



清华大学

EITab
Energy Intelligence Laboratory
智慧能源实验室

清华大学博士学位论文答辩

数据驱动的电力用户行为建模 与分析方法研究

专 业：电气工程

博士生：王 毅

导 师：康重庆 教授

2018年12月20日



1

论文研究背景

2

论文研究内容

3

论文研究结论

1. 广泛用户互动逐步成为智能电网的重要特征

- ✓ 系统安全可靠
- ✓ 经济调度



DSO



电力零售商

- ✓ 购电决策
- ✓ 电价设计
- ✓ 个性化服务
- ✓ 窃电检测



负荷集成商



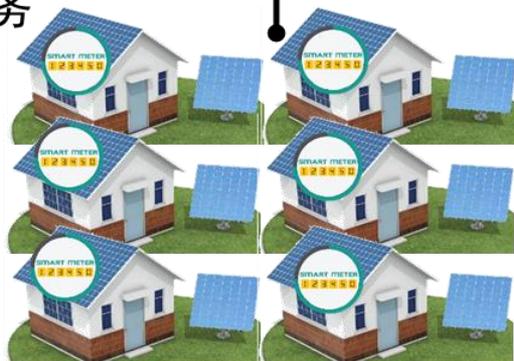
- ✓ 需求响应
- ✓ 能效管理

电力流
信息流



用户 (居民, 商业, 工业)

- ✓ 家庭能量管理系统
- ✓ 可交易能源



智能电表

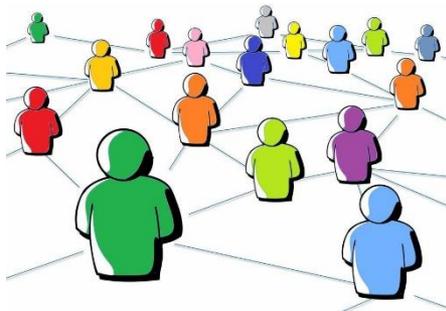


数据服务商



- ✓ 数据清洗
- ✓ 数据分析

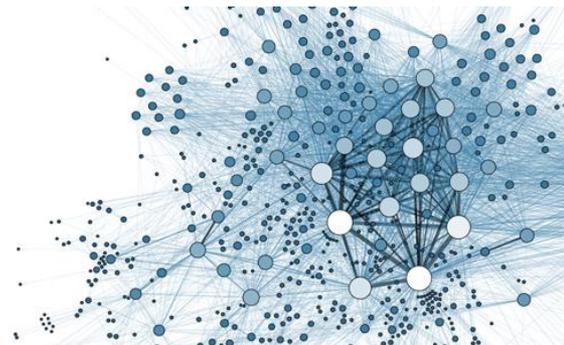
2. 未来电力系统将成为“物理-信息-社会”深度耦合的系统



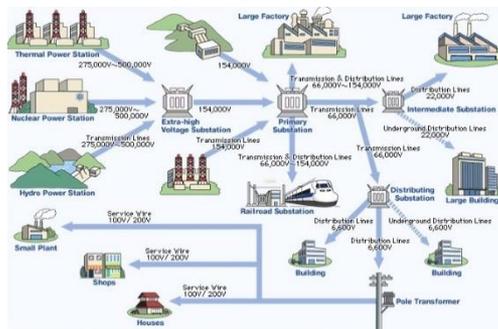
社会网络



信息网络



“物理-信息-社会” 深度耦合的系统



物理网络

自然 (Nature) 出版社专门开辟在线论坛供研究者探讨和分享行为学和社会学研究及其在各行业的应用。

对社会/人的行为建模不够!

3. 智能电表的广泛普及使得需求侧步入用电大数据时代

	系统/ 数据	数据源	数据类型	采集频率	数据结构
1	经济信息	统计局	GDP、CPI、PMI（采购经理指数）、销售额、景气指数	每月	非结构化
2	能源消费 数据	能效平台、 营销平台	电力负荷、电能质量	15分钟	非结构化/ 结构化
3	气象数据	气象部门	温度、湿度、降雨量	每天	结构化
4	电动汽充电 数据	充电线RTU	电流、电压、 充电速率、充电状态	15分钟	结构化
5	客户服务 语音数据	客服系统	客户语音数据	实时	非结构化

Variety（多样性）

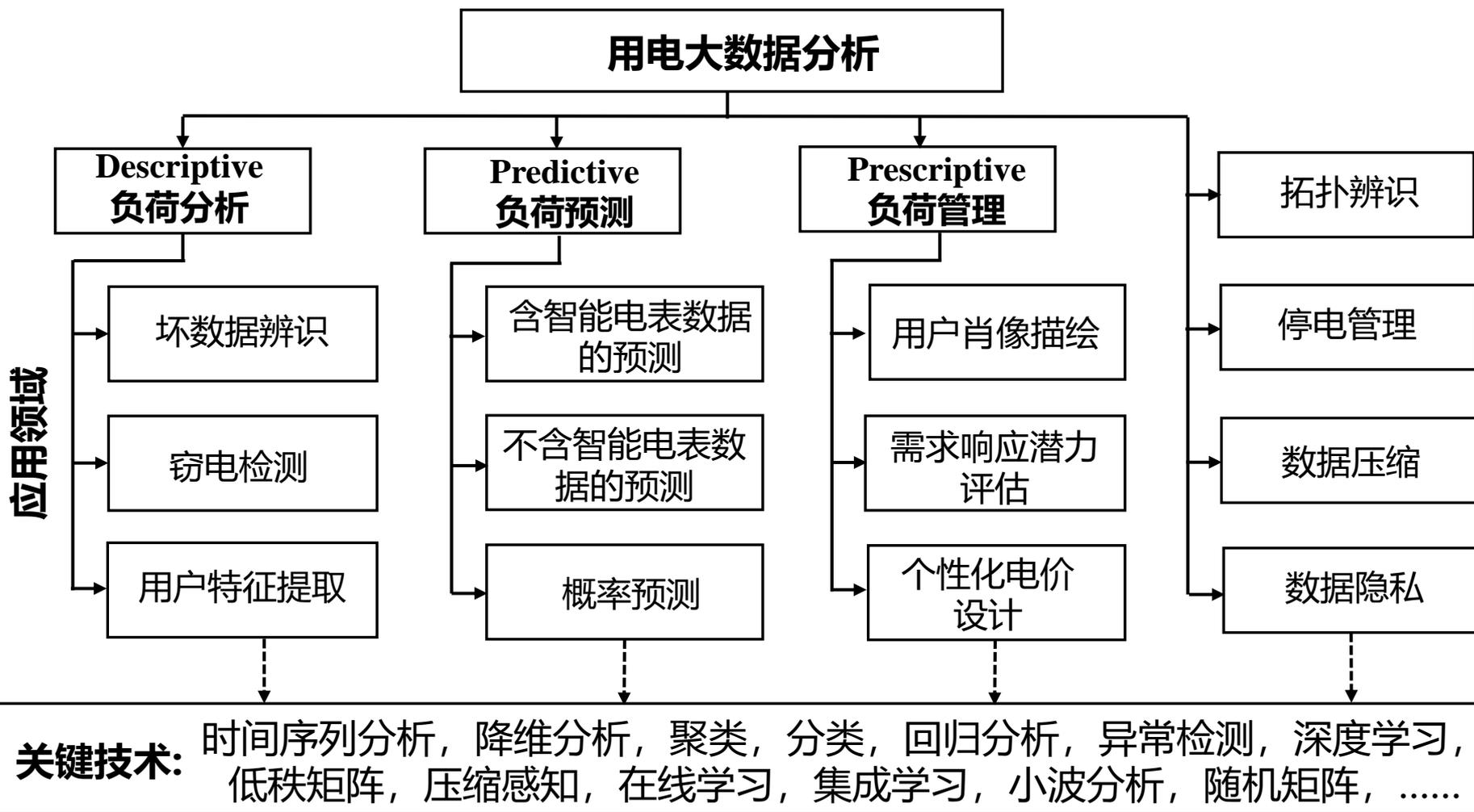
Velocity（产生速度快）

Value???（价值???)

Volume（海量）

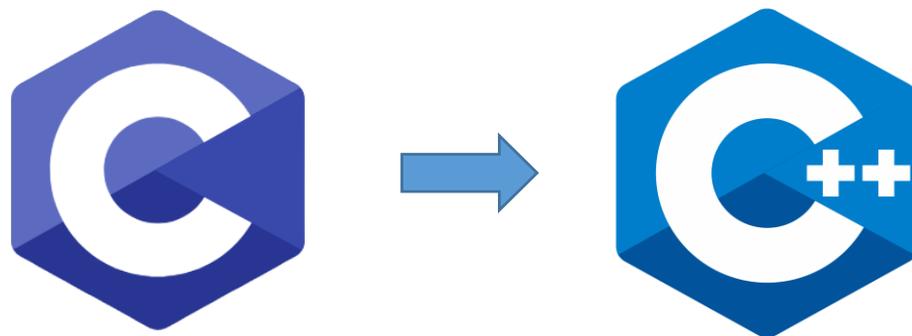
1000万智能电表每15分钟采一个点 每天产生60GB 数据，每年产生 21TB数据

◆ 国内外研究现状



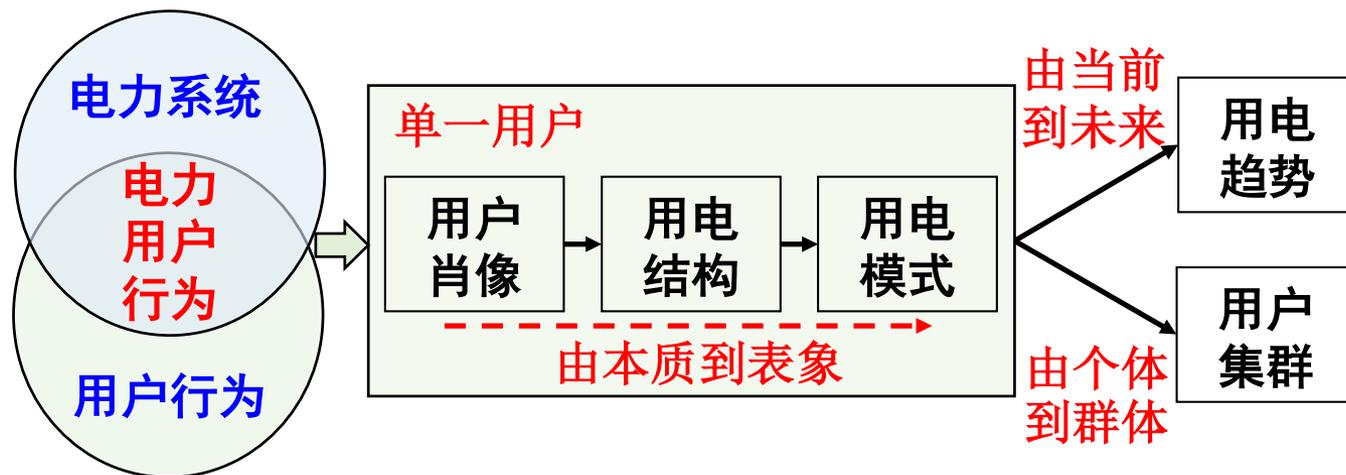
◆ 电力用户行为模型及其意义

目前研究类似于“面向过程”的程序编写，缺乏对电力用户行为体系化的认知，没有“面向对象”的整体设计。



以“用户行为”为核心，解读电力用户行为内涵，构建电力用户行为模型，并利用数据驱动的手段对用户行为进行全面的建模。

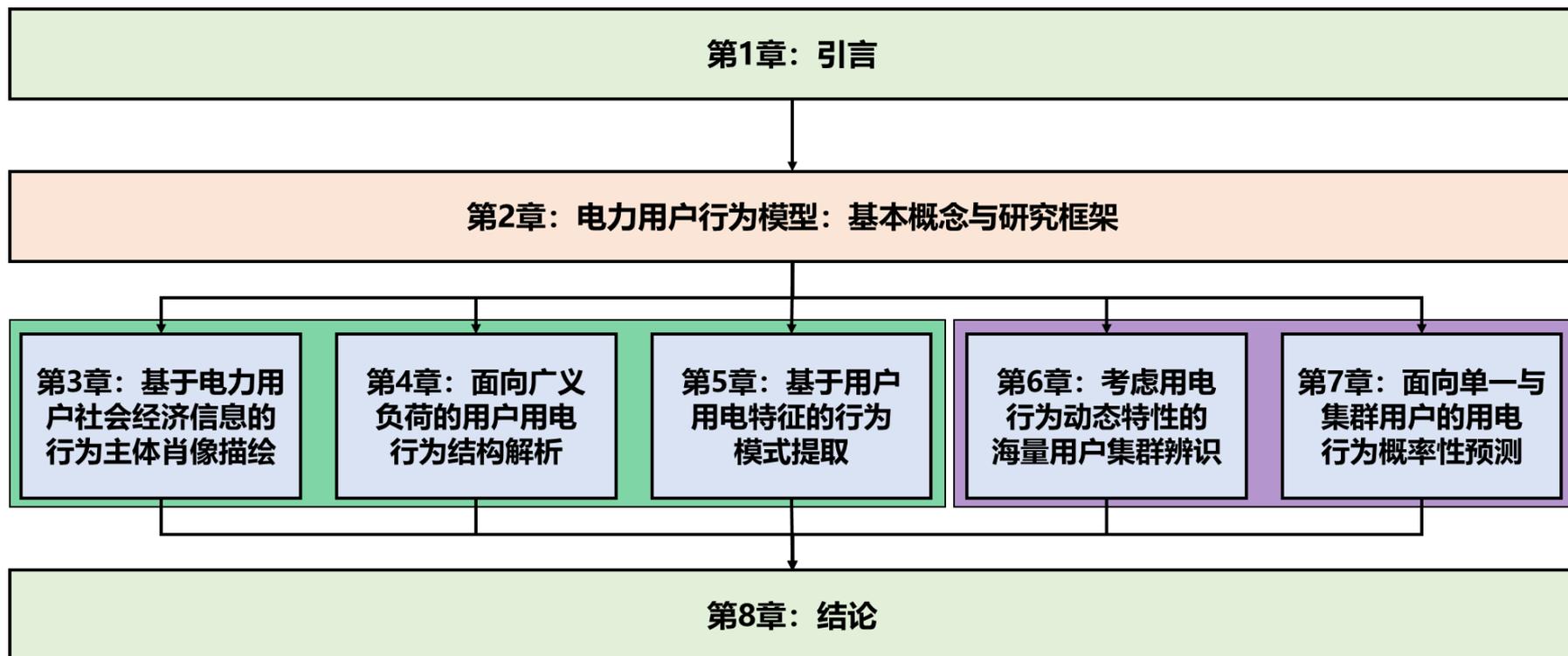
- 感知用户复杂变化
- 引导用户用电行为
- 提升用户管理水平
- 实现用户精准服务

