Smart Meter Data Analytics for Customer Behavior Modeling

Yi Wang Tsinghua University

2018-10-22

Backgrounds

No.	System/ Data	Data Source	Data Type	Frequency	Data Structure
1	Economic Information	Statistic Bureau	GDP、CPI、PMI（Purchasing Managers Index）、Sales Value、Prosperity Index	Per Month	Non structural
2	Energy Consumption Data	Energy Efficiency Platform	Electrical Load、Output、Power Quality、Temperature	15Min	Non structural /Structural
3	Meteorological Data	Meteorological Bureau	Temperature、Humidity、Rainfall	Per Day	Structural
4	EV Charging Data	Charging-Pile RTU	Current、Voltage、Charging Rate、State of Charge	15Min	Structural
5	Customer Service Voice Data	Customer Service System	Customer Voice Data	Real Time	Non structural

Variety

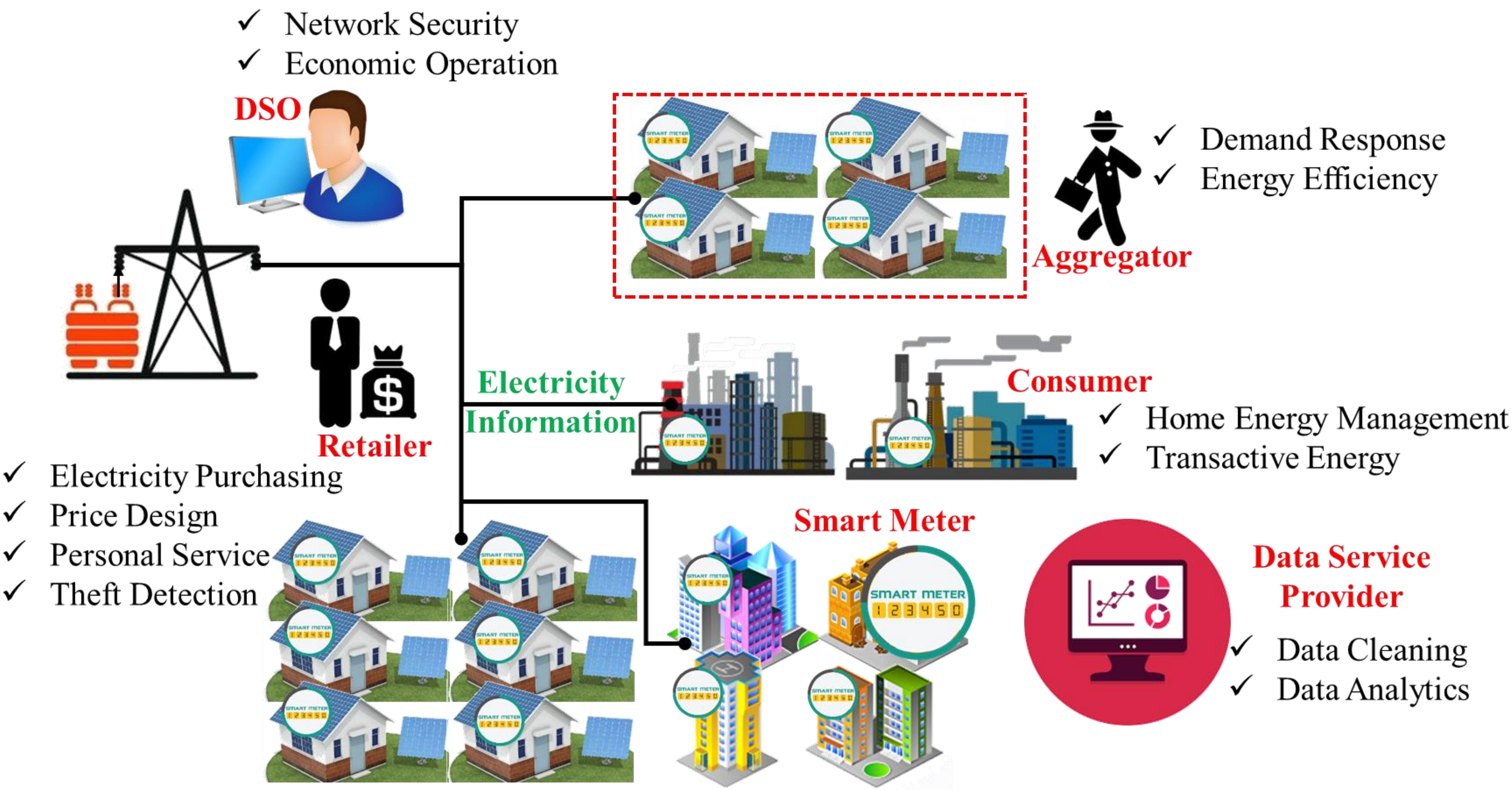
Value???

Velocity

10 million Smart Meters, 15min ⇨ 60GB per day, 21TB per year.

Volume

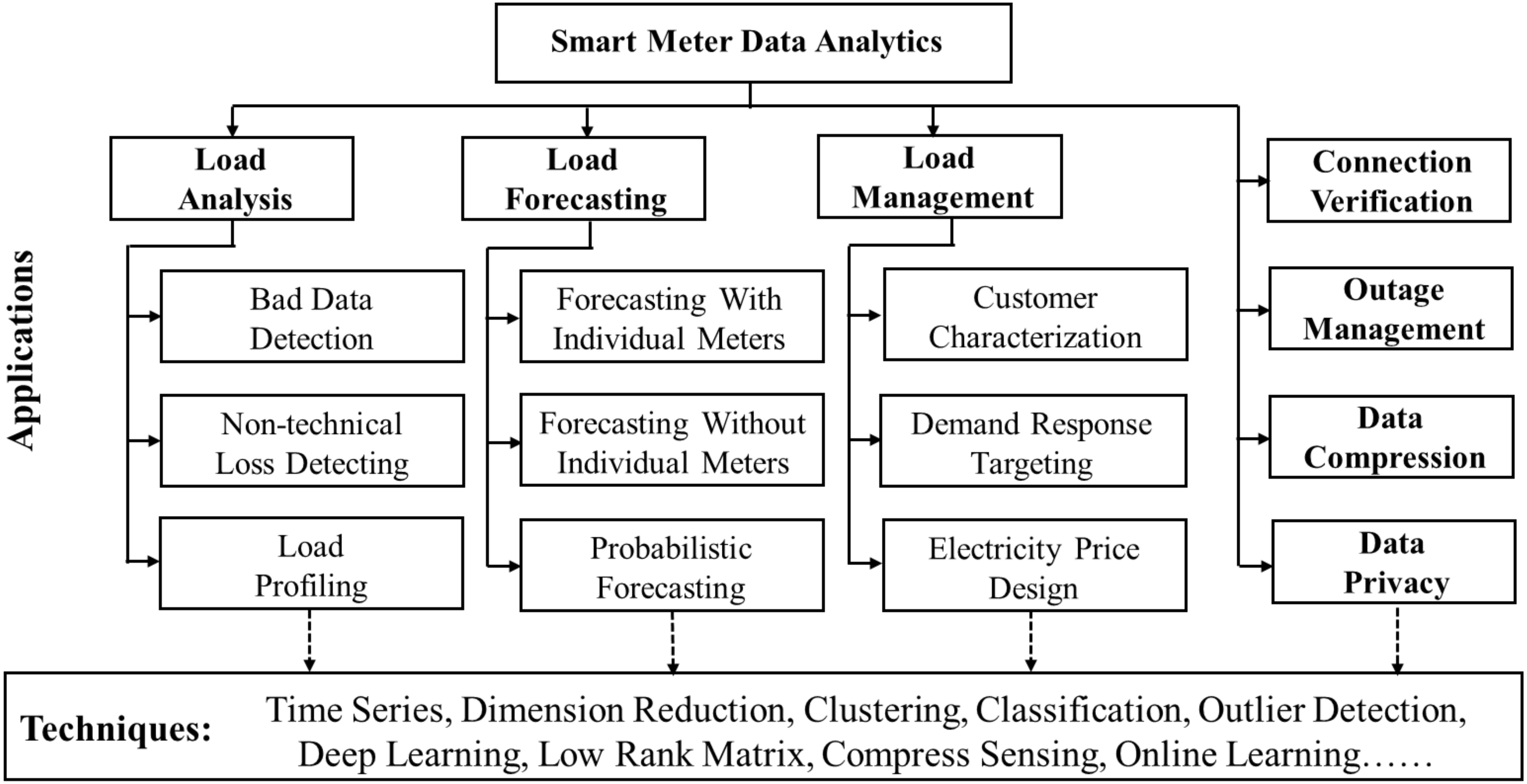
Participators and their businesses on the demand side



Chongqing Kang, **Yi Wang**, Yusheng Xue, Gang Mu, and Ruijin Liao, "Big Data Analytics in China's Electric Power Industry", *IEEE Power and Energy Magazine*, 2018, 16(3):54-65.

Backgrounds

Data Analytics is commonly dissected into three stages: **descriptive analytics** (what do the data look like), **predictive analytics** (what is going to happen with the data), and **prescriptive analytics** (what decisions can be made from the data).

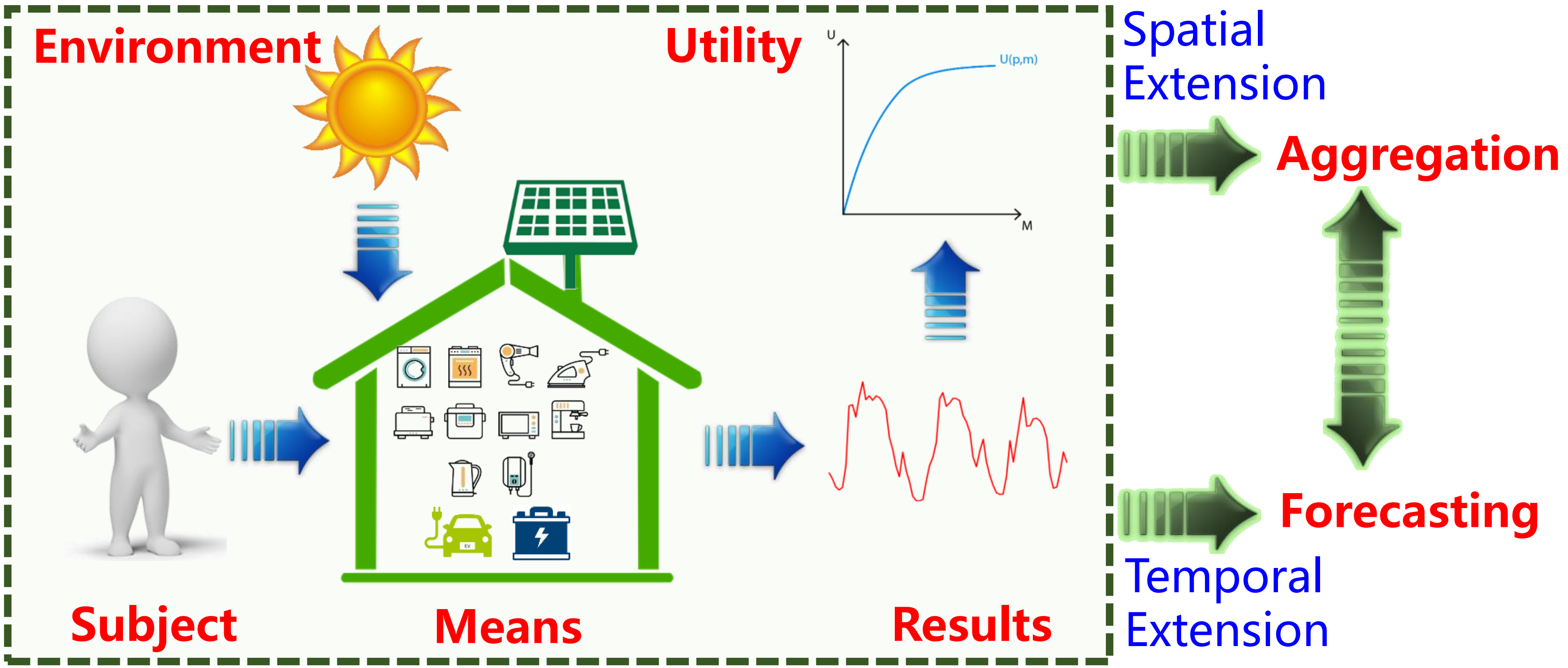


Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges", ***IEEE Trans. Smart Grid***, in press.

