14,21,28,35$
133–136	Average humidity observed over previous hourly periods	$avg\{W_{g-i}, \dots, W_{g-i[+1]}\}.\ i = 1,3,6,12,24$
137–142	Average humidity observed in the same hour over the previous week	$W_{g,d-i}, i = 2, \dots, 7$
143–146	Average humidity observed in previous i-day periods	$avg\{W_{g,d-i}, \dots, W_{g,d-i[+1]}\}.\ i = 7,14,21,28,35$

Load forecasting on the individual household level

Predictors

- The features presented in on the next slides are the outcome of segmentation and sequence analysis, and they describe, as widely as possible, the existing dependencies in the data.
- In particular, they revealed the following relations:
  - ✓The structures of the devices' profiles over days, weeks and months, which were discovered by analyzing appliances' switch on (activations) events in each hour, and the outcome of the analysis of the contributions from variables no. 147 to 196, which are associated with the characteristics of a single device.
  - ✓The sequential patterns and the time periods between successive activations are reflected in variables no. 197 to 221 as a result of the sequence mining approach.

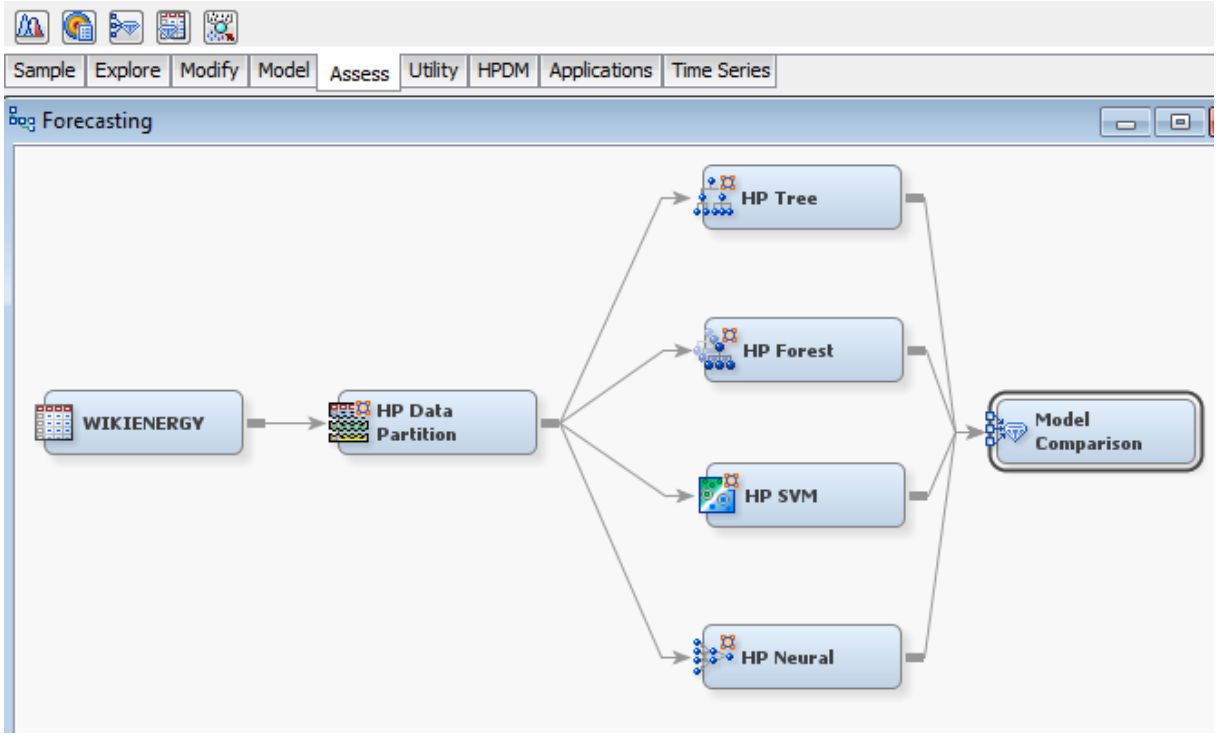
Load forecasting on the individual household level

Predictors

Attribute No.	Description	Formula
147–166	Number of switch on states (activations) for each appliance over previous hourly periods	$\sum_{ON}\left\{\text{Appliance}_{g-i}, \dots, \text{Appliance}_{g-i[+1]}\right\}$, $i = 1,3,6,12,24$
167–176	Number of switch on states (activations) for each appliance over previous daily periods	$\sum_{ON}\left\{\text{Appliance}_{d-i}, \dots, \text{Appliance}_{d-i[+1]}\right\}$, $i = 1,3,7$
177–196	Number of switch on states (activations) for each appliance in previous i-day periods	$\sum_{ON}\left\{\text{Appliance}_{d-i}, \dots, \text{Appliance}_{d-i[+1]}\right\}$, $i = 7,14,21,28,35$
197–221	Number of hours between the most recent five successive activations of each device	$\sum_{G}\left(\text{Appliance}_{ON}, \text{Appliance}_{ON[+1]}\right)$, $ON = 0, \dots, 5$