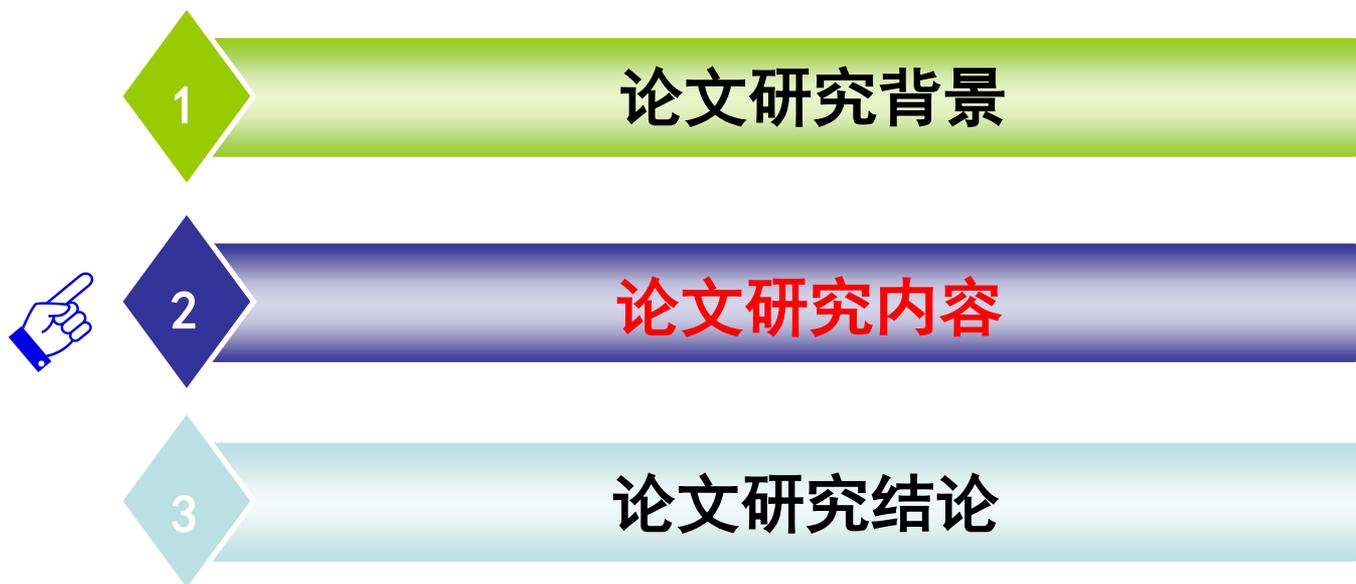


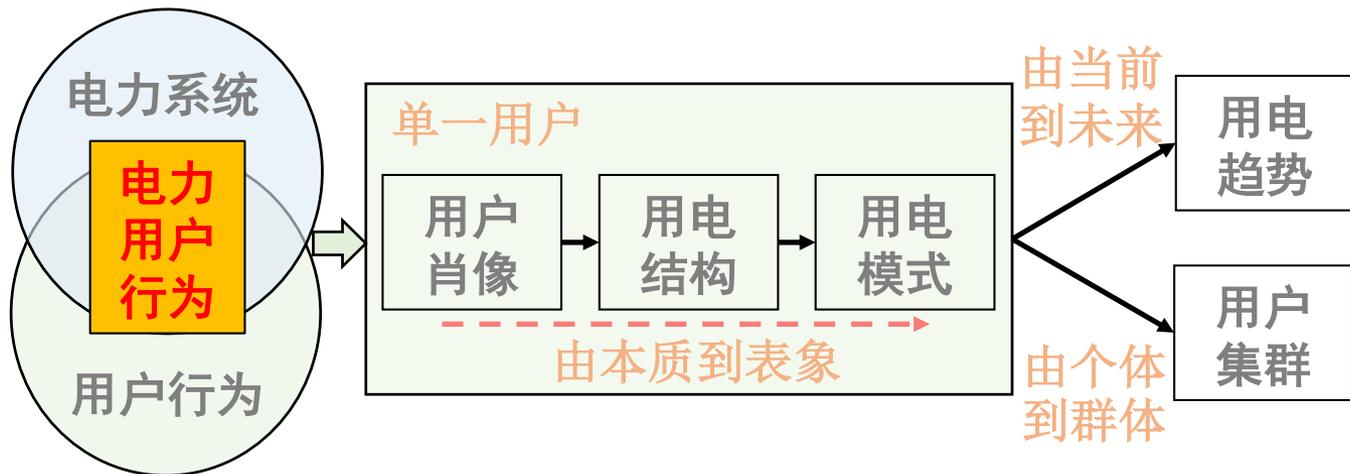
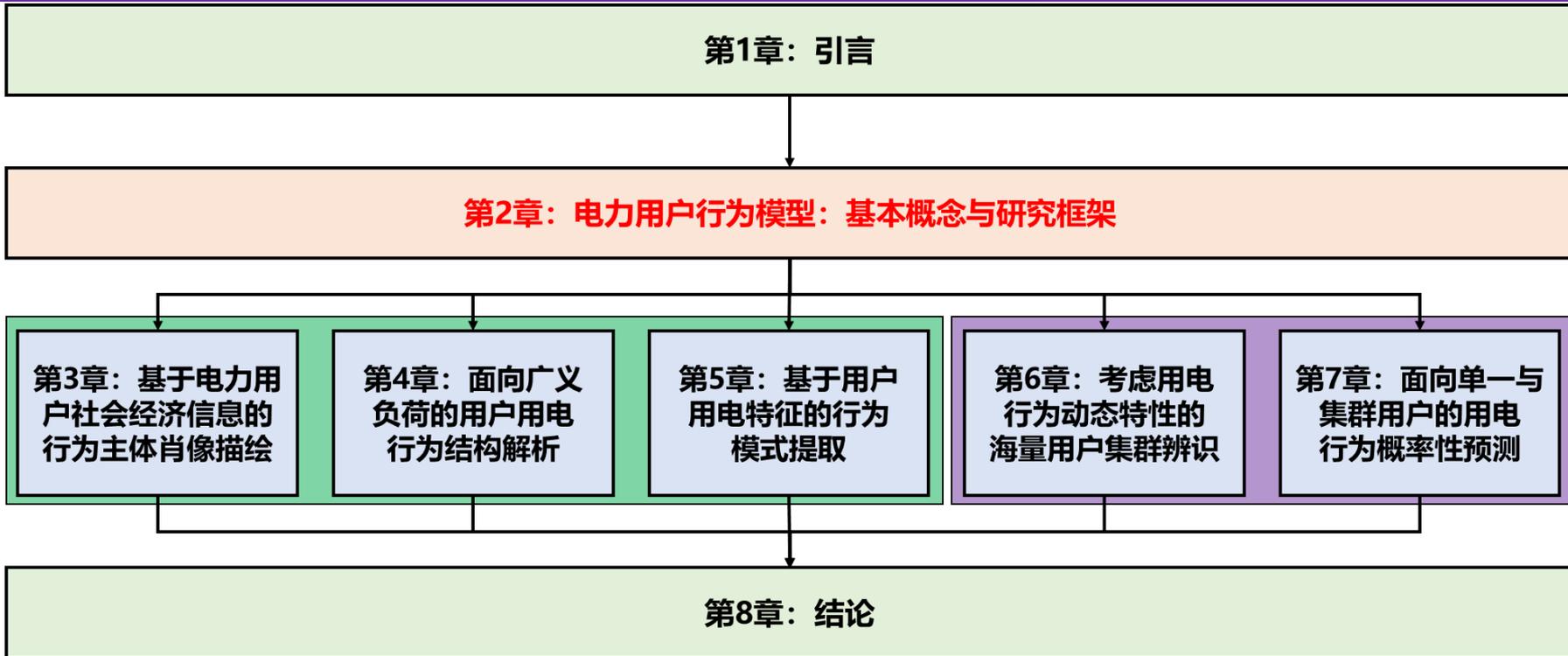
在“电力用户行为模型”核心理念引领下：

- 以广泛用户互动需求为问题导向，从多维度解读用户行为模型；
- 以先进数据挖掘方法为建模手段，从电力用户肖像、用电结构、用电模式、用电趋势和用户集群五个方面开展研究；
- 构建了由本质到表象、由个体到群体、由当前到未来的电力用户行为建模理论与方法研究框架。

数据驱动的电力用户行为模型与分析方法研究





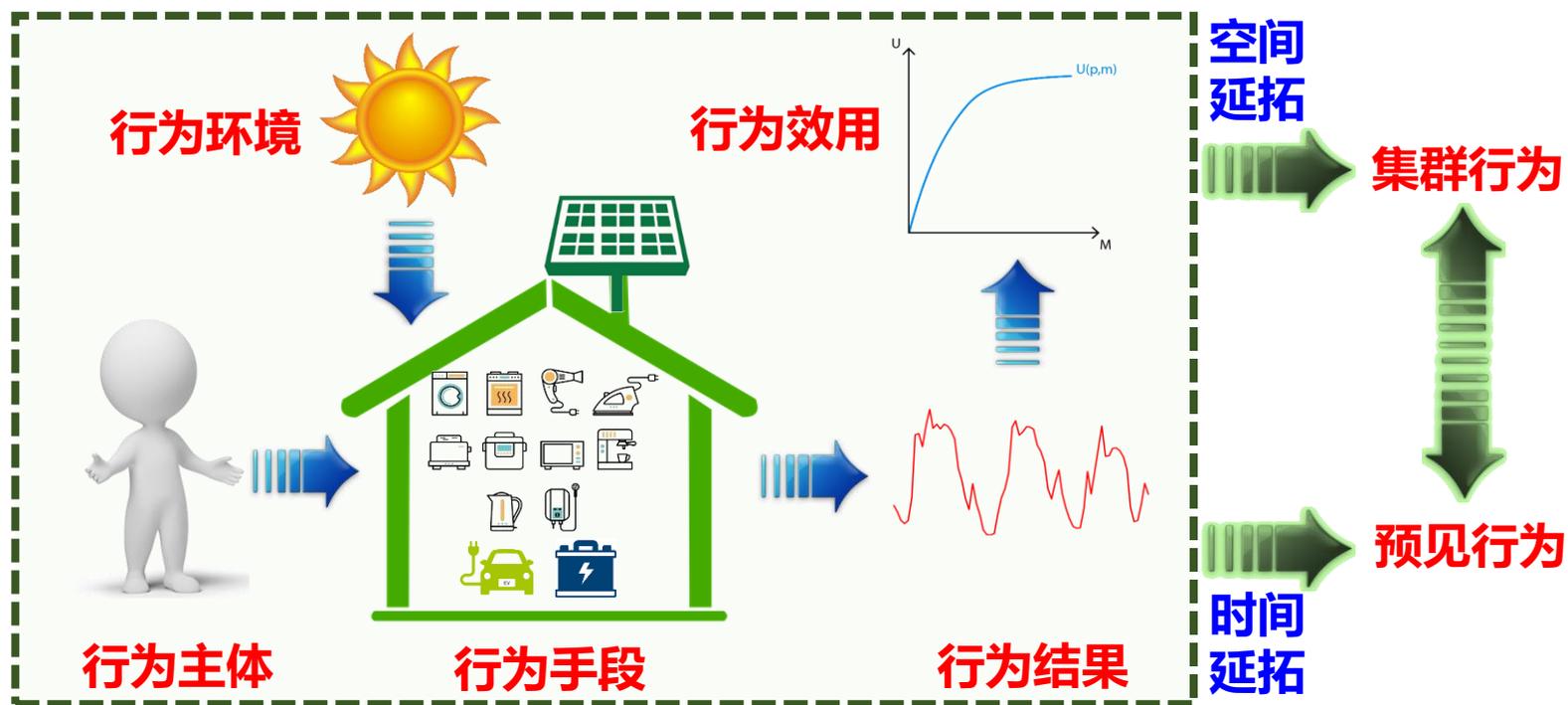


2. 电力用户行为模型：基本概念与研究框架

◆ 基本概念

电力用户行为是指电力用户在外界环境影响下所产生的用电活动和表现出来与用电相关的态度，一般包括**显性行为**和**隐性行为**。

在社会学领域，人的行为一般由行为主体、行为环境、行为手段、行为结果和行为客体五个部分组成



2. 电力用户行为模型：基本概念与研究框架

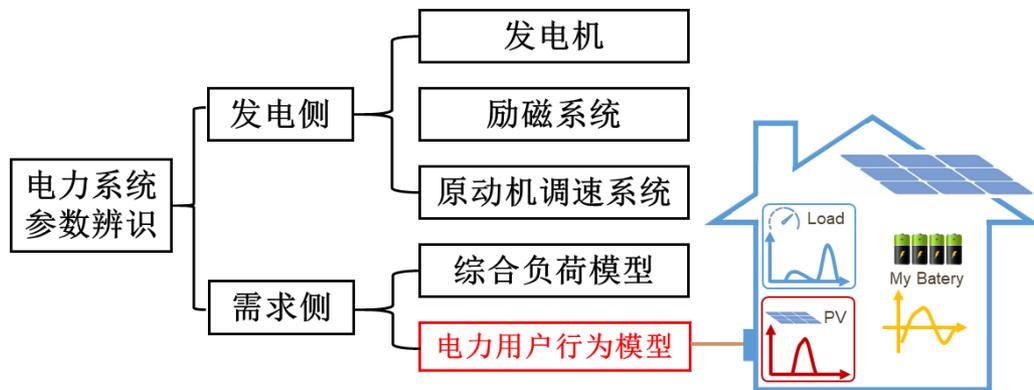
◆ 基本概念

电力用户行为模型就是**利用多元化信息，通过优化建模、数据分析等手段**，揭示和描述行为主体、行为环境、行为手段、行为结果、行为效用、预见行为、集群行为的**内在特性及其相互关联关系**，实现对电力用户行为的一种抽象和特征的规范化表达。