Customer Behavior

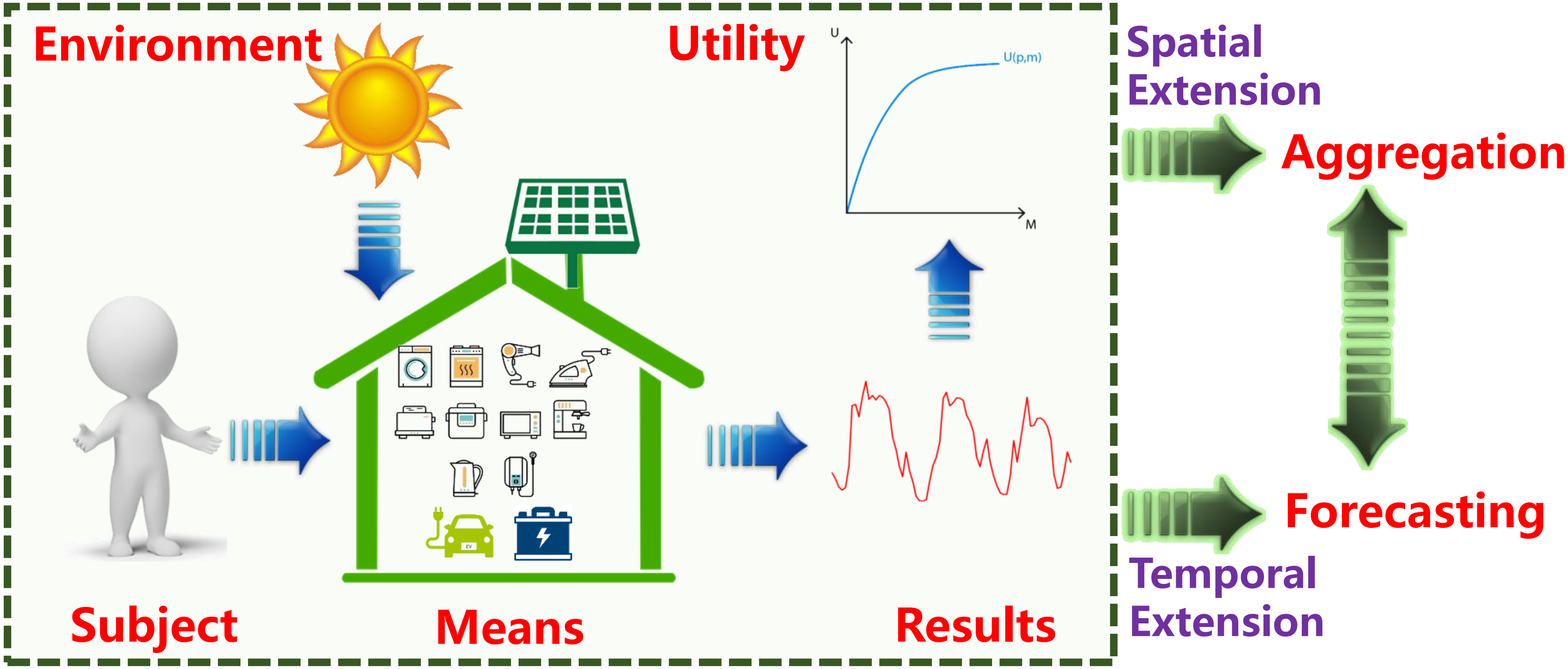
What is Customer Behavior ? One answer from sociological perspective:

- *Customer behavior refers to the electricity consumption* ***activities and related attitudes*** *of customers under a certain environment to maximize the overall utility.*



- *It has five basic parts: behavior subject, behavior environment, behavior means, behavior utility, behavior results.*
- *We can also have two extensions from spatial and temporal perspectives.*

Customer Behavior



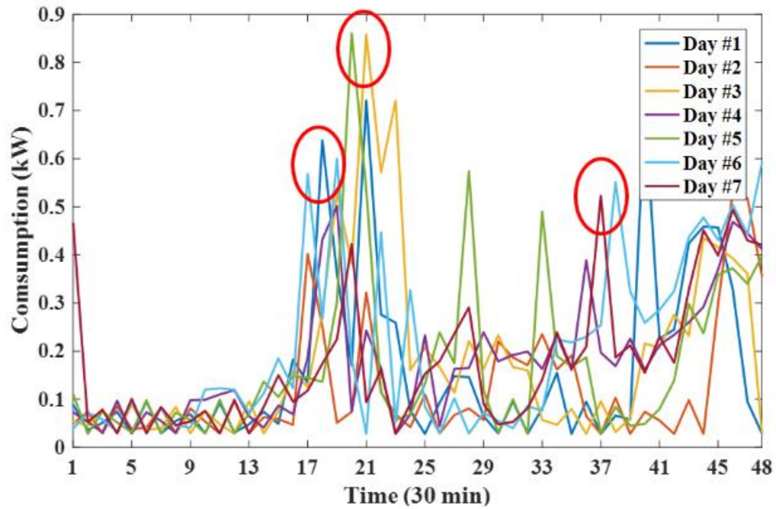
Socio-demographic Information Identification → Load Profile Disaggregation → Consumption Pattern Extraction

→ Customer Aggregation

→ Load Forecasting

Socio-demographic Information

Retailers attempt to analyze customers' electricity consumption behaviors, so that they can provide **diversified and personalized services.**



Can we identify the social-demographic information of the consumers?

Challenges:

1)Problem formulation;

2)High dimensional load data;

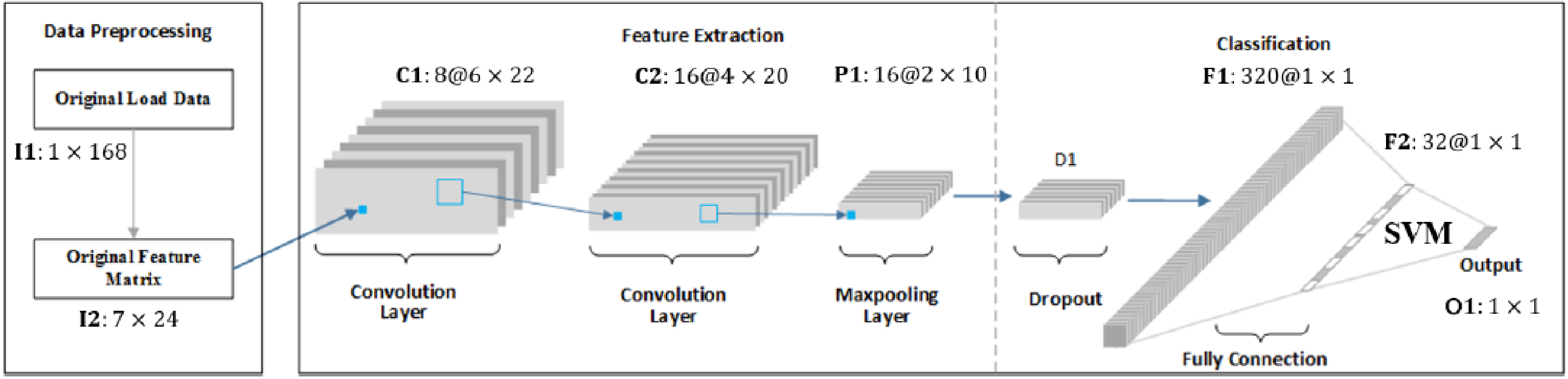
3)High time shift invariance;

<u>Yi Wang</u>, Qixin Chen, Dahua Gan, Jingwei Yang, Daniel Kirschen, and Chongqing Kang, "Deep Learning-Based Socio-demographic Information Identification from Smart Meter Data", ***IEEE Trans. Smart Grid***, in press.

Socio-demographic Information

No.	Question No.	Socio-demographic Information Question	Answers	Number
1	300	Age of chief income earner	Young(<35)	436
			Medium(35∼65)	2819
			Old(>65)	953
2	310	Chief income earner has retired or not	Yes	1285
			No	2947
3	401	Social class of chief income earner	A or B	642
			C1 or C2	1840
			D or E	1593
4	410	Have children or not	Yes	1229
			No	3003
5	450	House type	Detached or bungalow	2189
			Semi-detached or terraced	1964
6	453	Age of the house	Old(>30)	2151
			New(<30)	2077
7	460	Number of bedrooms	Very low(<3)	404
			Low(=3)	1884
			High(=4)	1470
			Very High(>4)	474
8	4704	Cooking facility type	Electrical	1272
			Not Electrical	2960
9	4905	Energy-efficient light bulb proportion	Up to half	2041
			Three quarters or more	2191
10	6103	Floor area	Small(<100)	232
			Medium(>100&<200)	1198
			Big(>200)	351

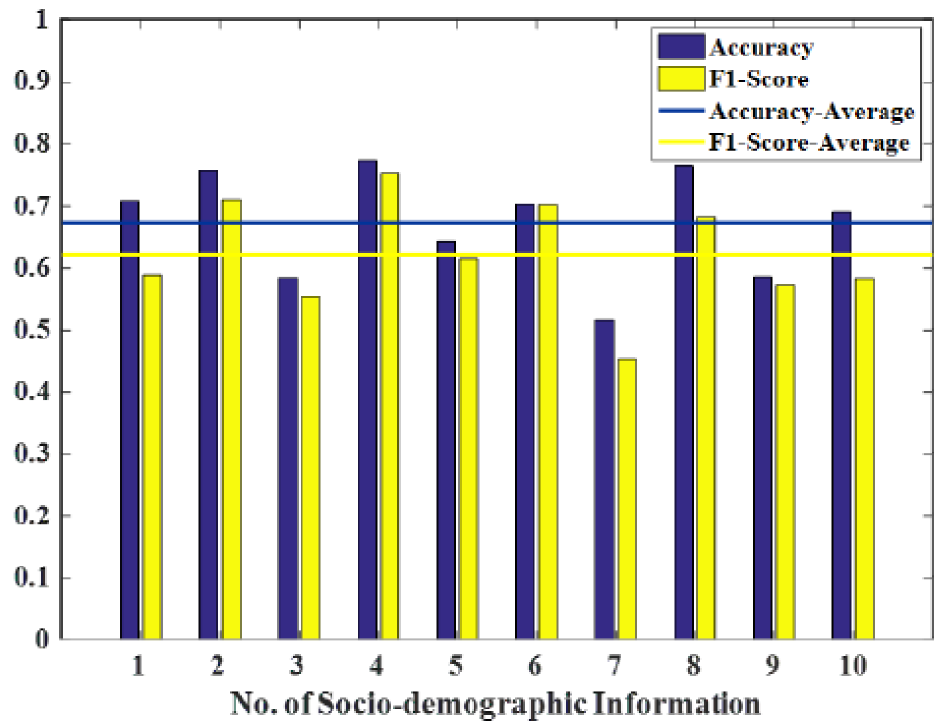
Socio-demographic Information



F_1 SCORES OF DIFFERENT METHODS

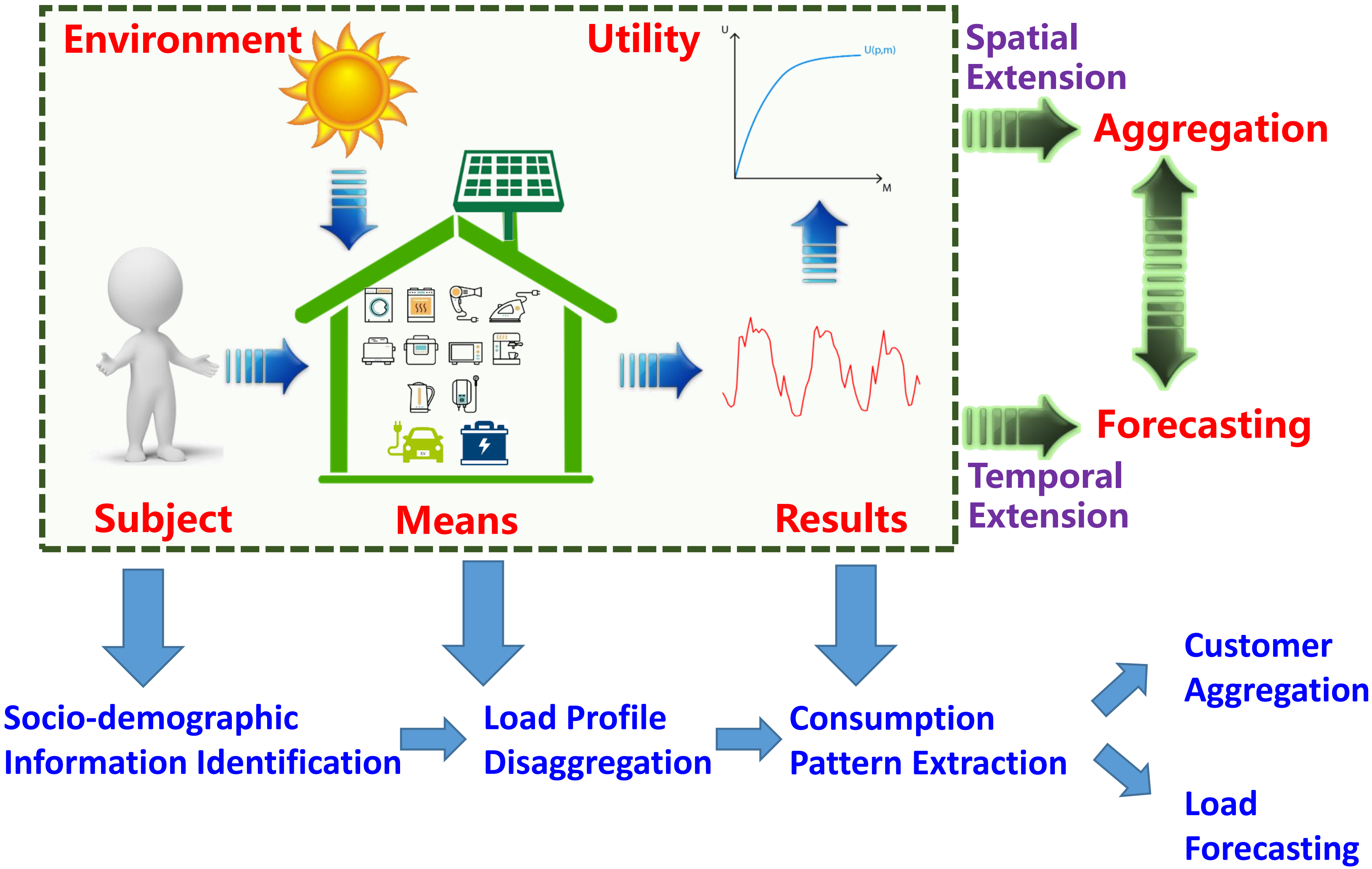
	SVM	LS	PS	SS	CS	Proposed	Improvement 1	Improvement 2
1	0.562	0.563	0.539	0.533	0.571	**0.589**	1.42%	**3.15%**
2	0.652	0.659	0.602	0.569	0.687	**0.71**	4.25%	**3.35%**
3	0.474	0.458	0.47	0.451	0.512	**0.554**	8.02%	**8.20%**
4	0.709	0.711	0.687	0.615	0.737	**0.752**	3.66%	**2.04%**
5	0.446	0.563	0.562	0.451	0.584	**0.616**	3.73%	**5.48%**
6	0.488	0.576	0.52	0.519	0.661	**0.702**	14.76%	**6.20%**
7	0.418	0.389	0.42	0.361	0.432	**0.454**	2.86%	**5.09%**
8	0.584	0.605	0.574	0.574	0.652	**0.683**	7.77%	**4.75%**
9	0.446	0.454	0.491	0.409	0.547	**0.572**	11.41%	**4.57%**
10	0.539	0.538	0.516	0.499	0.552	**0.583**	2.41%	**5.62%**

Socio-demographic Information



- Among these ten questions, the accuracies of #2 (**chief income earner has retired or not**), #4 (**have children or not**), and #8 (**cooking facility type**) are higher than 75%;
- The accuracies of #7 (number of bedrooms) and #9 (energy-efficient light bulb proportion) are lower than 60%;
- The accuracies of the remaining questions are between 60% and 75%.