Load forecasting on the individual household level

Prediction models - settings



General	
Node ID	HPPart
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Partitioning Method	Default
Random Seed	12345
Data Set Allocations	
Training	70.0
Validation	30.0
Status	
Create Time	20.04.19 23:36
Run ID	
Last Error	
Last Status	
Last Run Time	
Run Duration	
Grid Host	
User-Added Node	No

Load forecasting on the individual household level

Prediction models

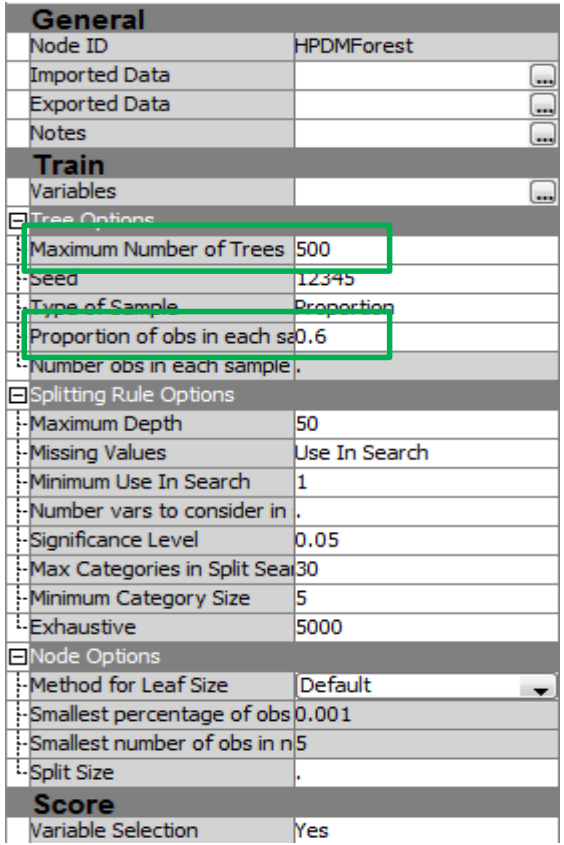
- Desicion trees:

Train	
Variables	
Splitting Rule	
Interval Target Criterion	Variance
Nominal Target Criterion	Entropy
Interval Bins	100
Minimum Distance	0.01
Significance Level	0.2
Bonferroni	No
Missing Values	Largest
Maximum Branch	2
Maximum Depth	10
Minimum Categorical Size	5
Leaf Size	5
Validation	
Create Validation	No
Validation	0.15
Partition Seed	12345
Split Search	
Exhaustive Search Compari	500000
Fast Search Comparisons	1000000
Subtree	
Subtree Method	Assessment
Confidence	0.25
Nominal Target Assessment	Entropy
Assessment Threshold Valu	1.0
Number of Leaves	1
Score	
Variable Selection	Yes
Node and Leaf Role	Segment

Load forecasting on the individual household level

Prediction models

- Random forest:



Load forecasting on the individual household level

Prediction models

- Support vector machine:

General	
Node ID	HPSVM
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Maximum Iterations	25
Use Missing as Level	No
Tolerance	1.0E-6
Penalty	1.0
Optimization Method	
Optimization Method	Interior Point
Interior Point Options	
Active Set Options	

Load forecasting on the individual household level

Prediction models

- Artificial neural network:

General	
Node ID	HPNNA
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Use Inverse Priors	No
Create Validation	No
Network Options	
Input Standardization	Range
Architecture	One Layer
Number of Hidden Neurons	3
Number of Hidden Layers	3
Hidden Layer Options	
Direct Connections	No
Target Standardization	Range
Target Activation Function	Identity
Target Error Function	Normal
Number of Tries	2
Maximum Iterations	300
Use Missing as Level	No

Load forecasting on the individual household level

Prediction models

- Dummy forecast (benchmarking):
  - ✓for the forecasting horizon of 24 hours, the value recorded on a previous day and at the respective hour was taken as the forecast.

Load forecasting on the individual household level

Prediction models

- Forecasting without behavioral variables.
- Forecasting with behavioral variables.