- 电力用户行为模型是用户行为模型在电力系统中的应用与拓展：



- 综合负荷建模和用户行为建模构成需求侧建模的一体两面：



2. 电力用户行为模型：基本概念与研究框架

◆ 数学表达

数学符号	物理内涵
\mathbb{C}/\mathbf{C}	用户属性空间/集合
c_j	用户第 j 个属性
\mathbb{E}/\mathbf{E}	环境因素空间/集合
e_k	第 k 个环境因素
I/i	用户集合/索引号
T/t	时间集合/索引号
A/a	产消元件集合/索引号
P	有功功率
O	用户总效用
g_i	第 i 个用户的效用函数
S_n	第 n 个用户集群

以上五个表达式将行为主体 \mathbf{C}_i^t 、行为环境 \mathbf{E}_i^t 、行为手段 $P_{i,a}^t$ 、行为结果 P_i^t 、行为效用 g_i^t 等进行耦合

➤ 电力用户的效用趋优性：

$$\arg \max_P O_i = \arg \max_P \sum_t g_i^t(P_{i,a}^t) \Big|_{\mathbf{C}_i^t, \mathbf{E}_i^t}$$

➤ 电力用户的线性叠加性：

$$P_i^t = \sum_{a \in A} P_{i,a}^t$$

➤ 电力用户的集群特性：

$$\begin{aligned} \max_{S_1, S_2, \dots, S_N} & \prod_{n=1}^N \prod_{i \in S_n} \text{Prob}(F_i^t | i \in S_n) \\ \text{s.t.} & S_1 \cup S_2 \cup \dots \cup S_N = I \\ & S_1 \cap S_2 \cap \dots \cap S_N = \emptyset \end{aligned}$$

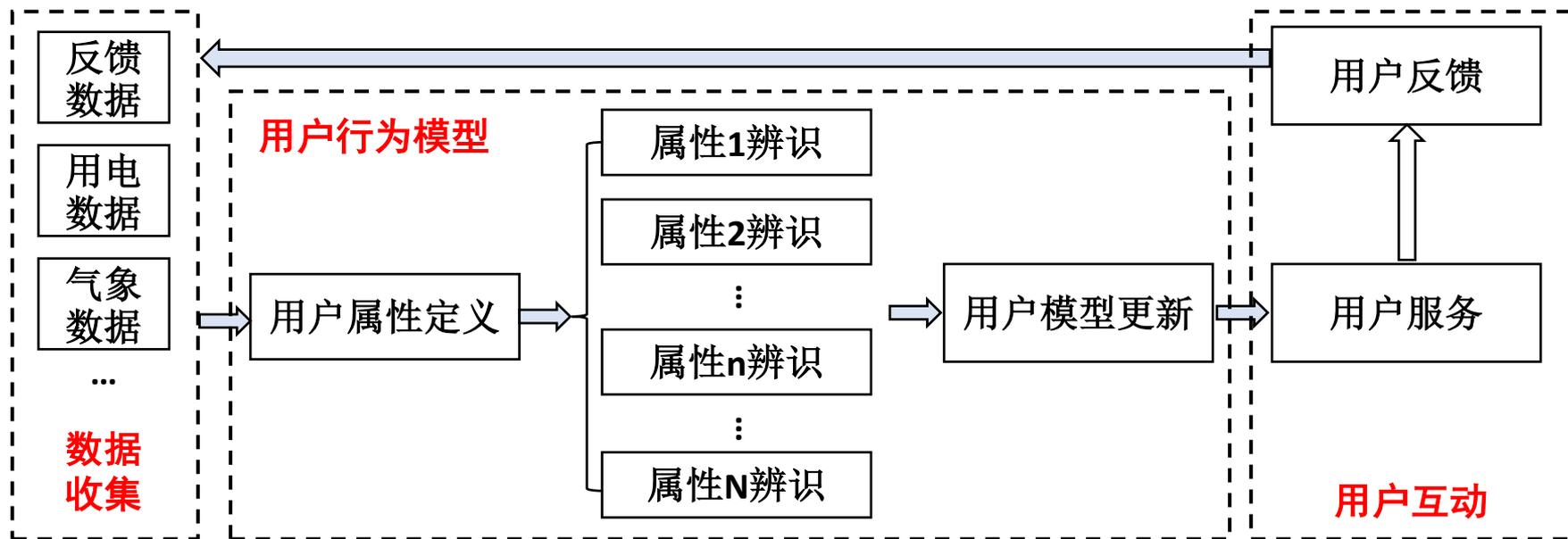
➤ 电力用户的行为可预见性：

$$\hat{P}_{i,a}^{t+h} = f_{i,a}(\mathbf{C}_i^t, \mathbf{E}_i^t, \hat{\mathbf{E}}_i^{t+h}, P_{i,a}^t, t)$$

$$\hat{P}_i^{t+h} = f_i(\mathbf{C}_i^t, \mathbf{E}_i^t, \hat{\mathbf{E}}_i^{t+h}, P_i^t, t)$$

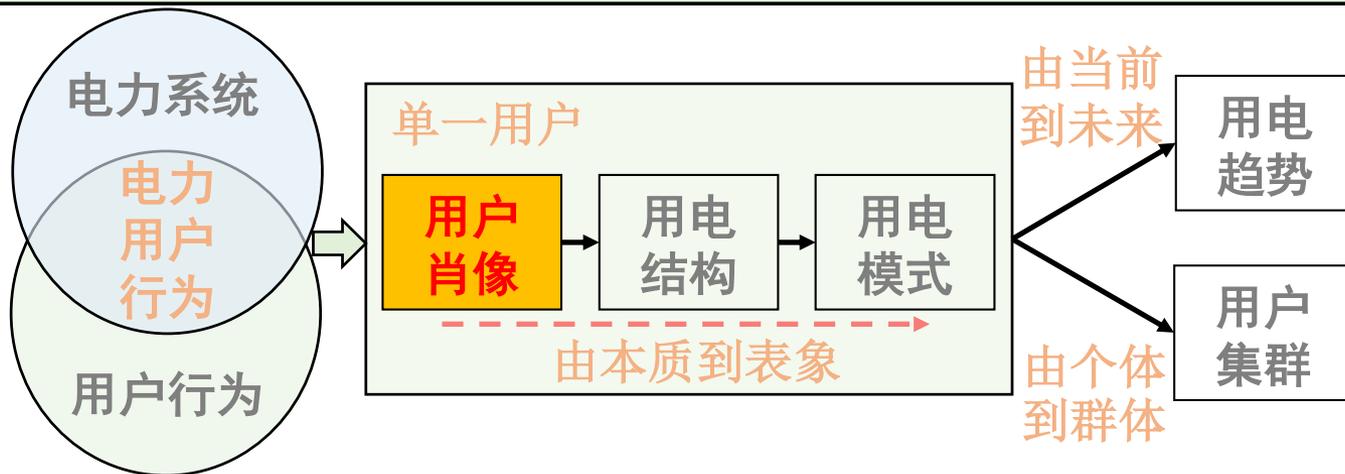
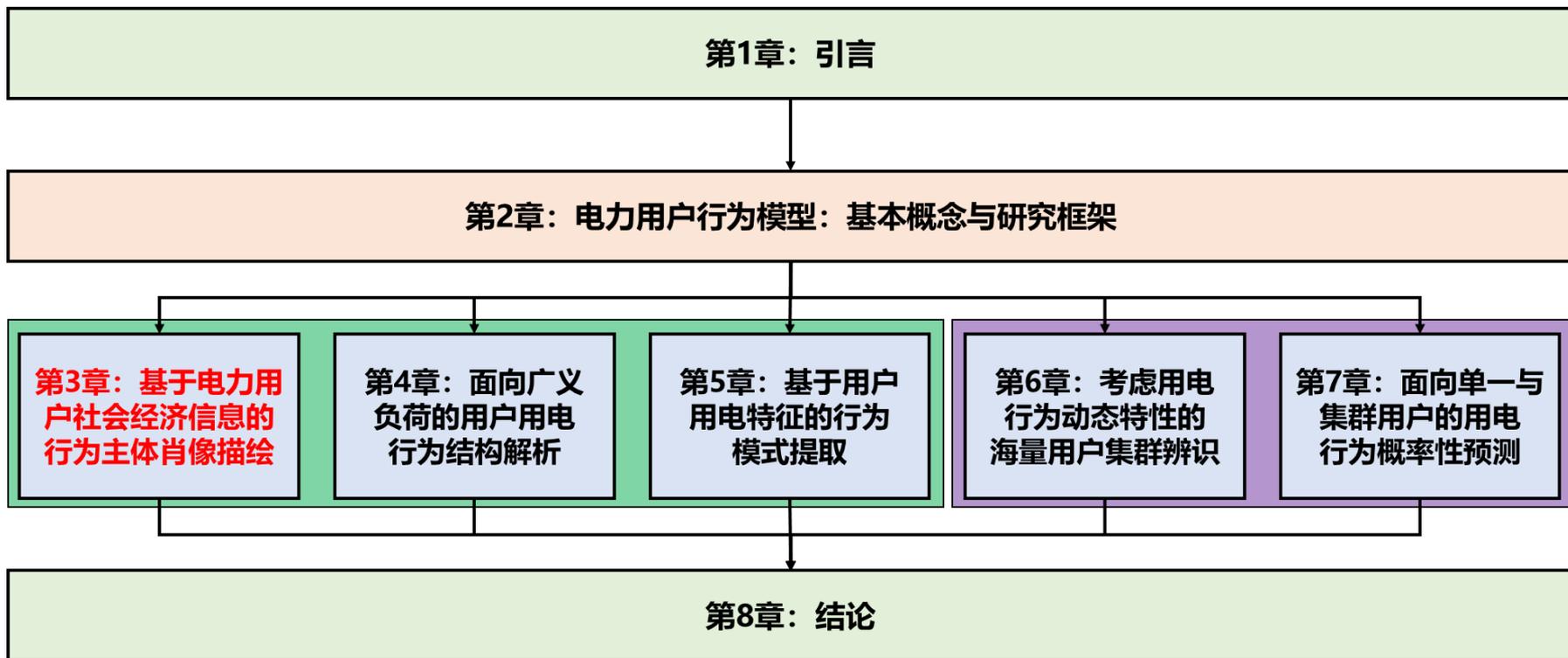
2. 电力用户行为模型：基本概念与研究框架

◆ 研究框架



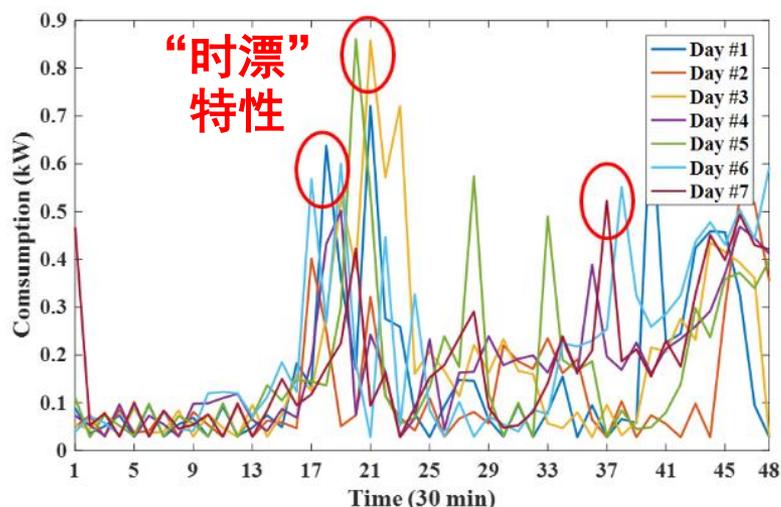
主要包括数据收集、用户行为模型和用户互动三个模块。

在这三个模块中，**数据收集是基础、用户行为模型是核心、用户互动是目的**，三者依次递进构成闭环，实现电力用户行为模型的不断更新和优化。



3. 基于电力用户社会经济信息的行为主体肖像描绘

用户用电行为受到多方面因素的影响，包括**内在因素**如用户自身的社会经济信息和**外在因素**如环境温度等，其中**用户自身的社会经济信息**包括用户的**职业、年龄、薪资水平、所居住房屋大小**等。



研究意义：

- 个性化用电服务；
- 第三方机构推销；
- 隐私保护的反思；



研究挑战：

- 如何建模？
- 高维复杂性？
- “时漂”特性？

3. 基于电力用户社会经济信息的行为主体肖像描绘

建模方案：

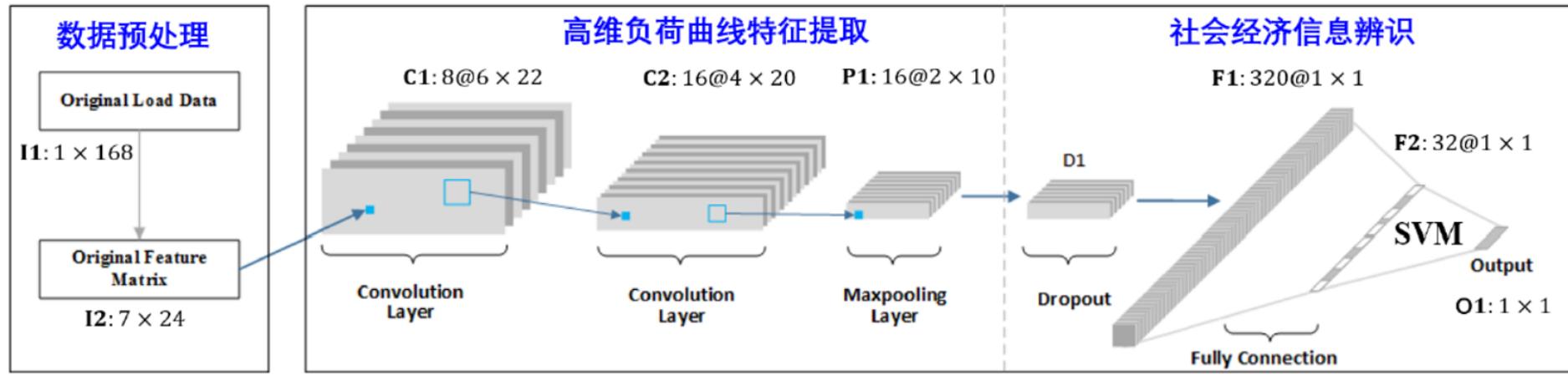
将用户的性别、年龄、就业情况、社会阶层和住所等信息进行离散化，构成用户的“**标签**”，进而**将用户社会经济信息辨识问题转化成为分类问题**。