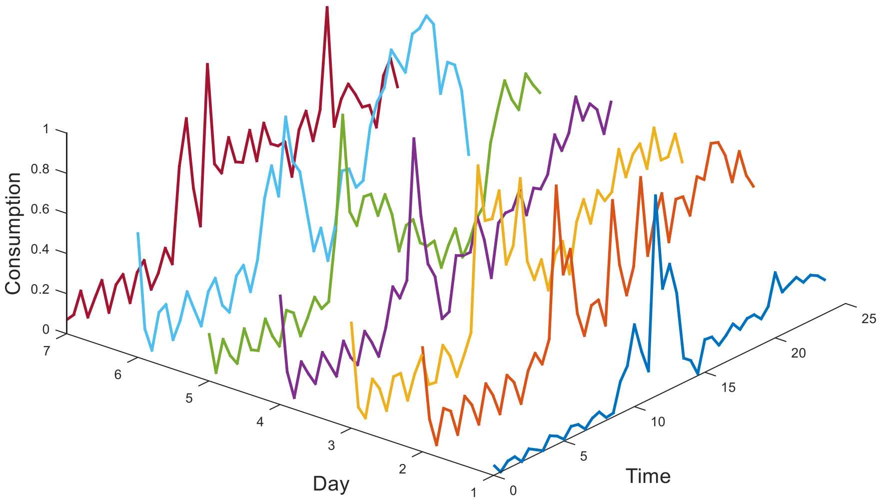
Customer Behavior

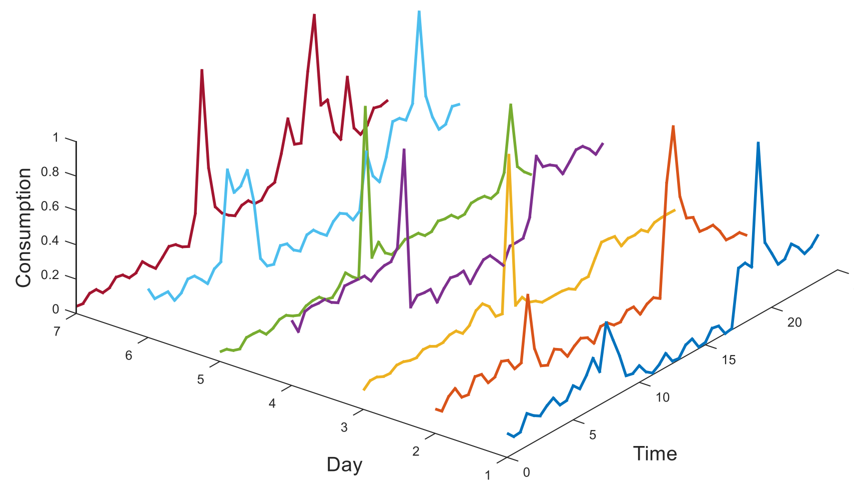


Electricity Consumption Patterns

Energy Intelligence Laboratory
智慧能源实验室

Characteristics of Individual Smart Meter Data



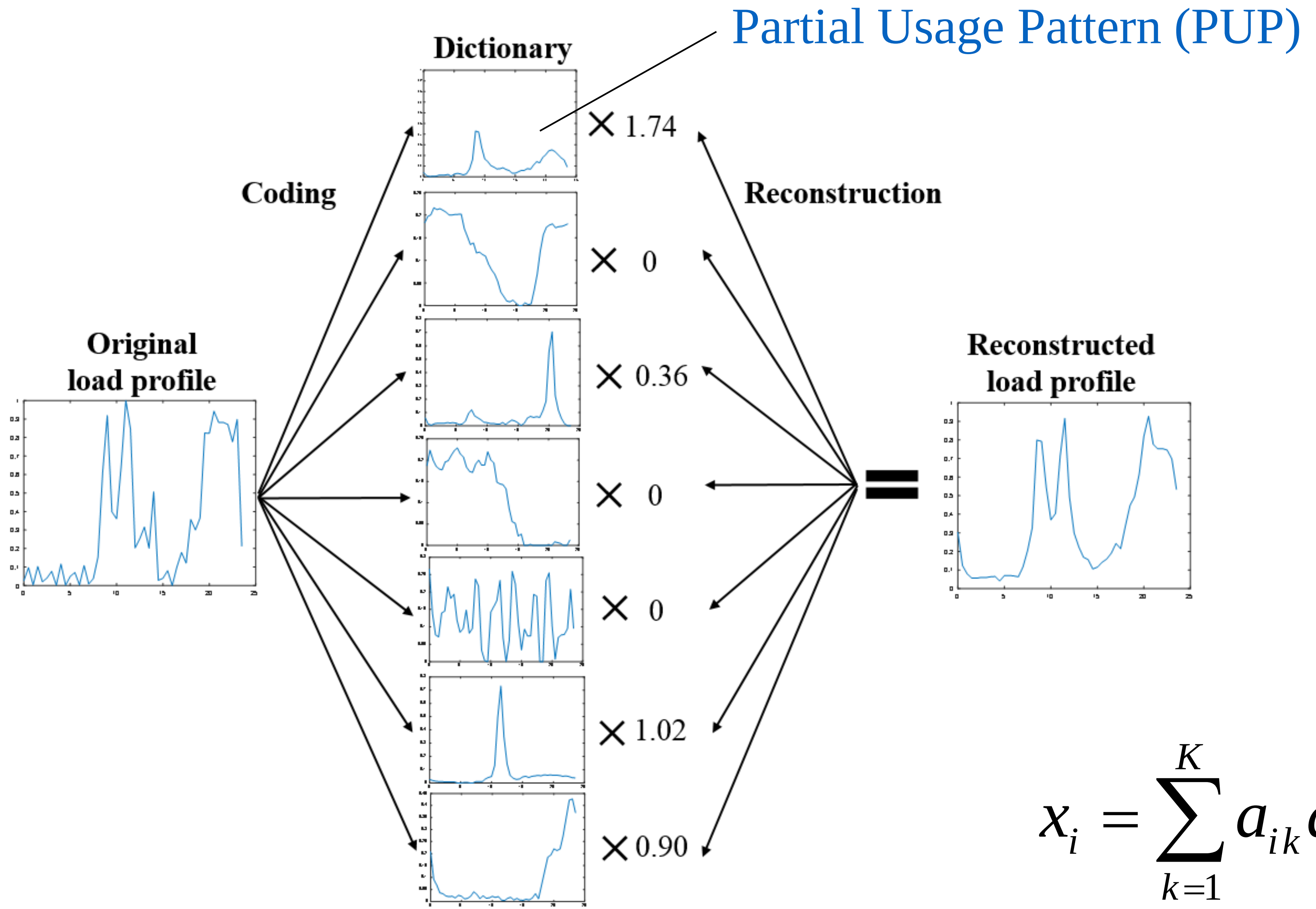


Sparsity: only a small fraction of the time has higher electricity consumption while the rest approximates to zero.

Diversity: load profiles are various with different customers and in different days, but it can be decomposed into different parts.

Yi Wang, Qixin Chen, Chongqing Kang, and Qing Xia, "Sparse and Redundant Representations-Based Smart Meter Data Compression and Pattern Extraction", ***IEEE Trans. Power Systems***, 2017, 32(3): 2142-2151.

Idea of Sparse Coding



$$x_i = \sum_{k=1}^{K} a_{ik} d_k$$

Non-Negative Sparse Coding

$$\min \quad \|\boldsymbol{X} - \boldsymbol{DA}\|_F^2$$

Minimize the recovery error

$$s.t. \quad \|a_i\|_0 \le s_0, \quad 1 \le i \le M$$

Sparsity Constrains

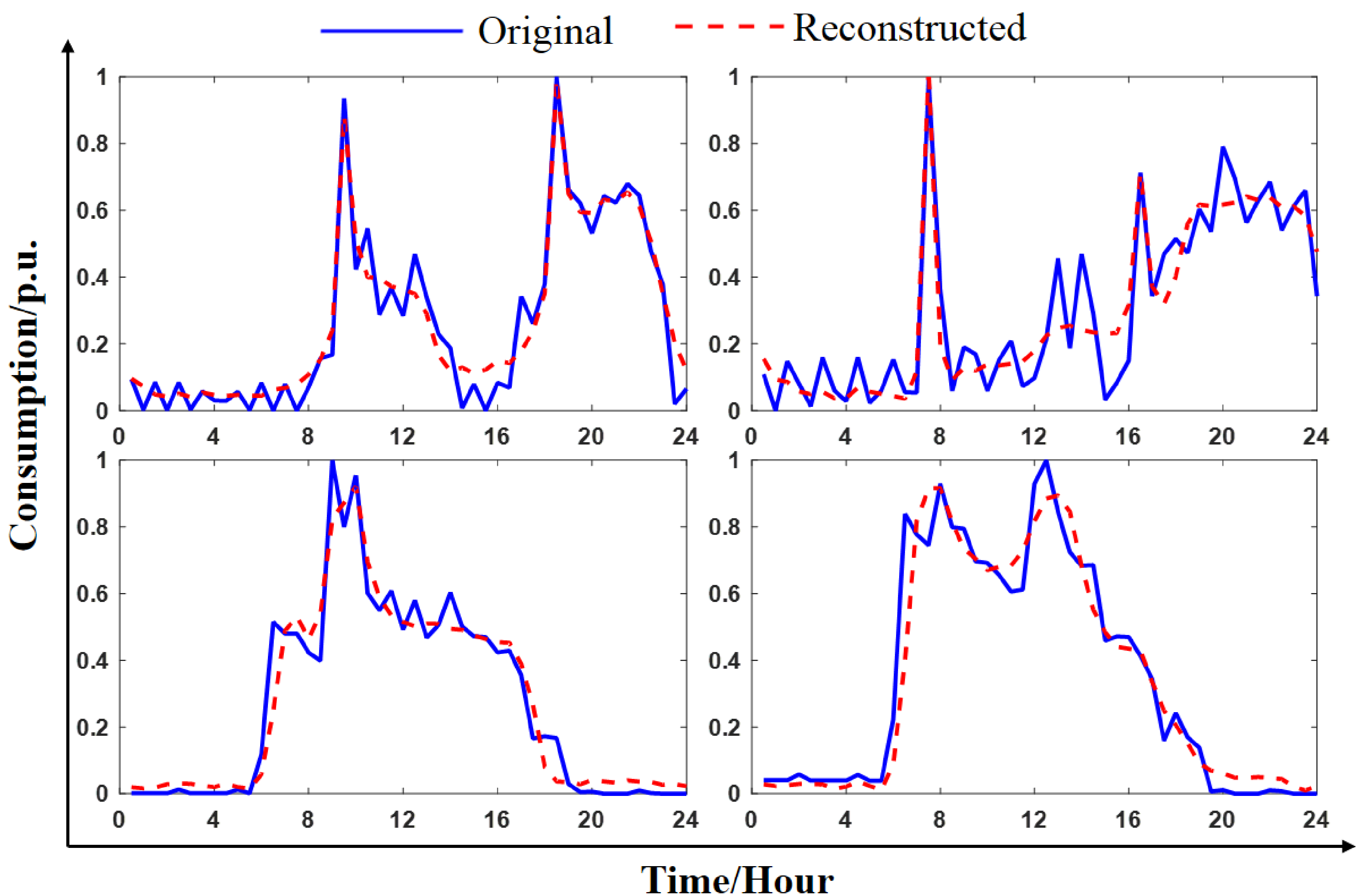
$$a_{i,k} \ge 0, \quad 1 \le i \le M, 1 \le k \le K$$

$$d_{k,n} \ge 0, \quad 1 \le k \le K, 1 \le n \le N$$

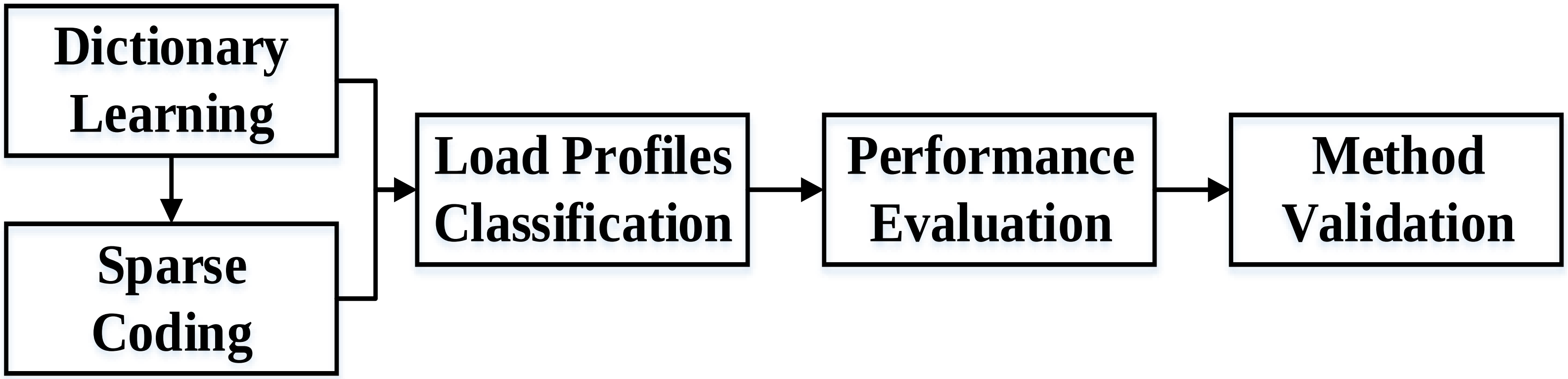
Non-Negative Constrains

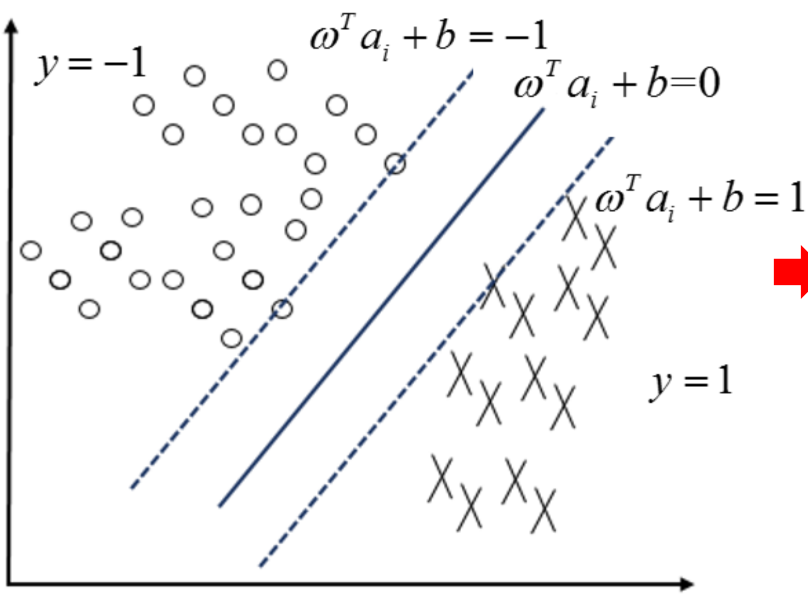
1) Search a redundant dictionary ***D*** that captures the features or PUPs of load profiles as well as possible
2) Optimize the coefficient vector ***A*** of each load profile to guarantee its sparsity and an acceptable reconstruction error.