Load forecasting on the individual household level

Results for AMPDS without behavioral variables

Model	MAPE (%)	Acc (%)	MSE	MAPE (%)	Acc (%)	MSE
	Training dataset			Validation dataset		
Naive	42.94	40.90	0.61	40.33	43.68	0.59
Decision trees	37.08	36.36	0.32	38.69	37.00	0.42
Random forest	0.48	100.00	0.00	32.72	37.89	0.39
Neural network	32.39	39.99	0.34	32.19	41.60	0.41
Support vector machines	28.73	43.98	0.35	28.95	44.43	0.46

Load forecasting on the individual household level

Results for AMPDS with behavioral variables

- To identify situations in which additional explanatory (behavioral) variables describing electricity usage patterns improved the final forecast, we introduced a percentage point sensitivity range, denoted with the following colors:
  - ✓green shows forecast improvements in terms of Acc and MAPE when building the model using the enhanced features dataset (with usage pattern variables), e.g., for the model that has 20% error, the improvement should be at least 0.5 p.p. so the model's error should be less than 19.5%.

Load forecasting on the individual household level

Results for AMPDS with behavioral variables

- ✓red shows forecast worsening in terms of Acc and MAPE when building the model using the enhanced features dataset (with usage pattern variables), e.g., for the model that has 20% error, the error increase should be at least 0.5 p.p. so the model's error should be greater than 20.5%.
- ✓no color shows neutral cases in which Acc and MAPE stayed at similar levels, e.g., for the model that has 20% error, we define 1 p.p. range (19.5%–20.5%) to say that no improvement is observed when building the model using enhanced features dataset (with usage pattern variables).

Load forecasting on the individual household level

Results for AMPDS with behavioral variables

Model	MAPE (%)	Acc (%)	MSE	MAPE (%)	Acc (%)	MSE
	Training dataset			Validation dataset		
Naive	42.94	40.90	0.61	40.33	43.68	0.59
Decision trees	39.18	35.16	0.36	39.86	35.36	0.42
Random forest	0.48	100.00	0.00	33.26	37.15	0.39
Neural network	29.65	42.05	0.36	27.61	46.21	0.45
Support vector machines	32.94	34.67	0.34	32.39	35.51	0.46

Load forecasting on the individual household level

Results for WikiEnery – scalability of the approach

Results	Modeling method			
	Decision trees	Random forest	Neural network	Support vector machines
Improving	39.13% (18)	15.22% (7)	82.61% (38)	19.57% (9)
Worsening	26.09% (12)	71.74% (33)	6.52% (3)	69.57% (32)
Neutral	34.78% (16)	13.04% (6)	10.87% (5)	10.87% (5)
Sum	100.00% (46)	100.00% (46)	100.00% (46)	100.00% (46)

Summary and conclusions

- In this paper, we presented an extensive analysis aimed at forecasting electricity loads on the individual household level, which potentially brings greater intelligence to smart meters and delivers value added for individual customers.
- The experiments were designed to find answers to research questions concerning the forecasting loads for individual customers. In particular, the findings are as follows:

Summary and conclusions

Research questions

- Based on the results, we can conclude that it **is possible to provide accurate load forecasting for 24 hours ahead on the individual household level, and this can be obtained with reasonable prediction accuracy.** The forecasts of the neural network models show that they have good performance characterized by low errors obtained on both datasets, i.e., a basic one with past usage data and a richer dataset with usage patterns.

Summary and conclusions

Research questions

- **The clustering and sequence recognition algorithms are good tools for identifying patterns of household behavior.** They allowed quickly grasping general trends in data and then clustering appliances based on their typical usage hours. Sequence analysis gave insight that can help understand how power consumption is influenced by certain activities and their sequences and how those activities are related to each other.

Summary and conclusions

Research questions

- We showed through experiments that a **combination of historical usage data and household behavioral data can greatly enhance the forecasting of individual consumer loads**. The richer data set can reduce MAPE by 8.5% on average and up to 41% for individual households (e.g., household #5 from the WikiEnergy dataset) as observed on the validation set for the neural network model.

Summary and conclusions

Research questions

- The results indicate that there are significant differences in forecasting in favor of the machine learning techniques, namely, SVM, NN, RF, and C4.5, in comparison to random forecasts.
- In particular, **artificial neural networks** through their hidden layers and ability to approximate complex nonlinear mappings directly from the input samples **seem to be very effective at solving short-term forecasting when dealing with high volatility data**.
- They are able to identify hidden trends and make use of richer data, thereby finding the trends in household consumption data.

Summary and conclusions

Future research

- As a future work, we will explore algorithmic approaches for mining typical usage patterns and utilize them for the purpose of energy consumption forecasting and the development of unique, individualized energy management strategies.

Summary and conclusions

Future research

- Additionally, considerable interest and high expectations worldwide are associated with attempts to combine research on forecasting systems utilizing non-intrusive appliance recognition and user patterns with multi agent systems.
- Such a multi-agent computer system can be used for managing the unbalanced energy in a microgrid, and the main goal of the system would be to control and minimize the differences between the current energy demand and the actual energy supply.

Thank you!

krzysztof_gajowniczek@sggw.pl
http://krzysztof_gajowniczek.users.sggw.pl/

Reminder:

Complete your session survey in the conference mobile app.