构建负荷曲线到用户信息的关系本质上是一个**高维且强非线性的分类问题**，构建分类模型前需要进行特征提取，以提升模型对非线性的描述能力。

编号	问卷编号	具体社会经济信息	选项	对应人数
1	300	主要收入者年龄	青年 (<35)	436
			中年 (35~65)	2819
			老年 (>65)	953
2	310	主要收入者是否退休	是	1285
			否	2947
3	401	主要收入者社会阶层	A 或 B	642
			C1 或 C2	1840
			D 或 E	1593
4	410	是否有小孩	是	1229
			否	3003
5	450	房屋类型	平房	2189
			楼房	1964
6	453	房屋新旧	旧 (>30)	2151
			新 (<30)	2077
7	460	房屋卧室数量	很少 (<3)	404
			少 (=3)	1884
			多 (=4)	1470
			较多 (>4)	474
8	4704	厨具类型	电厨具	1272
			非电厨具	2960
9	4905	节能灯占比	一半以下	2041
			一半以上	2191
10	6103	房屋面积	小 (<100)	232
			中等 (>100&<200)	1198
			大 (>200)	351

3. 基于电力用户社会经济信息的行为主体肖像描绘

构建深度卷积学习网络，自动提取有效特征，考虑用电内在关联。



1、数据重排：问题的化归转化

2、特征提取

$$S_{i,j} = G_j(\mathbf{c}_i, \mathbf{W}_{1,j})$$

3、“标签”辨识

$$\hat{y}_{i,j} = F_j(\mathbf{s}_{i,j}, \mathbf{W}_{2,j})$$

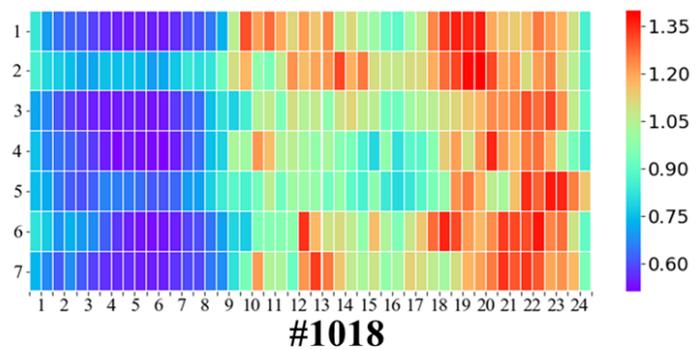
➤ 利用数据拓展、随机遗忘、权重衰减、随机梯度下降训练等手段减少模型的过拟合风险。

层编号	层类型	超参数	待训练的参数数量
C1	卷积层	输入数据大小: $7 \times 24 \times 1$ 卷积核大小: 2×3 卷积核数量: 8	56
C2	卷积层	输入数据大小: $6 \times 22 \times 8$ 卷积核大小: 3×3 卷积核数量: 16	160
P1	最大化池化层	输入数据大小: $4 \times 20 \times 16$	无
Dr1	Dropout 层	无	无
F1	平铺层	无	无
D1	全连接层	输入数据大小: 320 神经元数量: 32	10560
D2	Softmax 层	输入数据大小: 32 神经元数量: 1	32

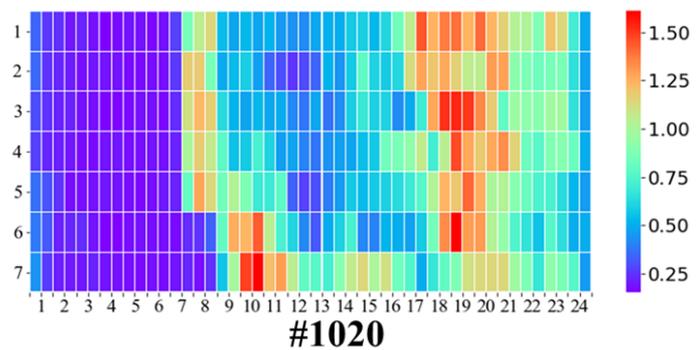
3. 基于电力用户社会经济信息的行为主体肖像描绘

若干简单的观察：

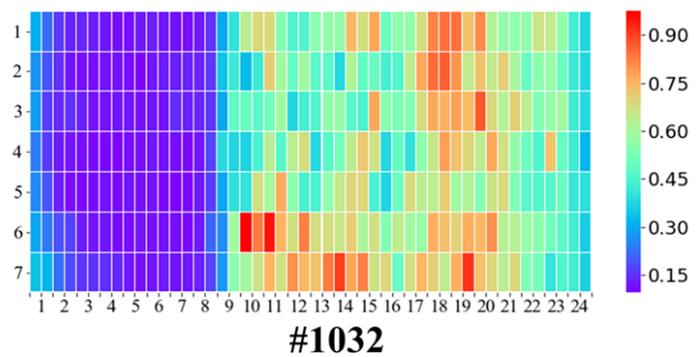
- 三个用户在晚间6点到8点之间都保持较为活跃的用电；
- 已退休的用户#1018 在工作时间用电水平也维持在较高水平，而未退休的用户#1020 和用户#1032 除周末外，在工作时间段用电水平较低；
- 用户#1032 的卧室数量较少，其用电水平也相对较低；
- 家中有小孩的用户#1018 在较晚的时候仍然有较高的用电水平，有家庭其他成员在夜间仍然保持活跃的用电，甚至为了在夜间照顾小孩。



老年 (>65) 已退休
家中无小孩 平房
卧室数量较多



中年 (35~65) 未退休
家中有小孩 楼房
卧室数量多



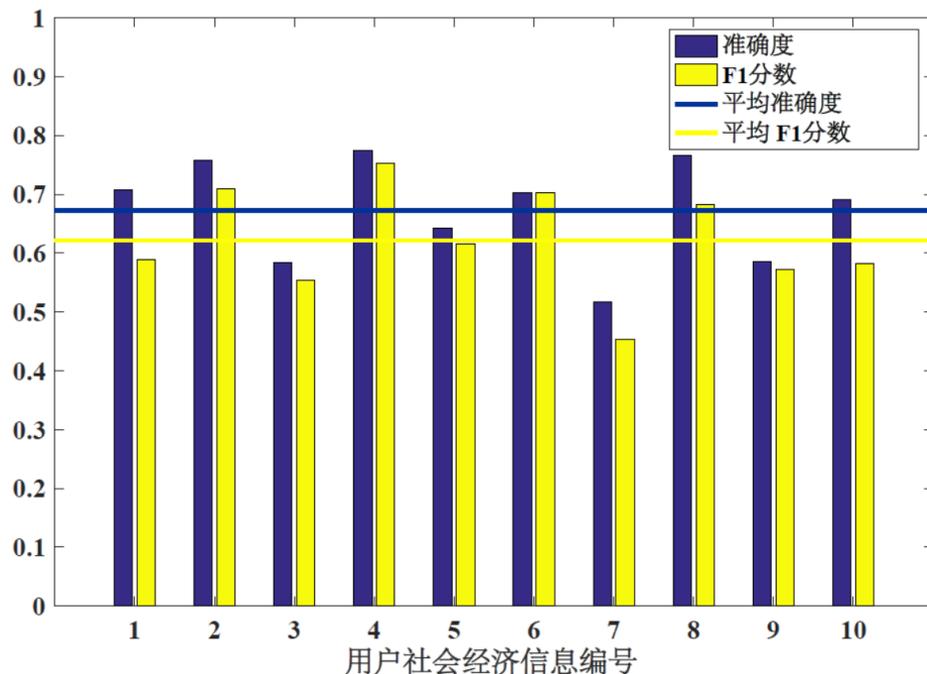
中年 (35~65) 未退休
家中有小孩 楼房
卧室数量少

3. 基于电力用户社会经济信息的行为主体肖像描绘

实验结果分析

特征2（家里主心骨是否退休）、特征4（家里是否有小孩）和特征8（厨房是用电还是用气）的精度都超过75%，说明这些特征与用户用电行为具有较强的相关性；

特征7（卧室数量）和特征9（节能灯比例）的精度都低于60%，说明这些特征对用电行为的影响不明显。



3. 基于电力用户社会经济信息的行为主体肖像描绘

实验结果分析

不同方法的分类准确度

	BG	MF	SVM	LS	PS	SS	CS	本文方法	改进 1	改进 2	改进 3
1	0.511	0.59	0.648	0.664	0.667	0.666	0.688	0.708	3.15%	2.95%	29.47%
2	0.577	0.73	0.697	0.702	0.696	0.693	0.748	0.758	2.47%	1.37%	24.39%
3	0.382	0.53	0.506	0.498	0.513	0.5	0.55	0.584	3.77%	6.09%	37.66%
4	0.588	0.73	0.709	0.715	0.73	0.714	0.748	0.774	2.47%	3.45%	24.39%
5	0.501	0.59	0.572	0.567	0.567	0.531	0.621	0.643	5.25%	3.53%	16.65%
6	0.5	0.64	0.564	0.577	0.582	0.566	0.665	0.703	3.91%	5.69%	22.77%
7	0.416	0.39	0.472	0.476	0.49	0.467	0.501	0.517	2.24%	3.17%	13.77%
8	0.579	0.71	0.687	0.675	0.694	0.698	0.739	0.766	4.08%	3.70%	22.60%
9	0.501	0.55	0.511	0.492	0.528	0.525	0.565	0.586	2.73%	3.70%	7.13%
10	0.508	0.5	0.643	0.649	0.644	0.639	0.658	0.691	1.39%	4.96%	24.41%

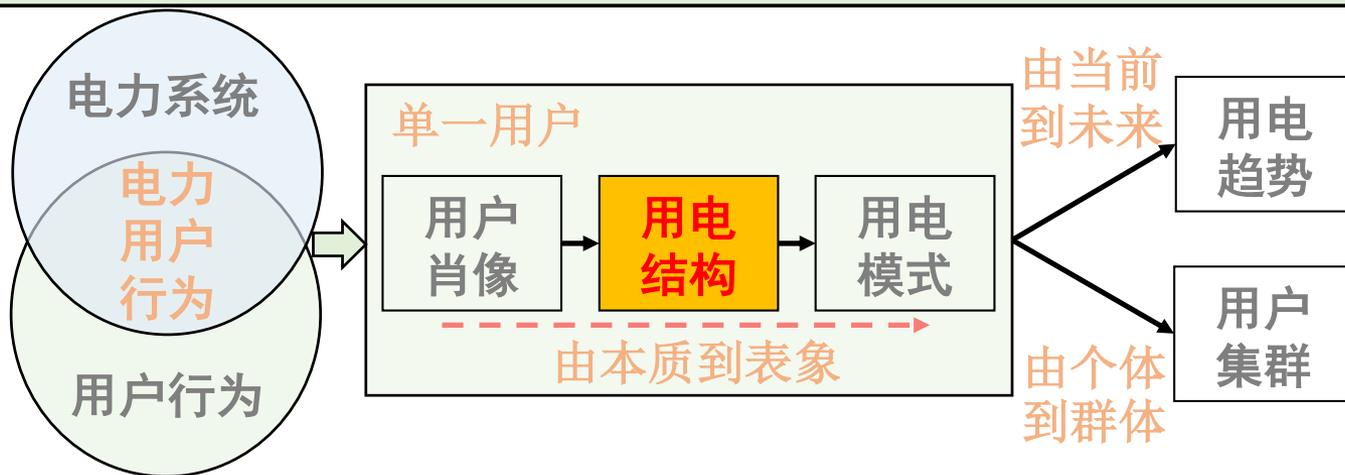
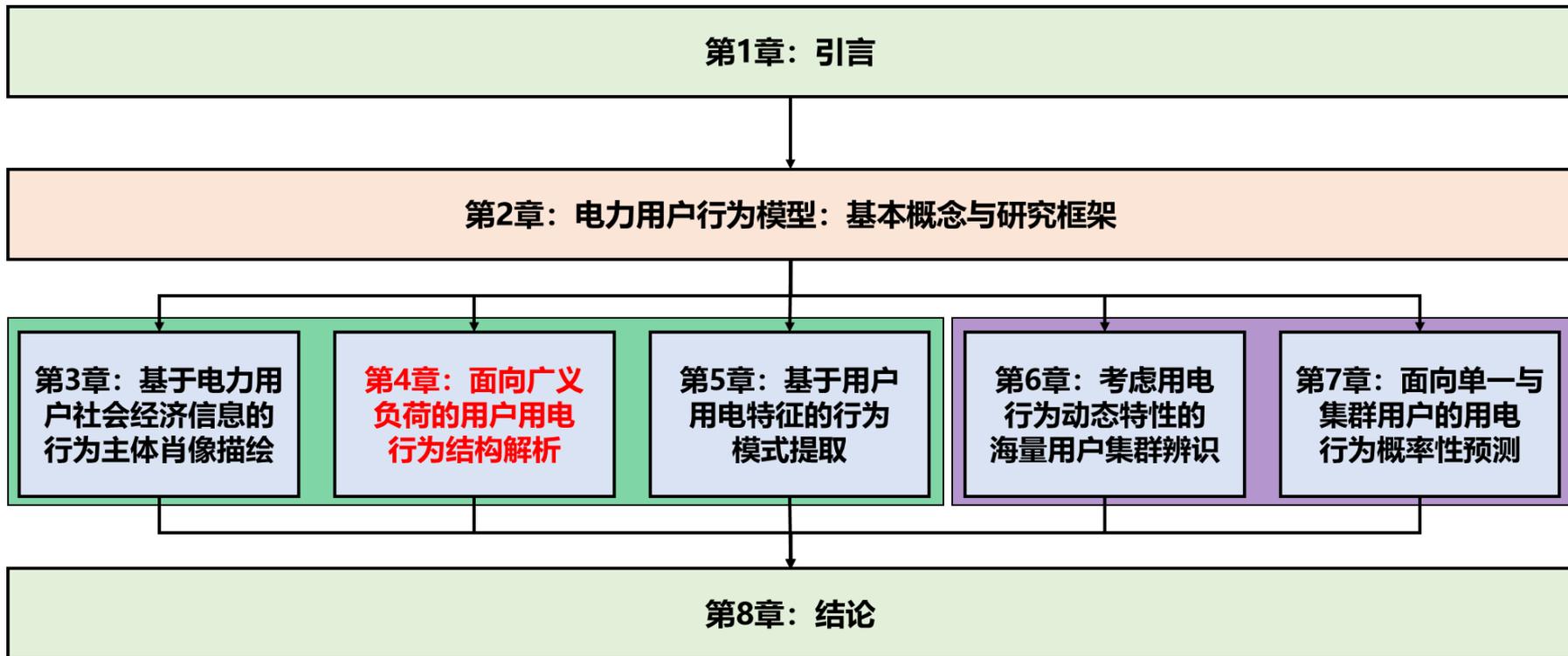
效果：各种用户信息的平均**识别精度高达67%**