Electricity Consumption Patterns

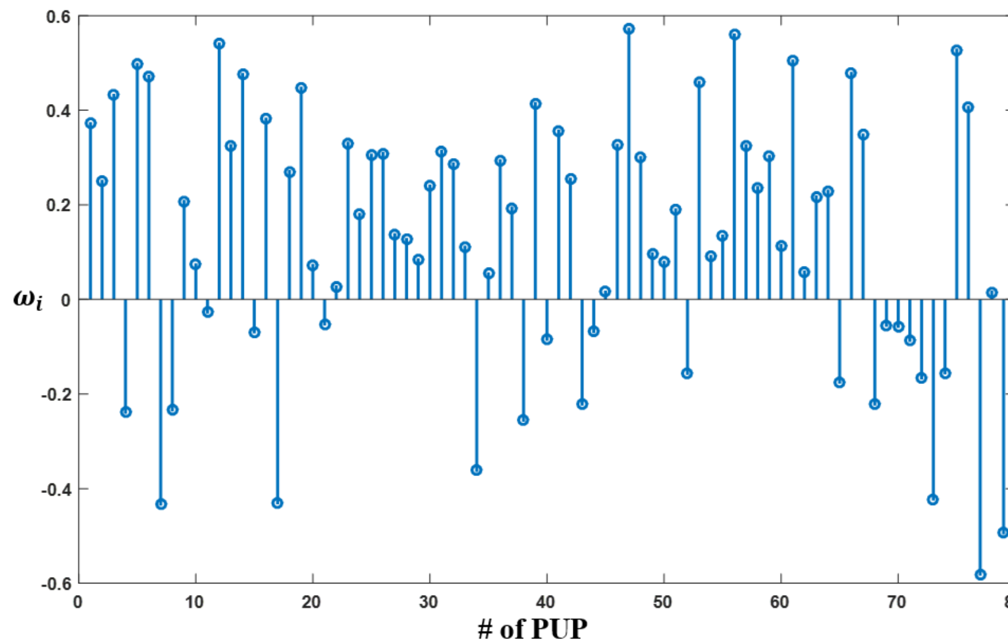


Electricity Consumption Patterns

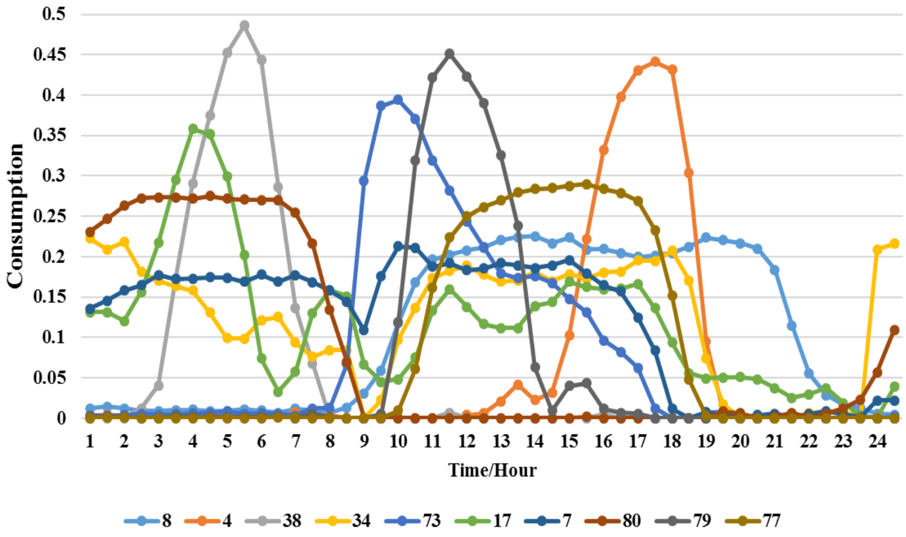


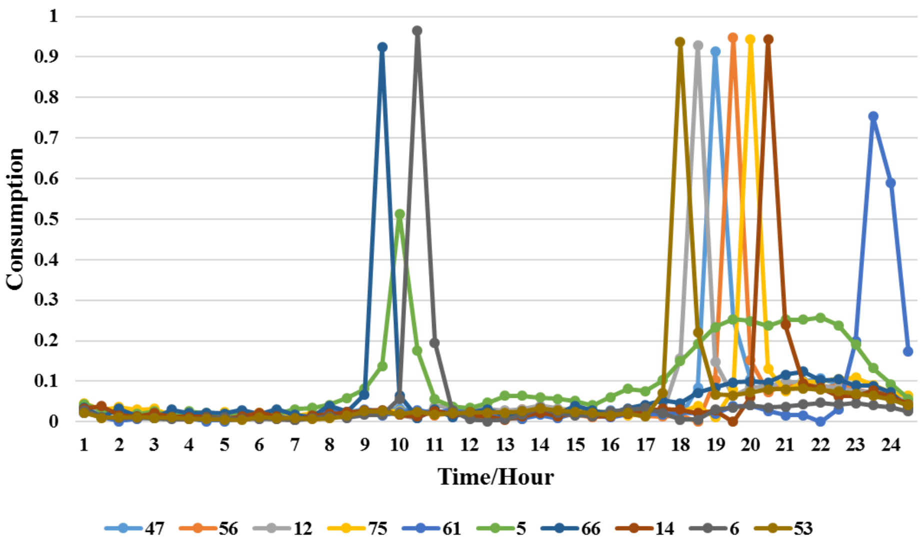


Linear SVM



◆ Ten most relevant PUPs for SMEs and residential customers





	Shape	Duration	Peak times
SME	Vaulted	Long	Dawn, working hours
Resident	Sharp peak	Short	Morning, night

$$RMSE = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(x_i - \sum_{i=1}^{M} a_i d_i)^2}$$

$$MAE = \frac{1}{K}\sum_{i=1}^{K}\left| x_i - \sum_{i=1}^{M} a_i d_i \right|$$

Data Compression-Based

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

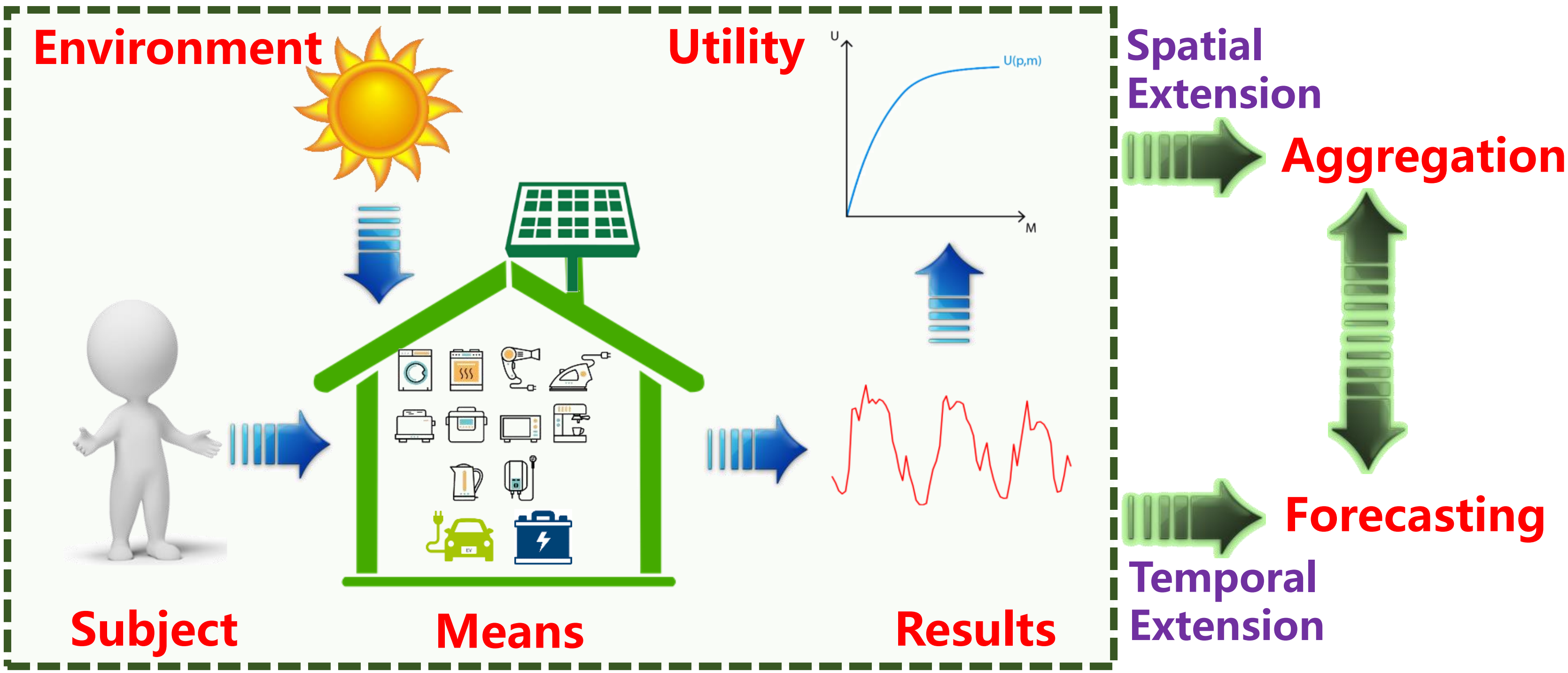
$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

Classification-Based

◆Comparison with Different Techniques

	Parameter	RMSE	MAE	Accuracy	F1
K-SVD	5, 80	0.099	0.060	0.874	0.793
k-means	80	0.120	0.180	0.786	0.752
PCA	5	0.111	0.167	0.771	0.764
DWT	5	0.141	0.327	0.667	0.688
PAA	6	0.112	0.181	0.706	0.725
Original	48	/	/	0.735	0.724

Customer Behavior

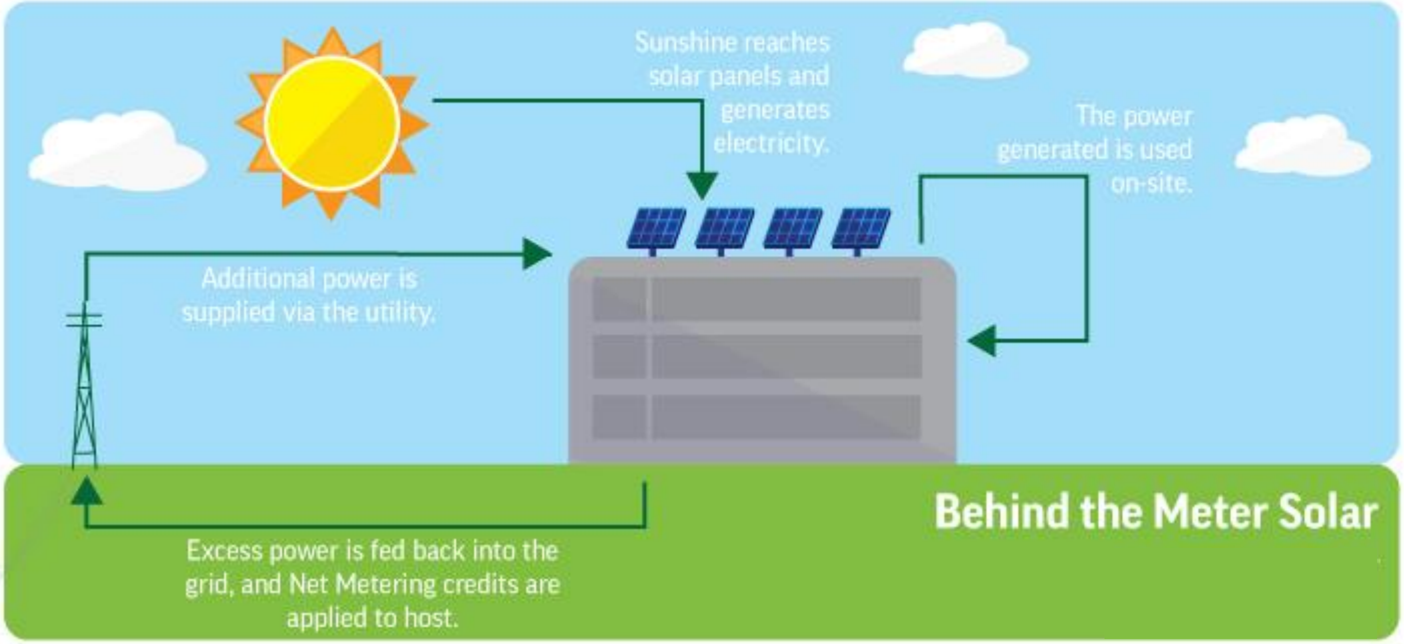


Socio-demographic Information Identification → Load Profile Disaggregation → Consumption Pattern Extraction → Customer Aggregation / Load Forecasting

Problem Statement & Basic Idea

Behind-the-meter (BtM) PV are invisible to DSO which poses great challenges to real time situation awareness.