与传统人工提取特征和浅层学习模型相比，肖像描绘的精度提高5%左右。

$$\arg \max_P O_i = \arg \max_P \sum_t g_i^t(P_{i,a}^t) \Big|_{C_i^t, E_i^t} \quad \longrightarrow \quad \mathbf{E}_i = f(P_i^t)$$

用户肖像 用电曲线

- 将行为主体肖像描绘问题转化为一个高维非线性的用户社会经济信息分类与辨识问题；
- 提出了基于深度卷积网络的用户用电曲线与用户社会经济关联关系构建方法；
- 本方法能够有效提升辨识精度，为个性化服务、用户互动和隐私保护研究提供参考。



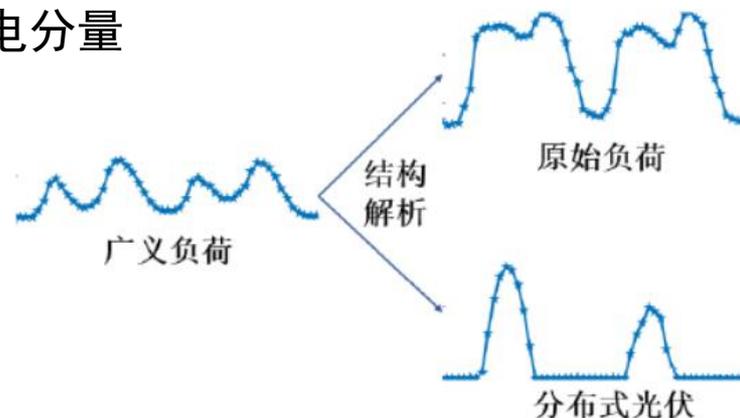
4. 面向广义负荷的用户用电行为结构解析

让用户用电行为变得更具“可观性”是增加用户用电行为“可控性”，开展家庭能量管理、实施需求响应等的重要手段。

用户用电行为结构解析的两种解读：

$$P_i^t = \sum_{a \in A} P_{i,a}^t$$

- 物理结构解析：直接根据总用电负荷对用户某一或某些设备的运行状态进行分解：
 - 非侵入式辨识
- 应用结构解析：面向不同应用对用户不同的分量进行解析：
 - 气象敏感分量/电价敏感分量/基础用电分量
 - 季节分量/周分量/日分量
 - 低频稳定分量/高频随机分量



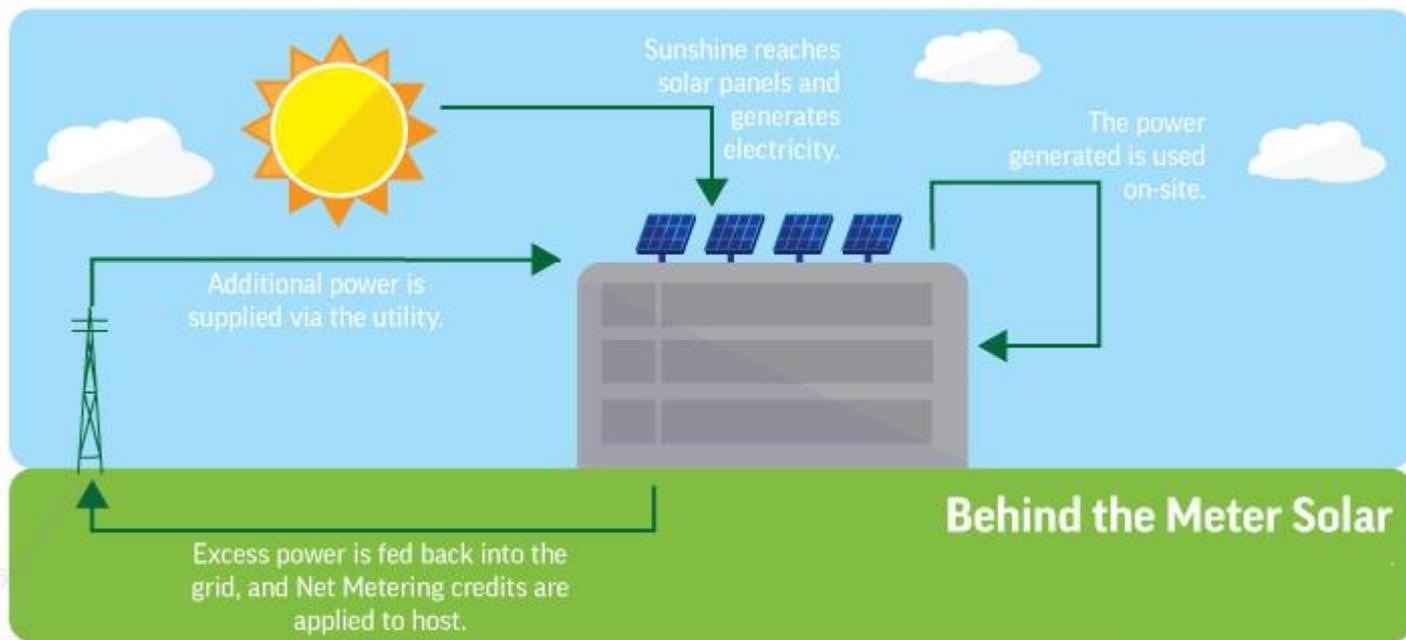
研究意义：

- 对用电行为及其变化规律的深入理解，增加用户“可控性”；
- 厘清不同负荷占比，为综合负荷建模、电力系统稳定分析提供支撑。

4. 面向广义负荷的用户用电行为结构解析

分布式光伏并网成为未来我国乃至全世界可再生能源发展的重要趋势，为了更好地消纳分布式可再生能源，保障电力系统安全稳定，需要对分布式可再生能源进行“实时感知”。

分布式不可见光伏的辨识是用电行为结构解析的重要部分，是传统非侵入式辨识的重要补充。



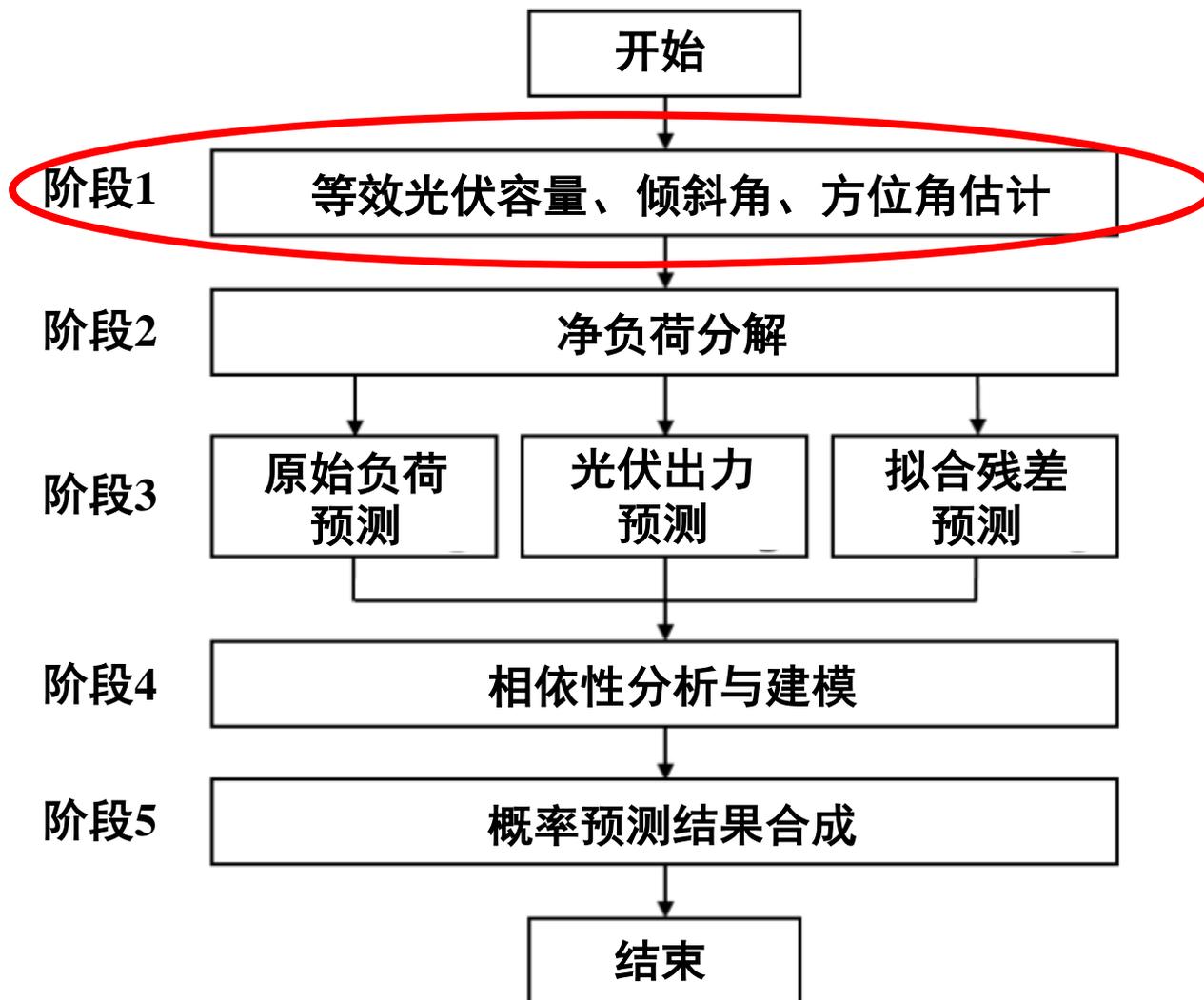
用电与气象等多元数据融合
数据模型与物理模型融合

不可见分布式光伏辨识方法特点：

- 融合多元数据
- 融合物理模型，引入等效的概念
- 用预测结果侧面反映辨识效果

4. 面向广义负荷的用户用电行为结构解析

含不可见光伏的广义负荷结构解析方案：



4. 面向广义负荷的用户用电行为结构解析

多元数据融合、物理模型与数据模型融合的不可见光伏辨识算法

初始化光伏容量



估计光伏出力

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



估计原始负荷

$$L_{E,t} = f_L(M_t, W_t, H_t, T_{A,t})$$



计算拟合残差

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



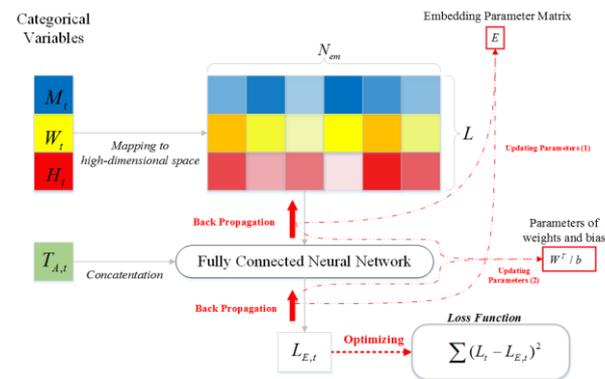
相关性分析

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



R达到最小时终止

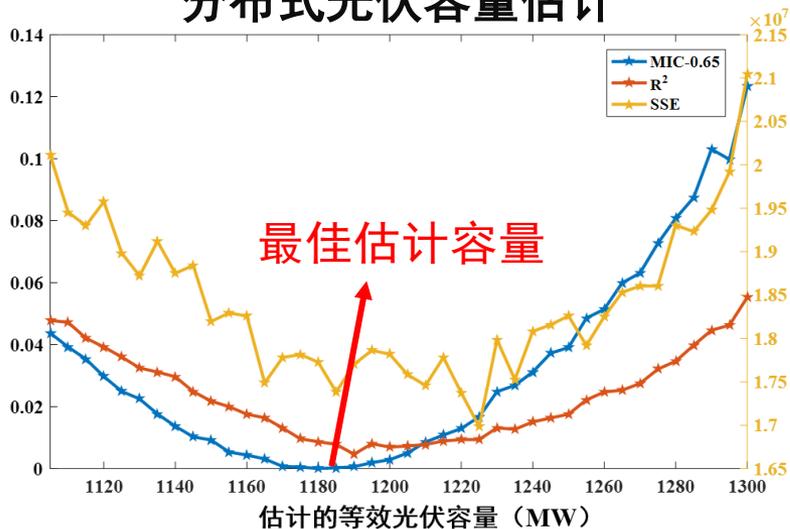
调整光伏容量



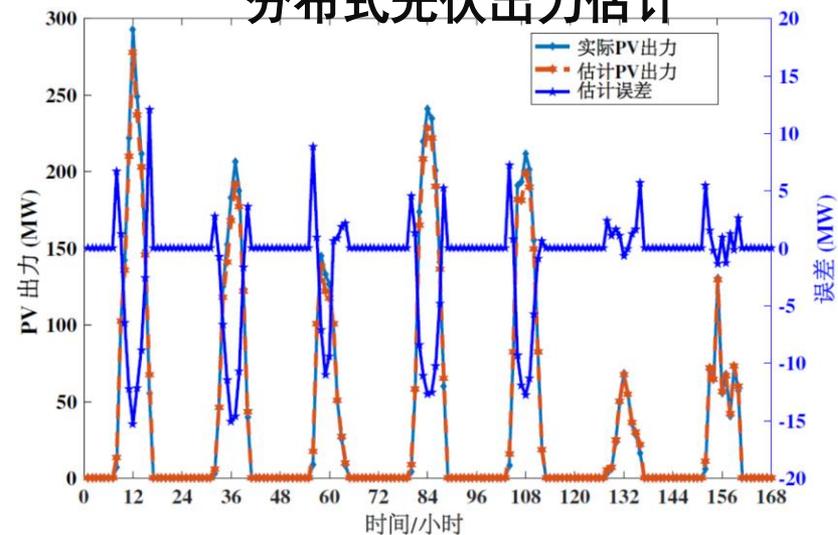
4. 面向广义负荷的用户用电行为结构解析

不可见分布式光伏辨识结果：

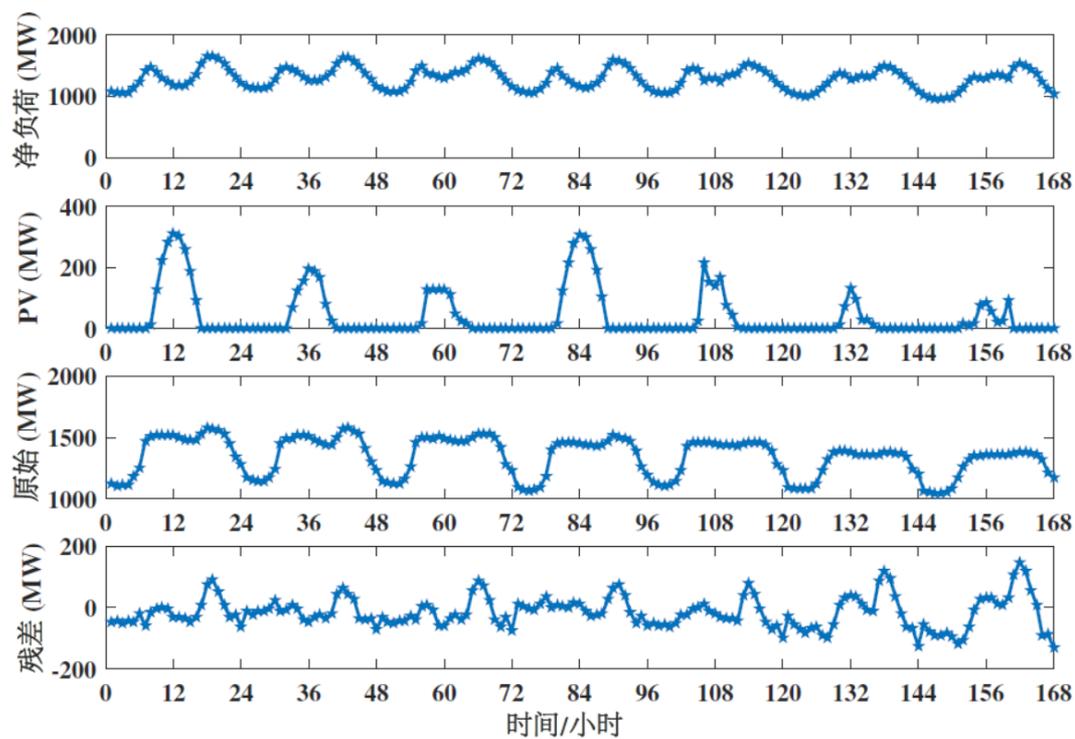
分布式光伏容量估计



分布式光伏出力估计



净负荷分离
(净负荷、光伏出力、实际原始负荷和残差)



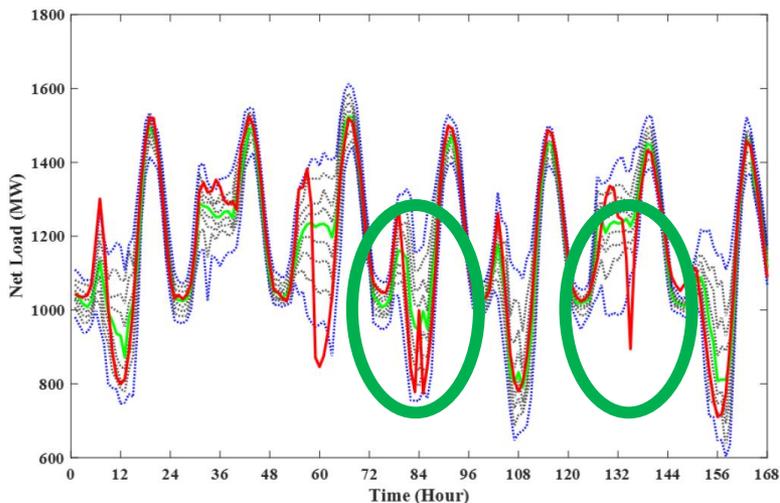
4. 面向广义负荷的用户用电行为结构解析

随着分布式光伏的增加，“检测-分离-预测-合成”的预测策略效果越明显

点预测效果(MW)

光伏渗透率	本章方法	方法 1	方法 2	方法 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

净负荷概率预测结果



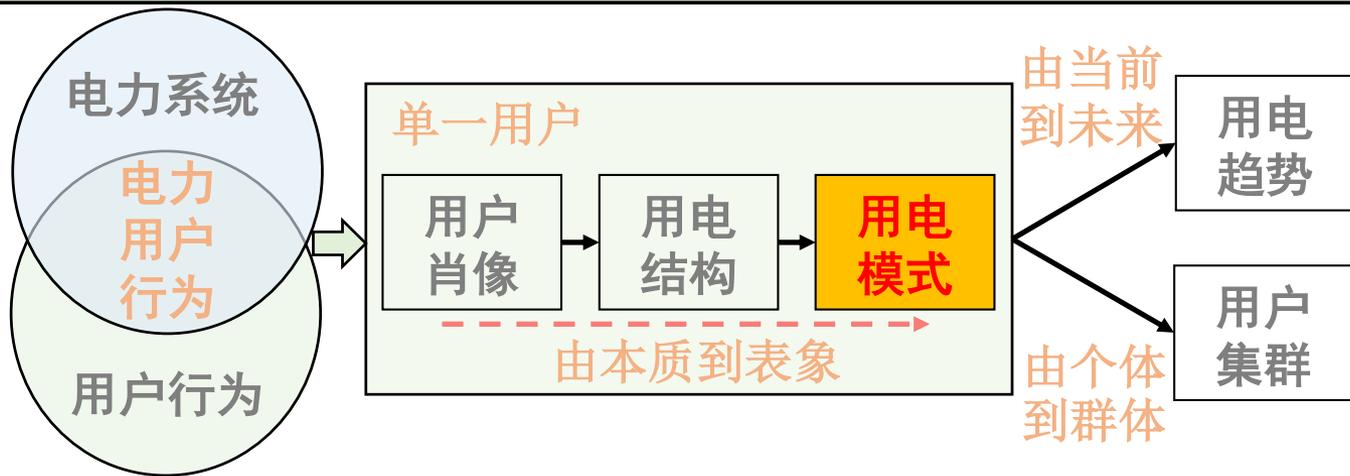
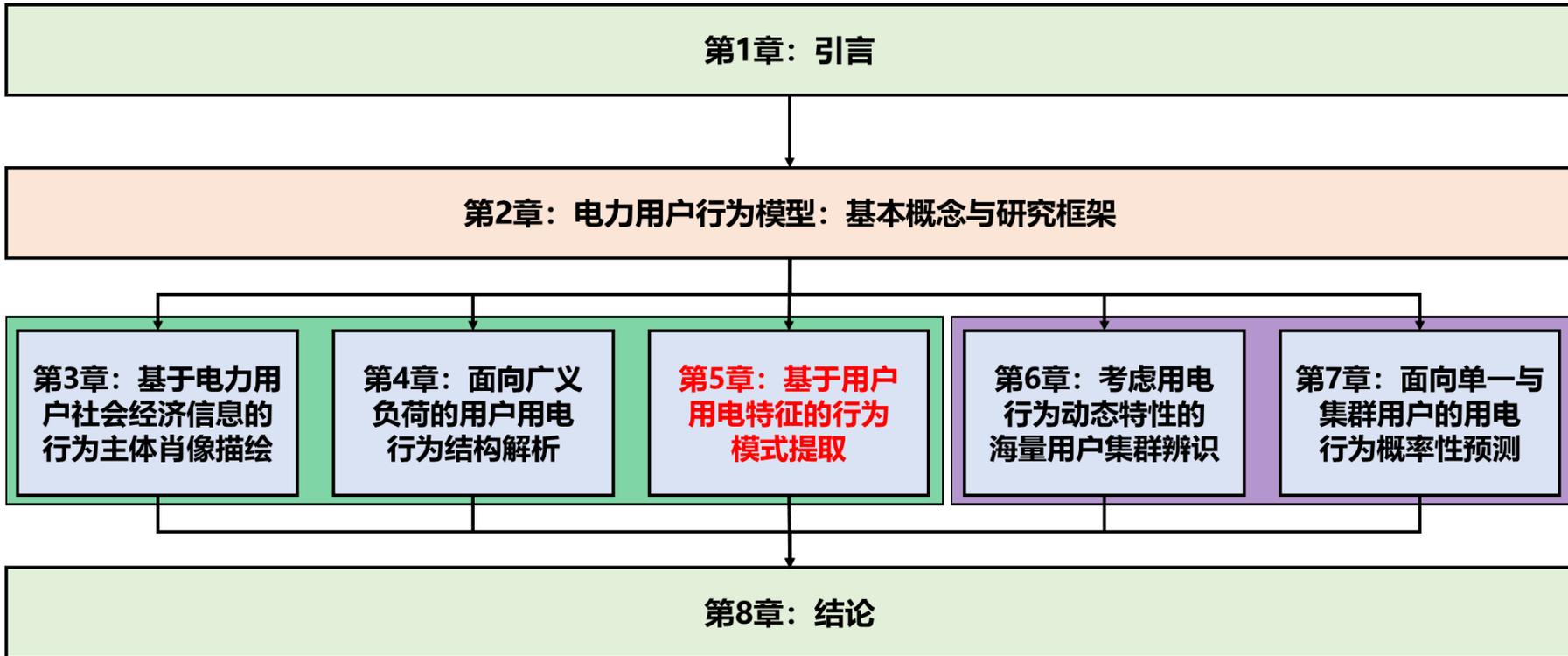
概率预测效果(MW)

光伏渗透率	本章方法	方法 4	方法 5	方法 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

$$P_i^t = \sum_{a \in A} P_{i,a}^t \quad \longrightarrow \quad P_{i,a}^t = f(P_{i,t}, \mathbf{E}_i)$$

不可见光伏 用电曲线 气象条件

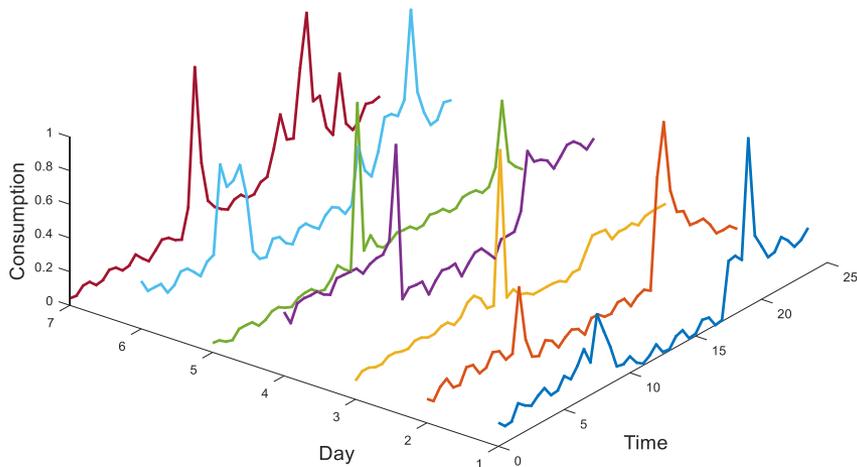
- 提出了基于多元数据融合、数据与物理模型融合的含不可见分布式光伏广义负荷结构辨识方法；
- 构建了“先分解预测，后概率合成”的全新概率预测框架，不可见光伏渗透率越高，预测改善程度越大，从侧面反映辨识“效益”；
- 提出了高维相依序列运算方法，在其他不确定性分析方面也有广泛潜在应用。



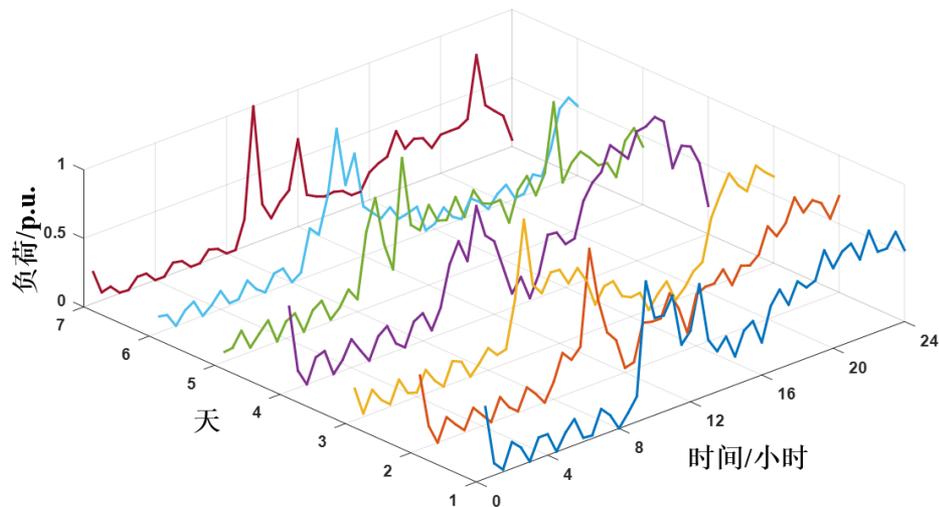
5. 基于用户用电特征的行为模式提取

用户用电受到多方面的影响，既表现出较大的波动性，也具有较强的不确定性，对用电曲线的认识可能需要“去粗取精”，发现用户用电的某些内在规律。

用户用电曲线的初步认知：



某居民用户一周负荷曲线



某工商业用户一周负荷曲线

用户用电特性

- 1) **稀疏性**：很多时段负荷值均接近于基荷，能“大致”看出一些固定的用电模式；
- 2) **多样性**：不同用户用电曲线差异大，甚至同一用户不同天也是多变的。