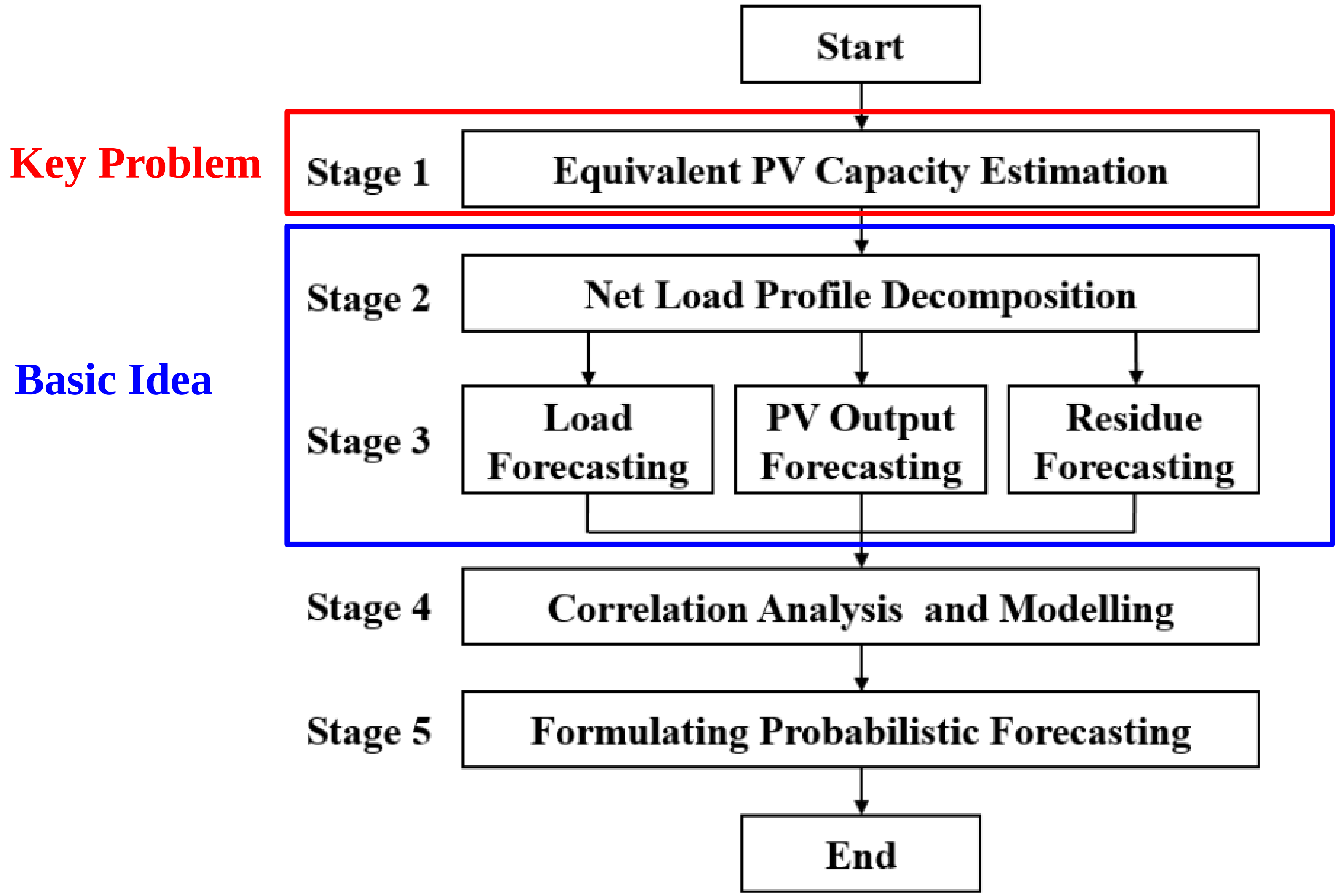
- How to estimate the capacity of BtM PV?
- How to further improve the forecasting accuracy?



Yi Wang, Qixin Chen, Chongqing Kang, and Qing Xia, "Sparse and Redundant Representations-Based Smart Meter Data Compression and Pattern Extraction", ***IEEE Trans. Power Systems***, 2017, 32(3): 2142-2151.

Energy Intelligence Laboratory 智慧能源实验室

Framework



PV Capacity Estimation

Initialize the capacity

Estimate PV output $$P_t \approx C \frac{I_{PV,t}}{1000}\left[1-\mu\left(T_{PV,t}-25\right)\right]$$

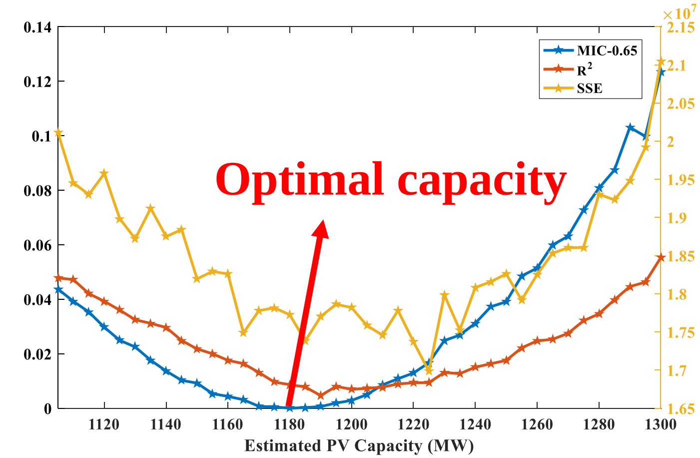
Estimate original load $$L_{E,t}=f_L\left(M_t, W_t, H_t, T_{A,t}\right)$$

Calculate residual $$Res_t = Net_t - L_{E,t} + P_{E,t}$$

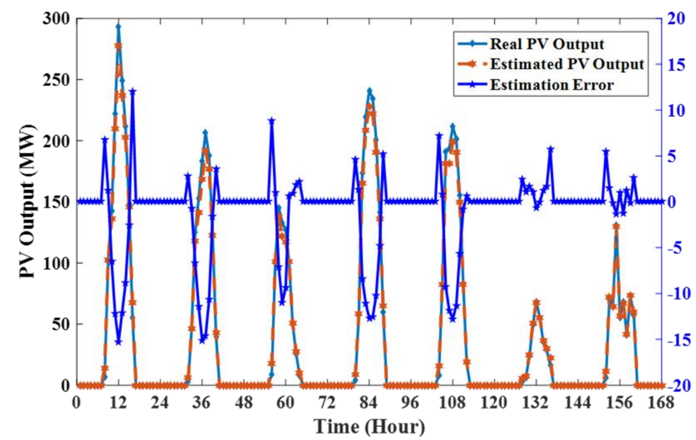
Correlation analysis $$R = \arg\min_{C_{eq},\beta_{eq},\gamma_{eq}} MIC\left(Res_t, I_{GHI,t}\right)$$

Adjust the capacity

Stop when R does not decrease



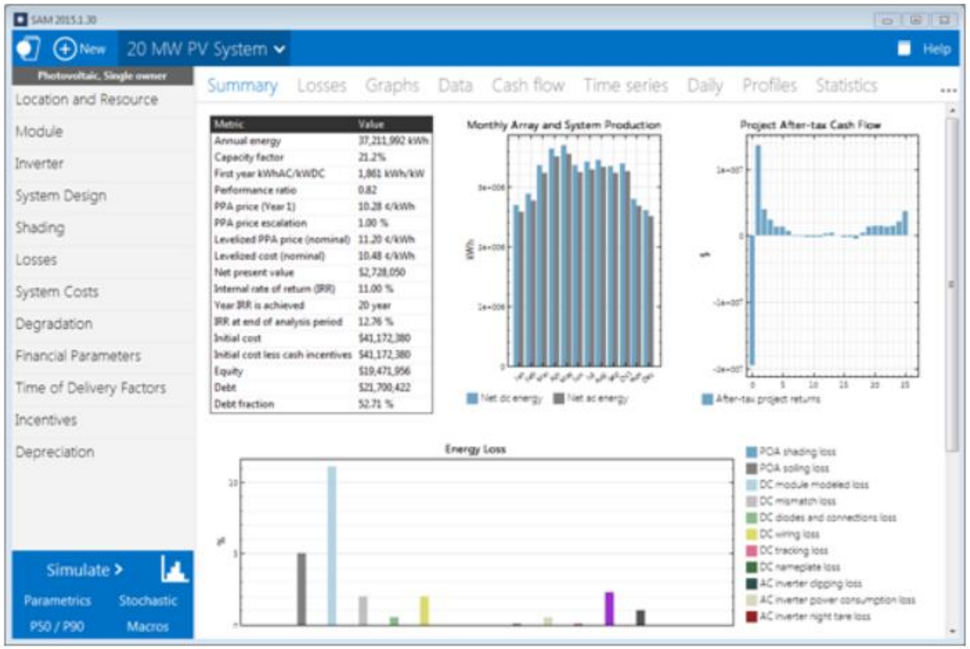
Estimated PV output



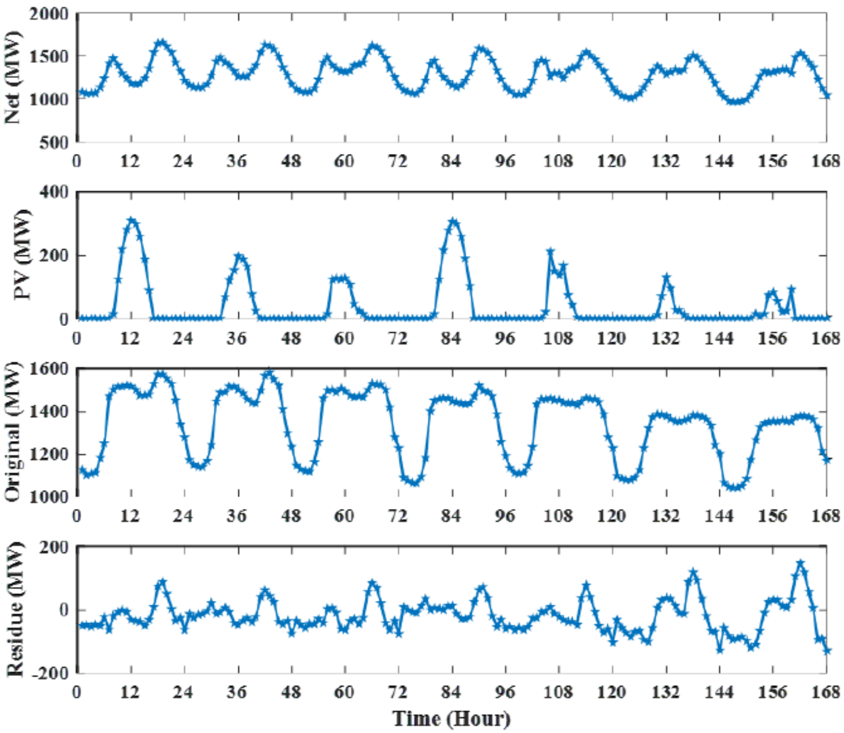
Yi Wang, Ning Zhang, Chongqing Kang, Miao Miao, Rui Shi, and Qing Xia, “An Efficient Approach to Power System Uncertainty Analysis with High-Dimensional Dependencies”, ***IEEE Trans. Power Systems***, 2018, 33(2): 2984-2994.