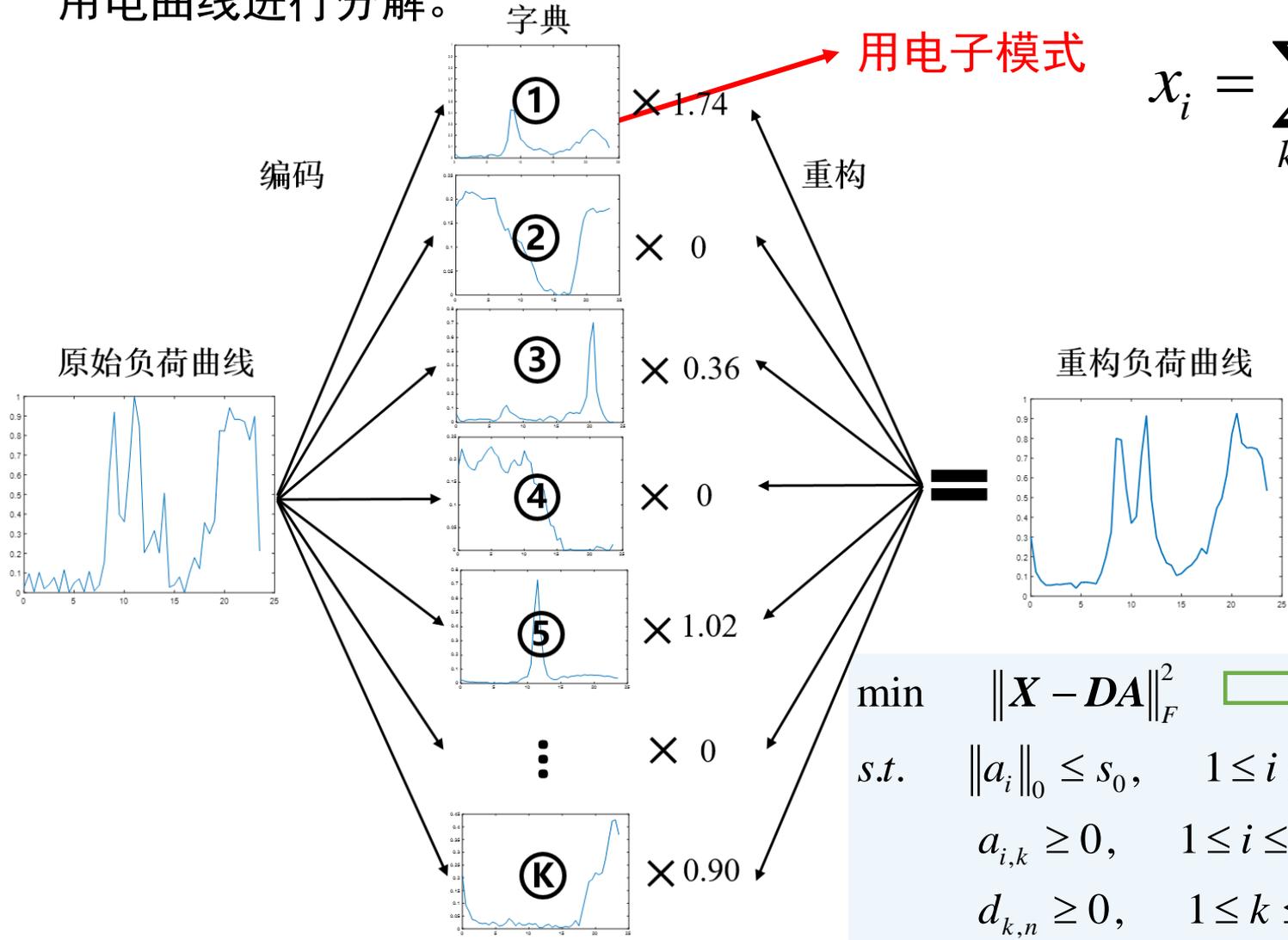
研究意义：

- 从稀疏性视角，对用户的用电特征进行提取，为用户类型辨识等提供基础；
- 从数据压缩视角，有效实现对海量用电数据的压缩，缓解数据通讯存储压力。

5. 基于用户用电特征的行为模式提取

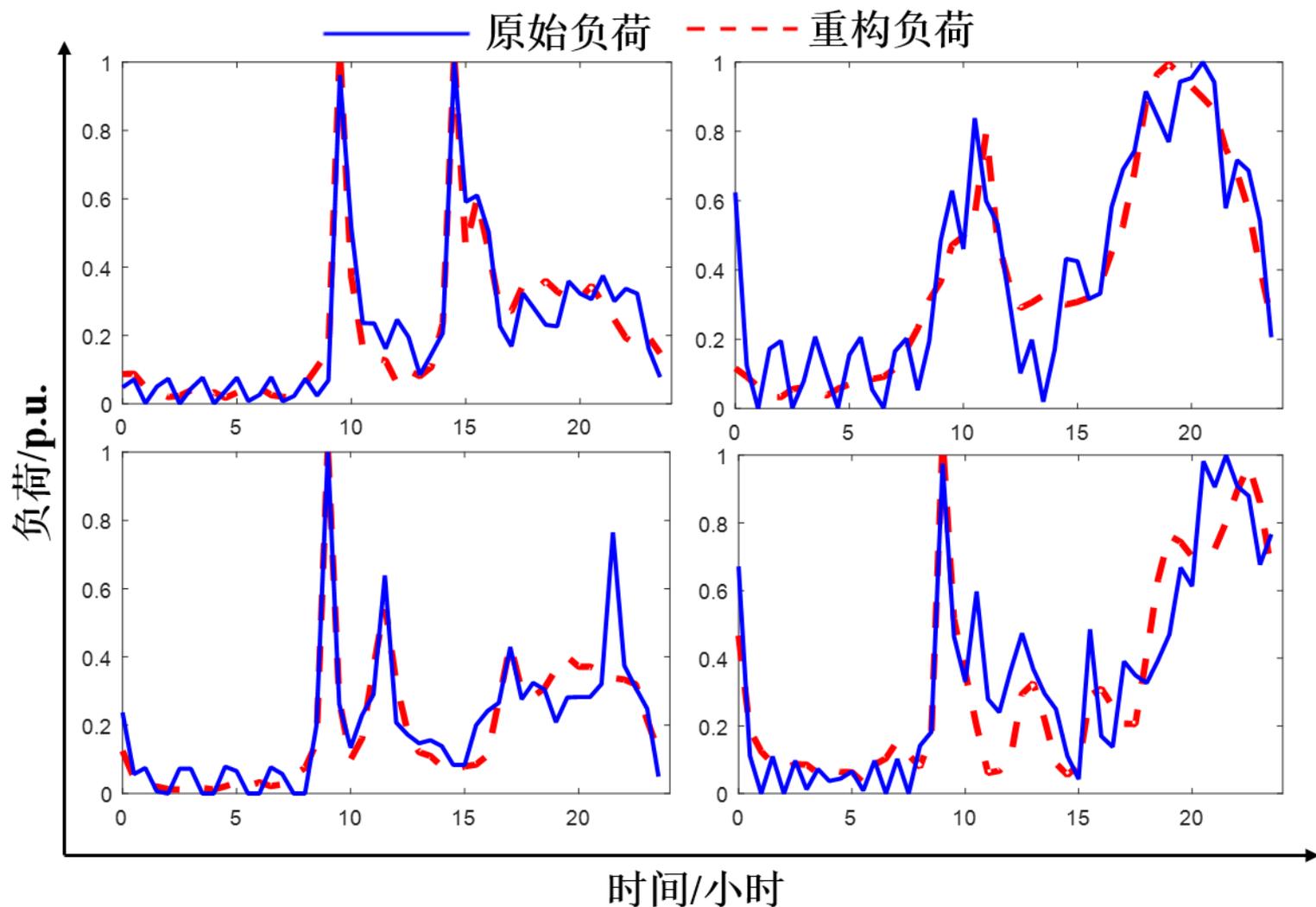
将用户用电曲线看作若干用电“子模式”的线性叠加，采用稀疏编码技术对用电曲线进行分解。

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



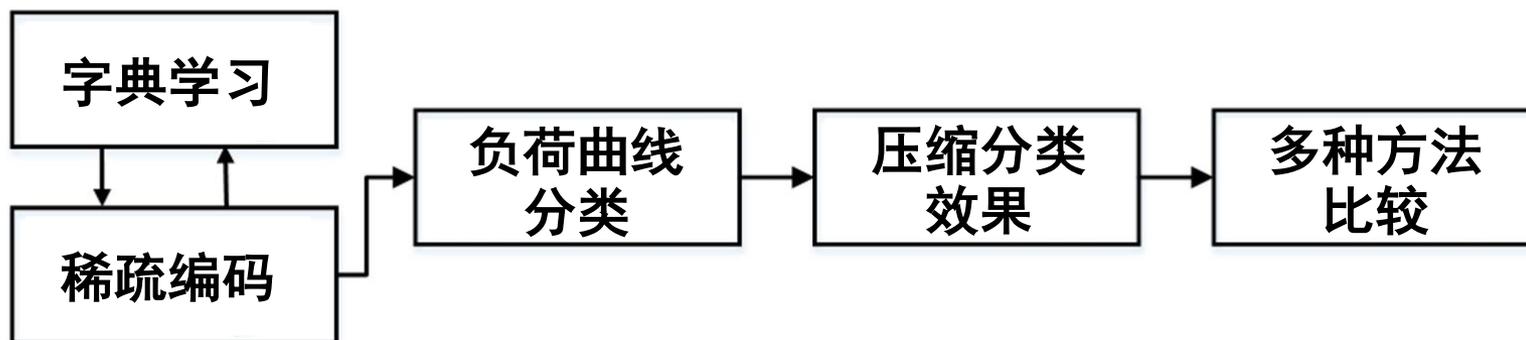
$\min \|X - DA\|_F^2$ ➔ 最小化恢复误差
 $s.t. \|a_i\|_0 \leq s_0, \quad 1 \leq i \leq M$ 稀疏约束
 $a_{i,k} \geq 0, \quad 1 \leq i \leq M, 1 \leq k \leq K$ 非负约束
 $d_{k,n} \geq 0, \quad 1 \leq k \leq K, 1 \leq n \leq N$ 非负约束

5. 基于用户用电特征的行为模式提取

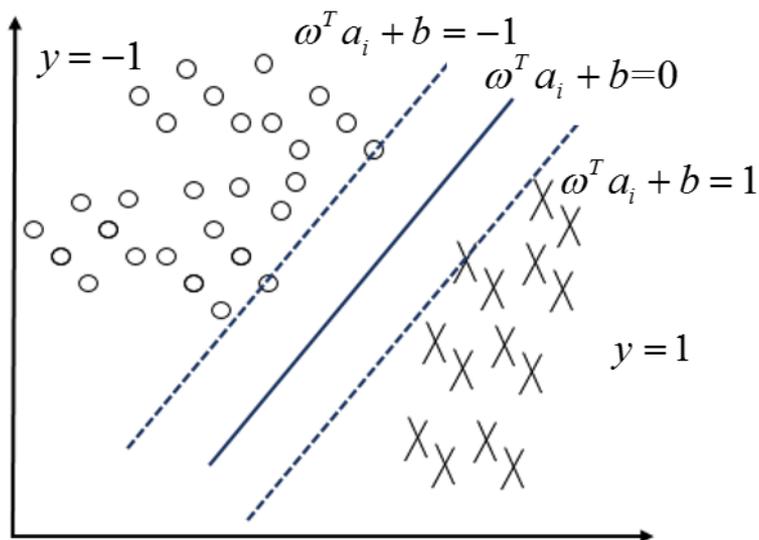


用电曲线恢复效果：仅用3-4个用电子模式就能够很好地重构负荷曲线；
意味着：用户用电曲线本质上是几种简单用电行为的叠加！

5. 基于用户用电特征的行为模式提取



基于线性SVM的用户负荷曲线分类：

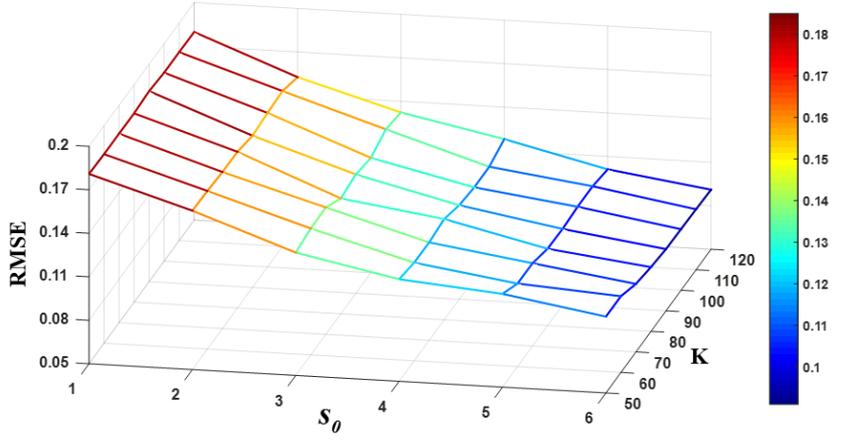


- 1) 计算量小，适合处理大型数据集；
- 2) 只有一个必须确定的参数C；
- 3) 训练样本数目足够大，精度较高；
- 4) 特征的权重具有一定的物理意义。