Data Simulation

System Advisor Model (SAM)
Developed by NREL



Net load separation



Results

Competing methods

	Point Forecasting	Probabilistic Forecasting
Time Series	#1	#4
Considering Temperature	#2	#5
Considering Temperature and Solar Irradiation	#3	#6

Point forecasting

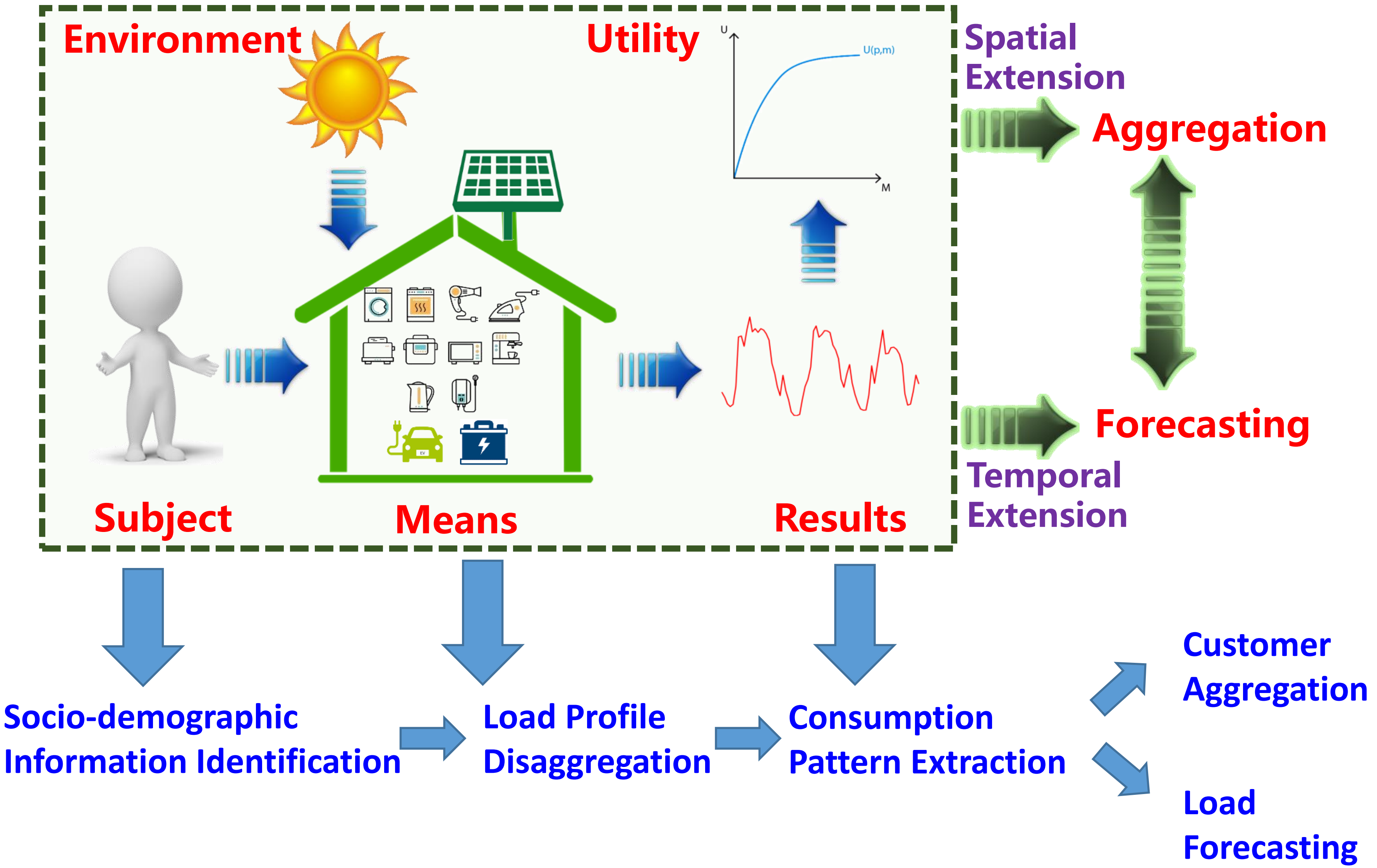
The higher, the better !

PV Penetration	Proposed Method	Method #1	Method #2	Method #3
0	34.3/2.60	38.3/2.85	40.7/3.06	**34.2/2.59**
5%	**60.1/3.37**	94.6/5.28	101.5/5.47	61.4/3.59
10%	**80.9/4.80**	145.8/8.17	157.5/8.50	83.6/5.23
15%	**109.1/7.28**	221.8/13.1	209.7/12.3	115.0/8.25
20%	**140.8/22.6**	279.1/109.2	267.1/84.1	162.8/43.6

Probabilistic forecasting

PV Penetration	Proposed Method	Method #4	Method #5	Method #6
0	34.2	42.1	38.8	**34.0**
5%	**43.4**	60.1	58.1	45.7
10%	**55.9**	82.7	80.5	63.2
15%	**69.2**	108.7	107.5	80.3
20%	**82.5**	135.2	133.7	97.7

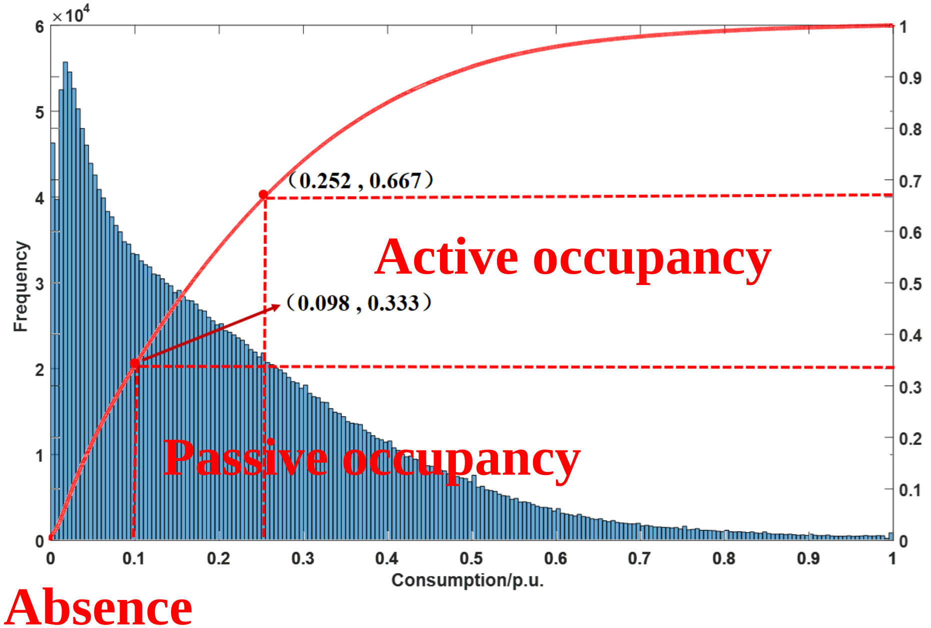
Customer Behavior

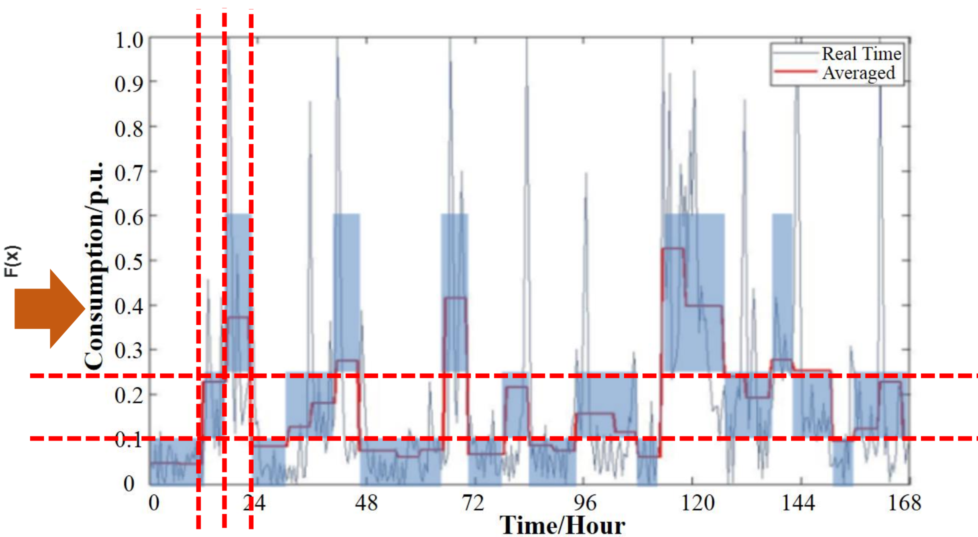


Customer Aggregation

The customers aggregated in the same cluster should share at least one similar characteristic.

- Characteristic Extraction
- Clustering Algorithm



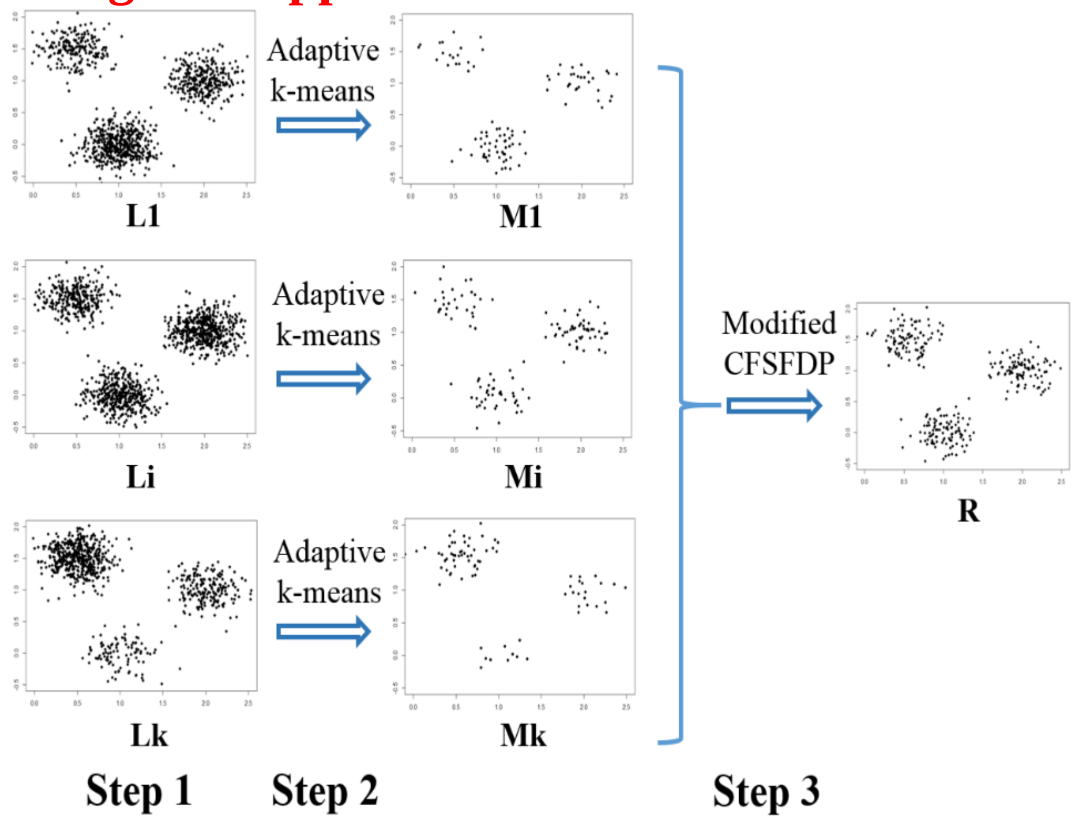


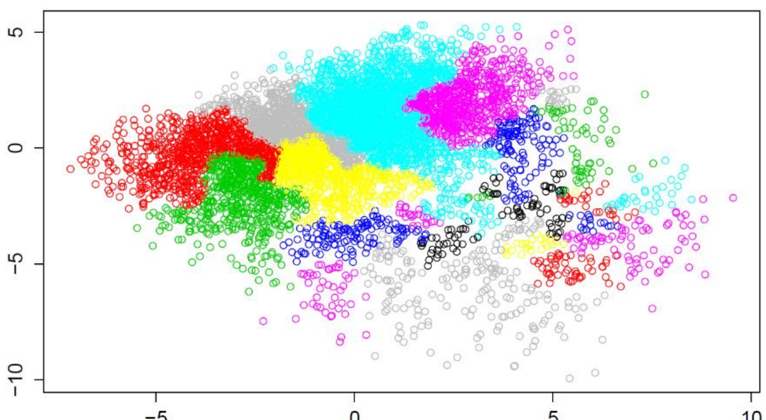
"abcabbcaaacabaabbaccbbbcbabb"

Yi Wang, Qixin Chen, Chongqing Kang, and Qing Xia, “Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications”, ***IEEE Trans. Smart Grid***, 2016, 7(5): 2437-2447.

$$\begin{array}{ccc} \mathbf{P}^1 & & \mathbf{P}^2 \\ \begin{bmatrix} p^1_{11} & p^1_{12} & \cdots & p^1_{1n} \\ p^1_{21} & p^1_{22} & \cdots & p^1_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p^1_{n1} & p^1_{n2} & \cdots & p^1_{nn} \end{bmatrix} & \Rightarrow & \begin{bmatrix} p^2_{11} & p^2_{12} & \cdots & p^2_{1n} \\ p^2_{21} & p^2_{22} & \cdots & p^2_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p^2_{n1} & p^2_{n2} & \cdots & p^2_{nn} \end{bmatrix} \\ \Uparrow & & \Downarrow \\ \begin{bmatrix} p^4_{11} & p^4_{12} & \cdots & p^4_{1n} \\ p^4_{21} & p^4_{22} & \cdots & p^4_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p^4_{n1} & p^4_{n2} & \cdots & p^4_{nn} \end{bmatrix} & \Leftarrow & \begin{bmatrix} p^3_{11} & p^3_{12} & \cdots & p^3_{1n} \\ p^3_{21} & p^3_{22} & \cdots & p^3_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p^3_{n1} & p^3_{n2} & \cdots & p^3_{nn} \end{bmatrix} \\ \mathbf{P}^4 & & \mathbf{P}^3 \end{array}$$

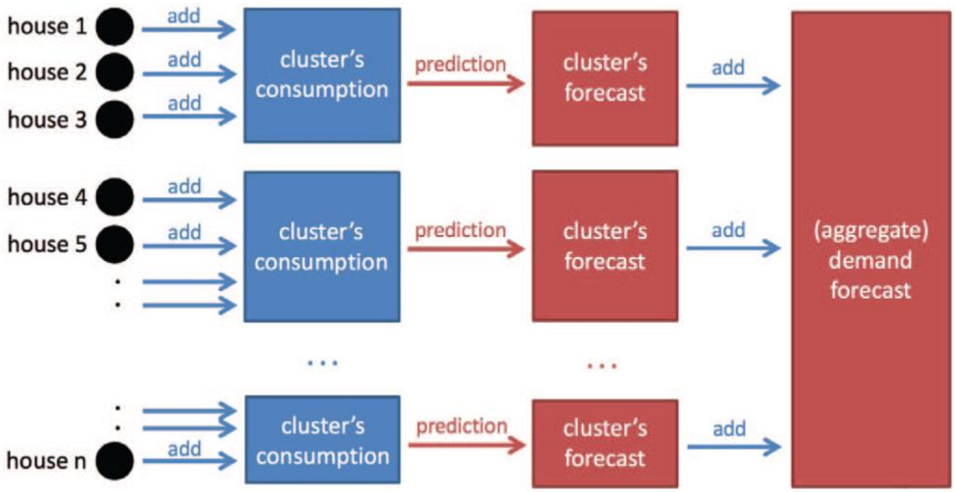
Distributed clustering towards big data applications



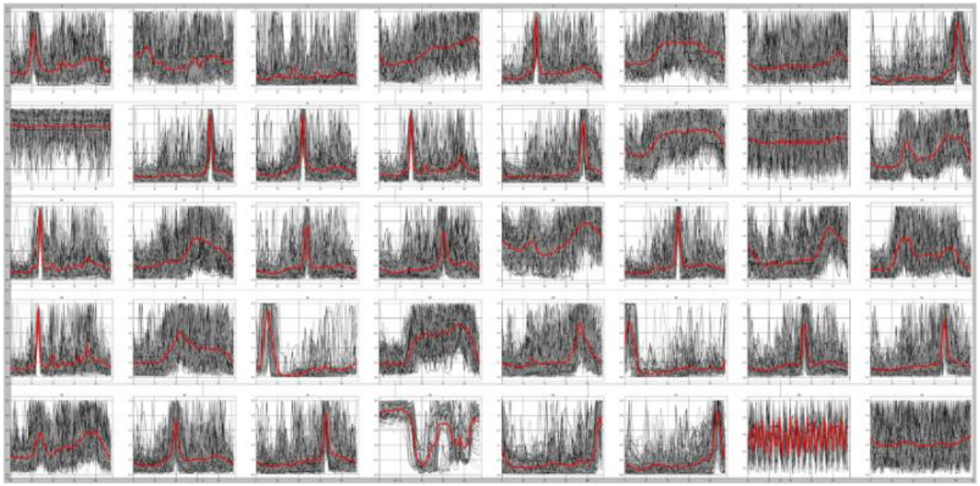


The distribution of over 6000 customers

Which customers are suitable for demand response?



With the prevalence of smart meters, fine-grained sub profiles reveal more information about the aggregated load and further help improve the forecasting accuracy.



Is it possible to utilize both ensemble techniques and fine-grained subprofiles to further improve the forecasting accuracy?

<u>**Yi Wang**</u>, Qixin Chen, Mingyang Sun, and Chongqing Kang and Qing Xia, "An Ensemble Forecasting Method for the Aggregated Load with Subprofiles", ***IEEE Trans. Smart Grid***, 2018, 9(4): 3906-3908.

Aggregated Load Forecasting

- How to determine the number of clusters?
- The optimal number of clusters is different for different dataset?

Real load **The n-th predicted load**

$$\hat{\omega} = \arg\min_{\omega} \sum_{t=1}^{T} \frac{1}{T} \frac{|L_{en,t} - \hat{L}_{en,t}|}{L_{en,t}}$$

Minimize MAPE

$$s.t.\ \hat{L}_{en,t} = \sum_{n=1}^{N} \omega_n \hat{L}_{en,n,t}, \quad \sum_{n=1}^{N} \omega_n = 1, \quad \omega_n \geq 0.$$

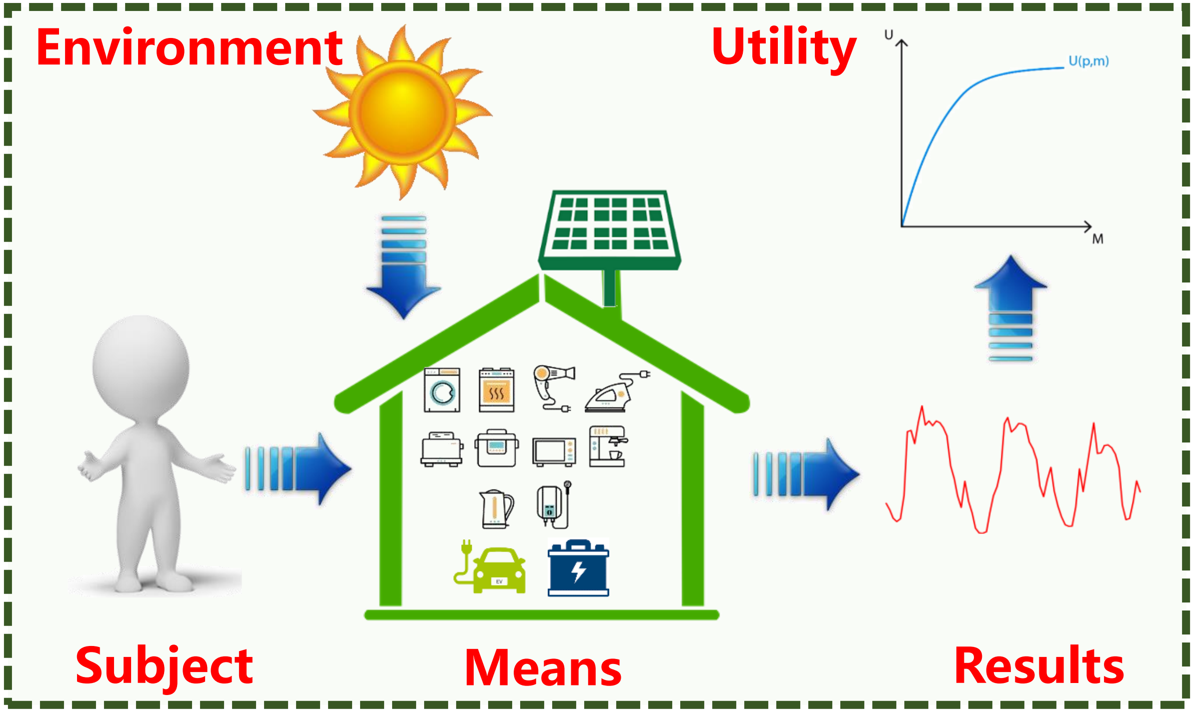
It can be formulated as LP problem.

To determine the weights for the forecasts

N	1	2	4	8	16	32	64	128	256	...	5237	Ensemble
ω	0.634	0	0	0.271	0	0	0.095	0	0	...	0	/
MAPE	***4.25%***	5.05%	5.29%	4.74%	5.55%	4.66%	4.79%	5.09%	5.59%	...	10.31%	**4.05%**
RMSE	***210.95***	229.73	228.01	217.68	244.9	217.64	227.36	232.61	250.27	...	441.33	**202.88**

Case studies on both residential load data and substation load data **demonstrate the superior performance of the proposed ensemble method** when comparing with the traditional direct or bottom-up forecasting strategies.

Customer Behavior



Spatial Extension

Aggregation

Forecasting

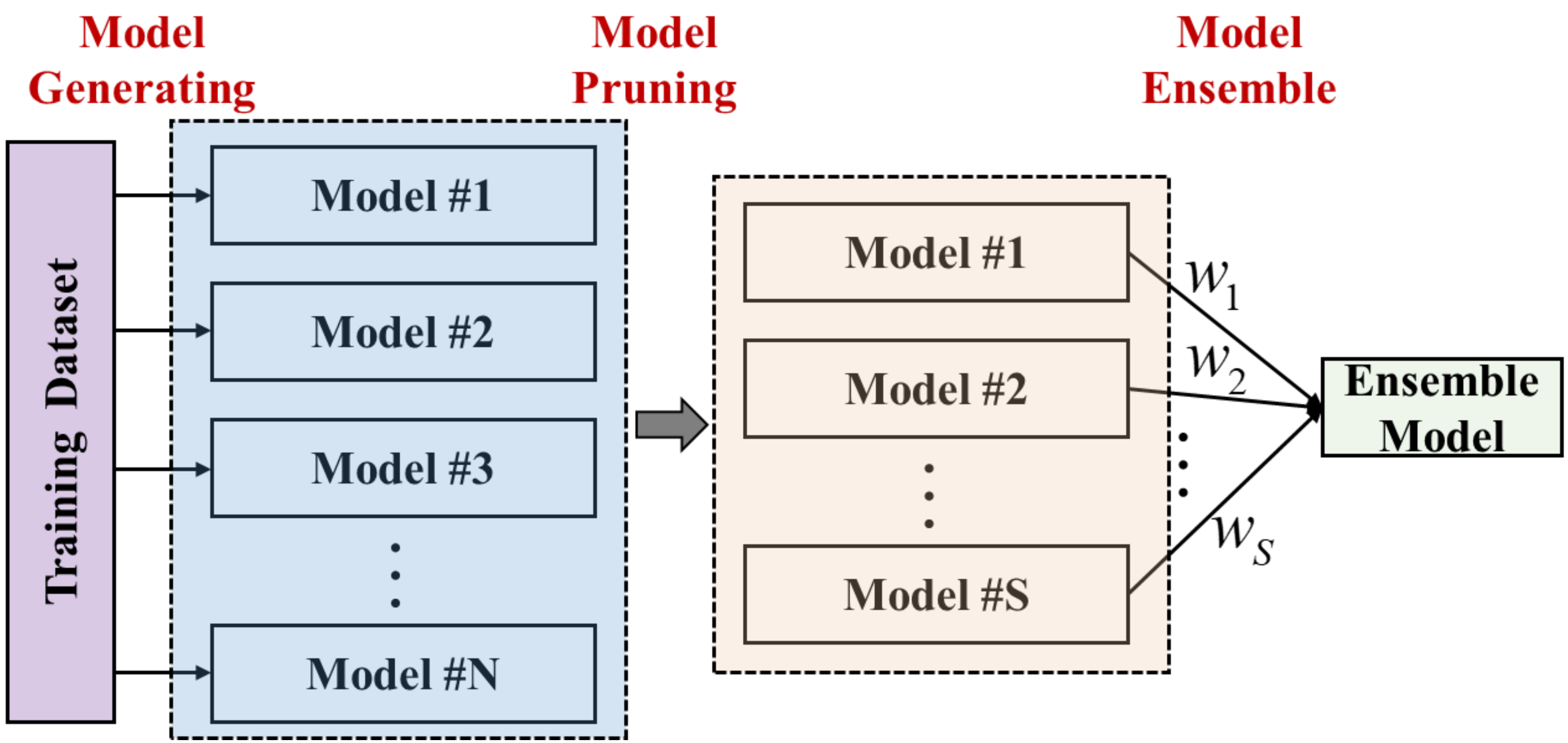
Temporal Extension

Socio-demographic Information Identification → Load Profile Disaggregation → Consumption Pattern Extraction

→ Customer Aggregation

→ Load Forecasting

Ensemble Learning



QRNN
QGBRT
QRF
…

1) Generate models;
2) Prune models;
3) Combine models

Yi Wang, Ning Zhang, Yushi Tan, Tao Hong, Daniel Kirschen, and Chongqing Kang, "Combining Probabilistic Load Forecasts", ***IEEE Trans. Smart Grid***, in press.

From point forecast to probabilistic forecast

$$f_e(\mathbf{X}_{n,t},\boldsymbol{\omega}) = \sum_{n=1}^{N} \omega_n f_n(\mathbf{X}_{n,t},\mathbf{W}_n).$$

$$\begin{aligned} \hat{\omega} = & \arg\min_{\boldsymbol{\omega}} \quad \sum_{t\in T} L_{n,t}\Big(\sum_{n=1}^{N} \omega_n f_n(\mathbf{X}_{n,t},\mathbf{W}_n), y_t\Big) \\ s.t. \quad & \sum_{n=1}^{N} \omega_n = 1, \\ & \omega_n \geq 0, \quad \forall n \in \{1,\cdots,N\}. \end{aligned}$$

Point Forecasts

$$f_{e,q}(\mathbf{X}_{n,t},\boldsymbol{\omega}_q) = \sum_{n=1}^{N} \omega_{n,q} f_{n,q}(\mathbf{X}_{n,t},\mathbf{W}_{n,q}).$$

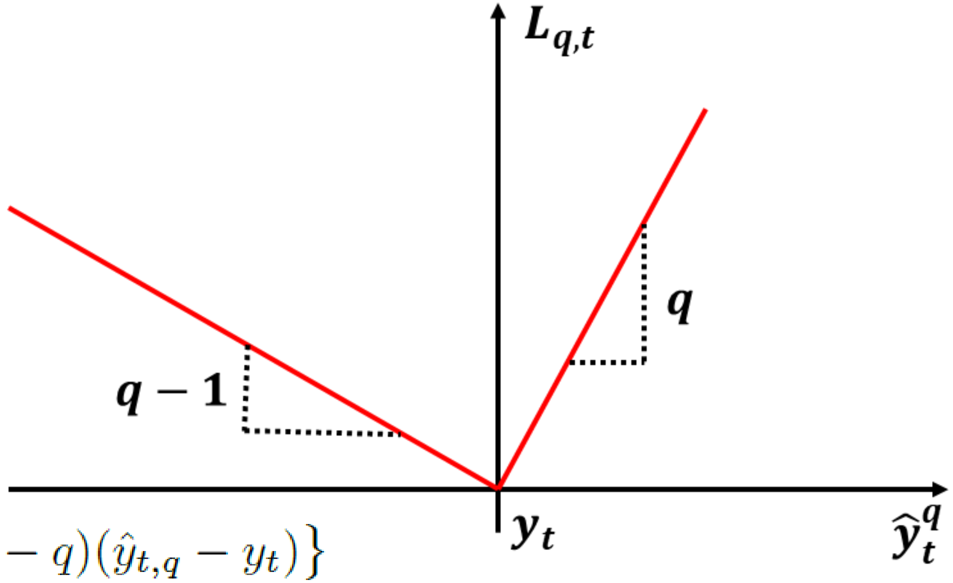
$$\begin{aligned} \hat{\boldsymbol{\omega}}_q = & \arg\min_{\boldsymbol{\omega}_q} \quad \sum_{t\in T} L_{n,t,q}\Big(\sum_{n=1}^{N} \omega_{n,q} f_{n,q}(\mathbf{X}_{n,t},\mathbf{W}_{n,q}), y_t\Big) \\ s.t. \quad & \sum_{n=1}^{N} \omega_{n,q} = 1, \\ & \omega_{n,q} \geq 0, \quad \forall n \in \{1,\cdots,N\}. \end{aligned}$$

Quantile Forecasts

Probabilistic Load Forecasting

Linear Programming Model

$$\hat{y}_{t,q} \approx \sum_{n \in N} \omega_{n,q} \hat{y}_{n,t,q}.$$



$$\hat{\omega}_q = \arg\min_{\omega_q} \sum_{t \in T} L_{t,q}(\hat{y}_{t,q}, y_t)$$

$$= \arg\min_{\omega_q} \sum_{t \in T} \sum_{q \in Q} \max\left\{q(y_t - \hat{y}_{t,q}), (1-q)(\hat{y}_{t,q} - y_t)\right\}$$

$$s.t.\ \ \hat{y}_{t,q} = \sum_{n \in N} \omega_{n,q} \hat{y}_{n,t,q}, \quad \sum_{n \in N} \omega_{n,q} = 1, \quad \omega_n \geq 0 \ \ \forall n.$$

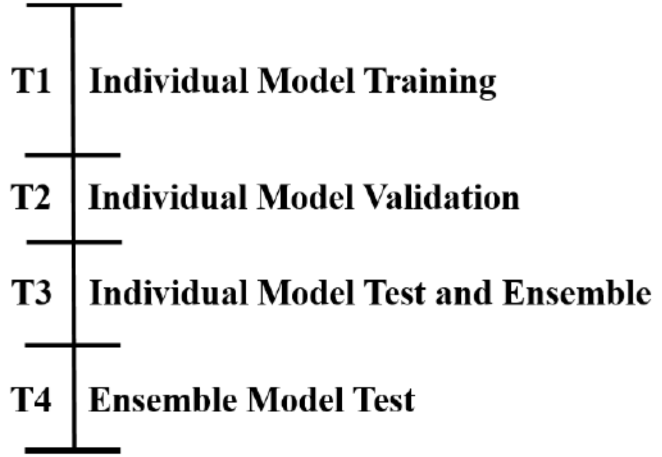
$$\hat{\omega}_q = \arg\min_{\omega_q} \sum_{t \in T} v_{t,q}$$

$$s.t.\ \ \hat{y}_{t,q} = \sum_{n \in N} \omega_{n,q} \hat{y}_{n,t,q}, \quad \sum_{n \in N} \omega_{n,q} = 1, \quad \omega_{n,q} \geq 0 \ \ \forall n.$$

$$v_{t,q} \geq q(y_t - \hat{y}_{t,q}), \quad v_{t,q} \geq (1-q)(\hat{y}_{t,q} - y_t)$$

$$\{v_{t,q} - q(y_t - \hat{y}_{t,q})\}\{v_{t,q} - (1-q)(\hat{y}_{t,q} - y_t)\} = 0.$$

1) LP problem;
2) Model Selection;



Comparisons
Nine models

1) Naïve Sorting (NS): With each forecasting model producing Q quantiles, a total of $N \times Q$ quantiles can be observed (in some sense) by N forecasting models. By sorting these observations by descending order, a new sequence $\mathbf{S}_t = \{S_{t,j}, j = [1, Q \times N]\}$ can be obtained. And therefore the q-th quantile is estimated as follows:

$$\hat{y}_{t,q} = S_{t,1+(q-1)N}.$$

2) Median Value (MED): The median value of the N q-th quantiles is selected as the final quantile:

$$\hat{y}_{t,q} = S_{t,1+(q-1)N+[N/2]}.$$

3) Simple Averaging (SA): The simple averaging strategy applies equal weights to different methods:

$$w_{n,q} = 1/N. \quad ($$

Then, the final combined forecasts are calculated according to Eq. (15).

4) Weighted Averaging (WA): The basic idea of the weighted averaging method is that methods with higher accuracy should be given higher weights:

$$w_{n,q} = \frac{\frac{1}{L_{n,q}}}{\sum_{n \in N} \frac{1}{L_{n,q}}}. \quad ($$

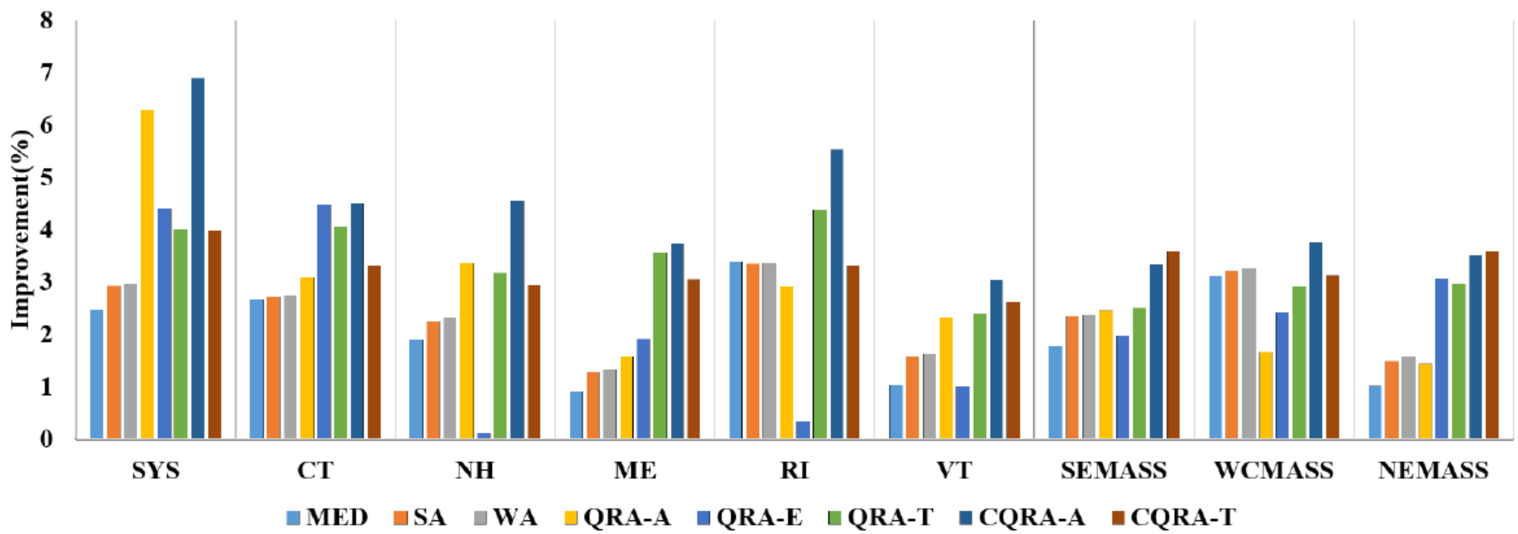
Constraints / Quantiles	With Constraints	Without Constraints
Averaged Quantiles	5) QRA-E	8) CQRA-E
All Quantiles	6) QRA-A	9) CQRA-A
Targeted Quantiles	7) QRA-T	CQRA-T (Proposed)

Probabilistic Load Forecasting

Results

PINBALL LOSSES OF THE INDIVIDUAL AND COMBINATION METHODS FOR DIFFERENT ZONES

Methods \ Zones	SYS	CT	NH	ME	RI	VT	SEMASS	WCMASS	NEMASS
BI	288.563	81.478	27.216	18.146	21.756	12.426	42.307	41.939	63.685
NS	327.569	95.058	31.586	19.003	25.738	13.247	48.817	47.041	71.873
MED	281.607	79.359	26.713	17.981	21.044	12.300	41.570	40.676	63.048
SA	280.375	79.322	26.618	17.916	21.053	12.233	41.336	40.638	62.752
WA	280.266	79.306	26.600	17.908	21.049	12.227	41.329	40.616	62.706
QRA-E	276.417	77.995	27.184	17.806	21.683	12.303	41.484	40.949	61.793
QRA-A	271.519	79.037	26.330	17.864	21.140	12.145	41.295	41.252	62.783
QRA-T	277.487	78.313	26.380	17.523	20.847	12.135	41.271	40.752	61.849
CQRA-E	356.527	100.925	33.829	22.767	26.540	15.616	51.765	51.544	79.131
CQRA-A	277.510	78.870	26.437	17.610	21.059	12.109	**40.847**	40.672	**61.491**
CQRA-T	**269.953**	**77.961**	**26.034**	**17.492**	**20.619**	**12.061**	40.941	**40.422**	61.524



Results

Models \ Quantiles	10-th	20-th	30-th	40-th	50-th	60-th	70-th	80-th	90-th
#1	0	0	0	0.128	0.123	0	0.015	0	0.102
#2	0	0	0	0.177	0.022	0.236	0.154	0.004	0
#3	0.036	0	0	0.041	0.255	0	0.123	0.302	0
#4	0.385	0.444	0.281	0	0	0.030	0	0	0.068
#5	0.165	0	0	0.200	0.298	0.339	0.092	0	0.134
#6	0.037	0.093	0.537	0.264	0	0	0.000	0.251	0
#7	0	0.131	0	0.071	0	0	0.265	0.051	0.218
#8	0	0.207	0.152	0	0.158	0.003	0.350	0.133	0
#9	0.377	0.047	0.030	0.117	0.143	0.392	0	0.206	0.333
#10	0	0.078	0	0	0	0	0	0	0
#11	0	0	0	0	0	0	0	0.052	0.145
#12	0	0	0	0	0	0	0	0	0
#13	0	0	0	0	0	0	0	0	0

Models \ Zones	SYS	CT	NH	ME	RI	VT	SEMASS	WCMASS	NEMASS
#1	0.102	0.144	0.231	0.015	0.001	0.355	0	0	0.196
#2	0	0	0	0.082	0.074	0.146	0.071	0	0
#3	0	0	0.031	0	0	0.079	0	0.196	0
#4	0.068	0	0.089	0.349	0	0	0.038	0	0
#5	0.134	0	0	0	0.272	0	0.199	0.318	0.199
#6	0	0	0.283	0.231	0.226	0.096	0	0	0.136
#7	0.218	0	0.058	0.058	0	0.082	0.166	0.218	0.049
#8	0	0.129	0.308	0.079	0.197	0	0.173	0.076	0.087
#9	0.333	0.341	0	0.185	0.021	0.243	0.290	0.192	0.333
#10	0	0	0	0	0	0	0	0	0
#11	0.145	0.267	0	0	0	0	0	0	0
#12	0	0	0	0	0.210	0	0	0	0
#13	0	0.119	0	0	0	0	0.062	0	0

Smart Meter Data Analytics

Data Management

1. Smart Meter Data Compression Based on Load Feature Identification
2. A Combined Data-Driven Approach for Electricity Theft Detection

Pattern Extraction

3. Typical Electricity Consumption Pattern Identification for Massive Consumers
4. Sparse and Redundant Representation-Based Partial Usage Pattern Extraction

Socio-demographic Information

5. Deep Learning-Based Socio-demographic Information Identification
6. Cross-domain Feature Selection and Coding for Household Energy Behavior

Customer Management

7. Clustering of Electricity Consumption Behavior Dynamics
8. Load Profiling Based Personalized Price Design

Load Forecasting

9. Short-term Probabilistic Residential Load Forecasting with Quantile LSTM
10. An Ensemble Forecasting Method for the Aggregated Load With Subprofiles

Probabilistic Load Forecasting

Fundamentals	1. Introduction to Probabilistic Load Forecasting
Uncertainty Modeling in Three Phases	2. Sparse Penalized Feature Selection 3. Quantile Regression Neural Network Model 4. Modeling Conditional Residual
New Forecasting Objects in Smart Grid	5. Residential Load Forecasting 6. Net Load Forecasting with Behind-the-Meter PV
Forecasts Transformation	7. From Point Forecasts to Probabilistic Forecasts 8. From Quantile Forecasts to Density Forecasts
Forecasting Model Combination	9. Combining Quantile Probabilistic Forecasts 10. Combining Density Probabilistic Forecasts

References

1. <u>**Yi Wang**</u>, Qixin Chen, Tao Hong, and Chongqing Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges", ***IEEE Trans. Smart Grid***, in press.
2. <u>**Yi Wang**</u>, Qixin Chen, Dahua Gan, Jingwei Yang, Daniel Kirschen, and Chongqing Kang, "Deep Learning-Based Socio-demographic Information Identification from Smart Meter Data", ***IEEE Trans. Smart Grid***, in press.
3. <u>**Yi Wang**</u>, Ning Zhang, Qixin Chen, Daniel Kirschen, Pan Li, and Qing Xia, "Data-Driven Probabilistic Net Load Forecasting with High Penetration of Behind-the-Meter PV", ***IEEE Trans. Power Systems***, 2018, 33(2): 3255-3264.
4. <u>**Yi Wang**</u>, Ning Zhang, Chongqing Kang, Miao Miao, Rui Shi, and Qing Xia, "An Efficient Approach to Power System Uncertainty Analysis with High-Dimensional Dependencies", ***IEEE Trans. Power Systems***, 2018, 33(2): 2984-2994.
5. <u>**Yi Wang**</u>, Qixin Chen, Chongqing Kang, and Qing Xia, "Sparse and Redundant Representations-Based Smart Meter Data Compression and Pattern Extraction", ***IEEE Trans. Power Systems***, 2017, 32(3): 2142-2151.
6. <u>**Yi Wang**</u>, Qixin Chen, Chongqing Kang, and Qing Xia, "Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications", ***IEEE Trans. Smart Grid***, 2016, 7(5): 2437-2447.
7. <u>**Yi Wang**</u>, Qixin Chen, Mingyang Sun, and Chongqing Kang and Qing Xia, "An Ensemble Forecasting Method for the Aggregated Load with Subprofiles", ***IEEE Trans. Smart Grid***, 2018, 9(4): 3906-3908.
8. <u>**Yi Wang**</u>, Dahua Gan, Ning Zhang, Zongxiang Lu, and Chongqing Kang, "Probabilistic Individual Load Forecasting Using Pinball Loss Guided LSTM", ***Applied Energy***, in press.
9. <u>**Yi Wang**</u>, Qixin Chen, Ning Zhang, and Yishen Wang, "Conditional Residual Modeling for Probabilistic Load Forecasting", ***IEEE Trans. Power Systems***, in press.
10. <u>**Yi Wang**</u>, Ning Zhang, Yushi Tan, Tao Hong, Daniel Kirschen, and Chongqing Kang, "Combining Probabilistic Load Forecasts", ***IEEE Trans. Smart Grid***, in press.

Q & A

Yi Wang Tsinghua University

2018-10-22