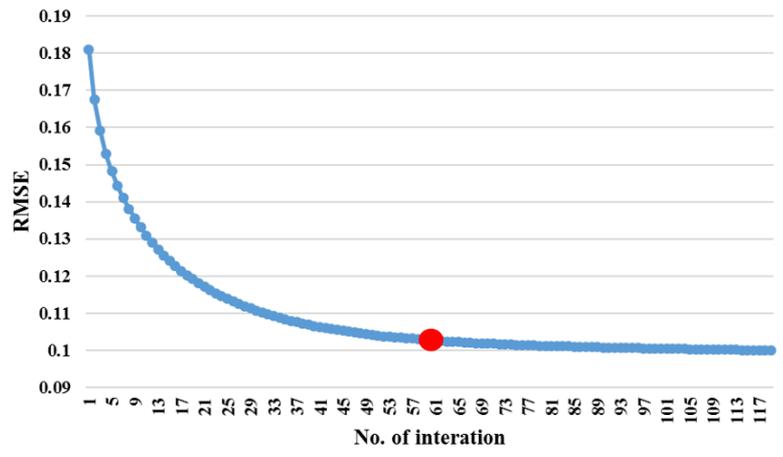
5. 基于用户用电特征的行为模式提取



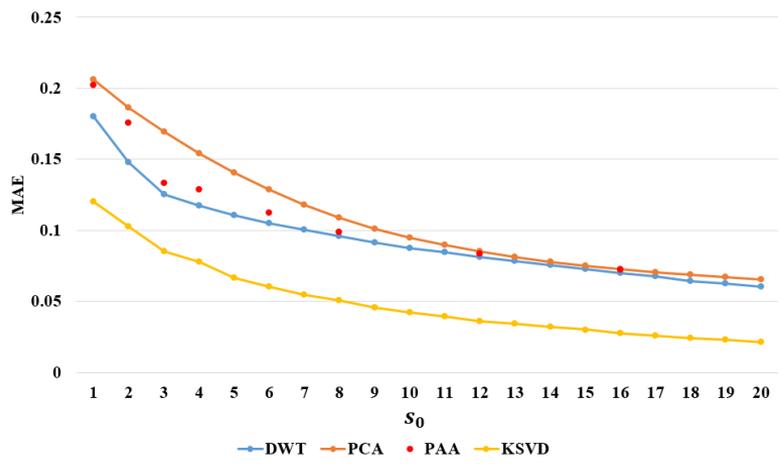
用电曲线分类准确度变化情况



重构误差变化情况

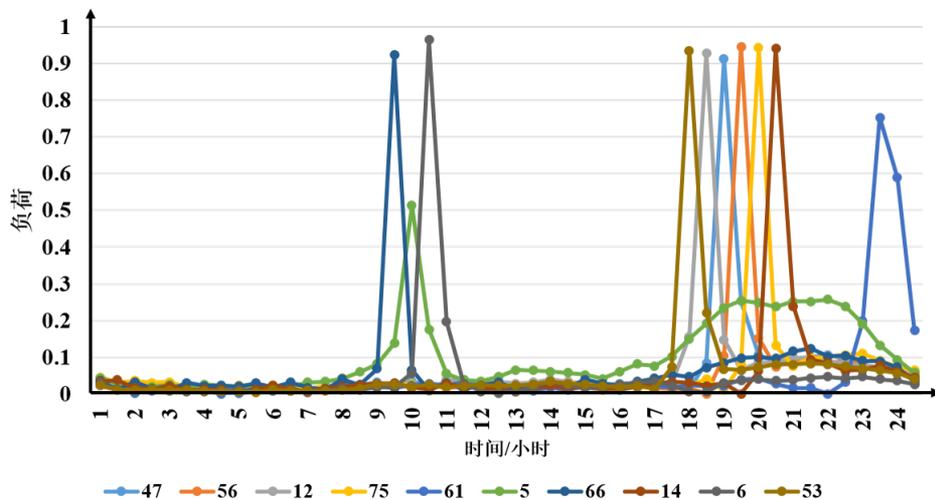
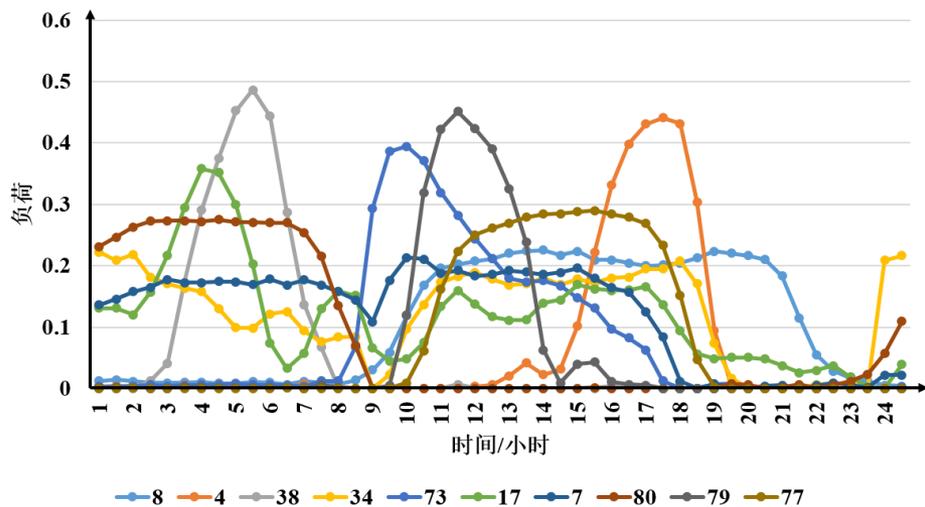


RMSE随迭代次数的变化情况



各方法重构误差变化

5. 基于用户用电特征的行为模式提取



- 工商业用电子模式持续时间较长，主要出现在工作时间和凌晨，可能是由于轮班导致；
- 而居民用电子模式持续时间较短，主要集中在早上和晚上非上班时间，与居民作息规律一致。

5. 基于用户用电特征的行为模式提取

稀疏编码技术能够对用户用电模式进行有效提取，该技术本质上也是一种特征提取或者信号变换技术，**将时域负荷曲线映射到高维特征域。**

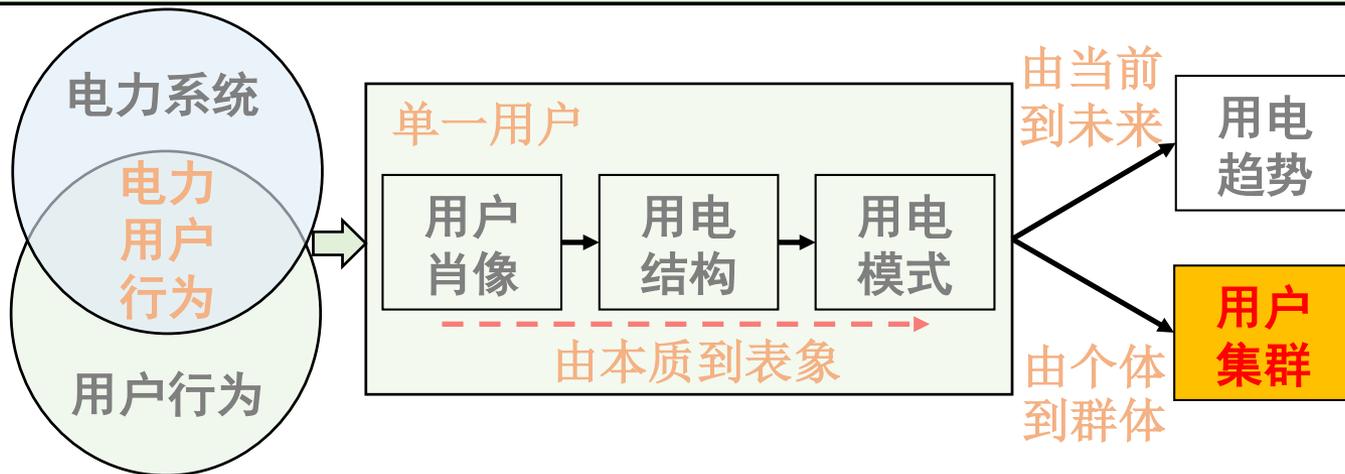
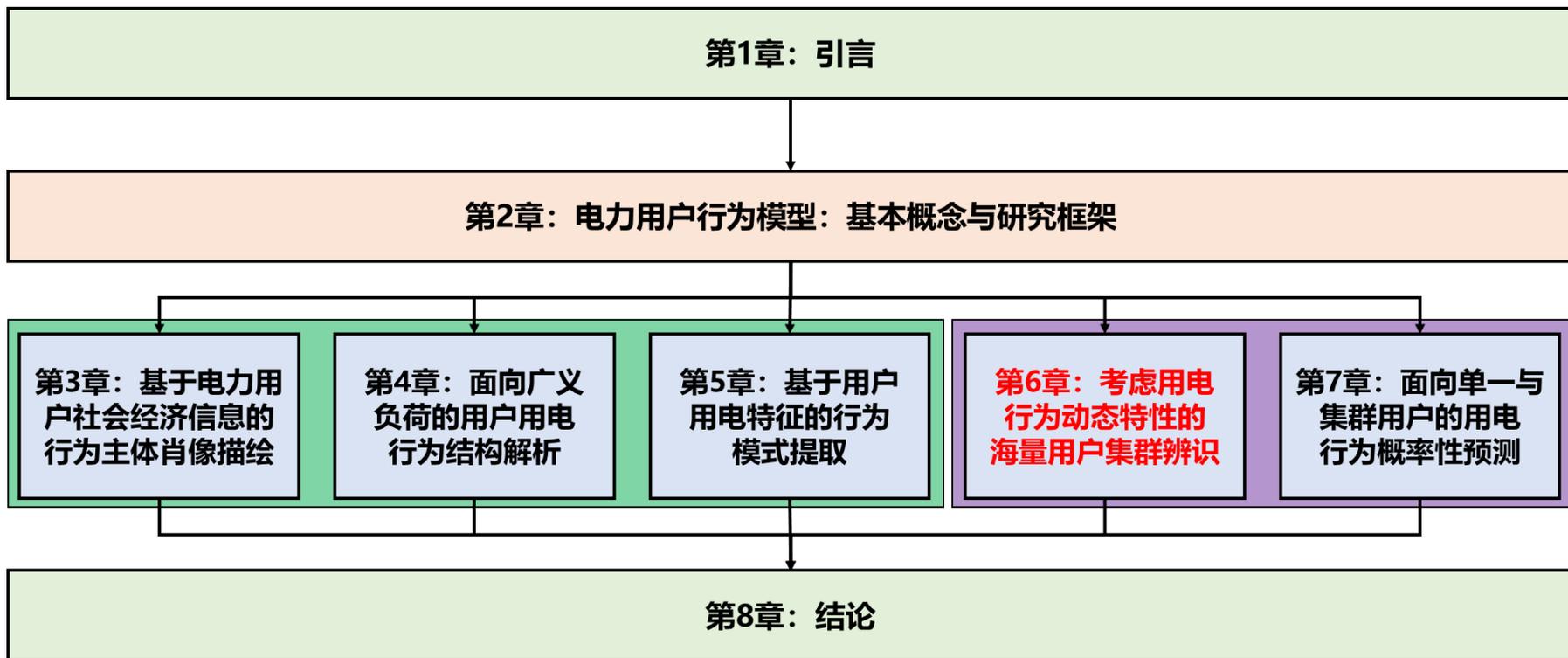
通过与传统聚类、特征提取、信号处理的方法相比，稀疏编码技术对负荷曲线具有良好的恢复效果，**将提取的用电子模式作为分类器输入，能够有效提升用户分类精度。**

	参数	RMSE	MAE	准确度	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
原始曲线	48	/	/	0.735	0.724

$$P_i^t = \sum_{a \in A} P_{i,a}^t \quad \longrightarrow \quad P_{i,a}^t = f(P_{i,t})$$

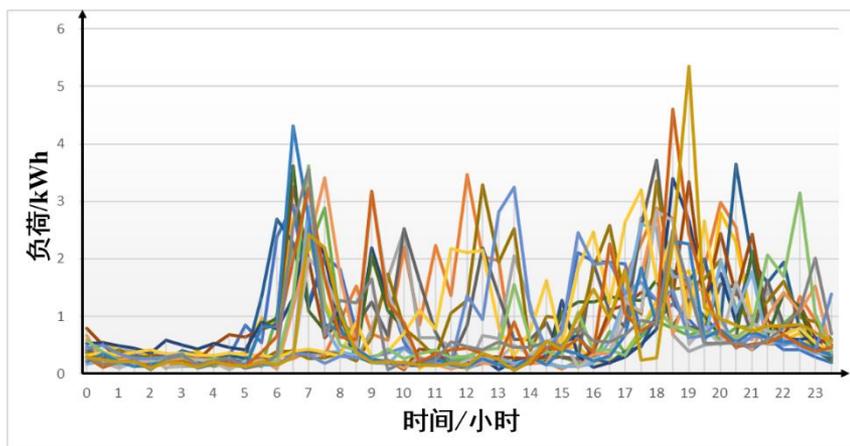
用电子模式 用电曲线

- 提出了用户用电特征的行为模式提取方法和基于稀疏表示的实现技术；
- 能够有效提取用电子模式，为负荷曲线分析提供全新视角；
- 提升数据压缩率和用户分类准确率。

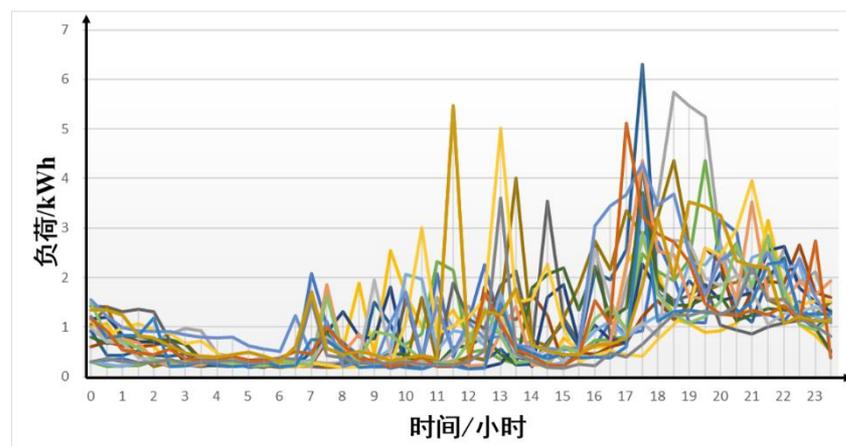


6. 考虑用电行为动态特性的海量用户集群辨识

用户集群辨识是将具有相似特征的用户聚集为一类，在“一视同仁”的粗放管理和“个性化关怀”的过于精细化管理中折衷是一种经济性和可行性较高的策略。



某居民用户#1



某居民用户#2

研究意义：

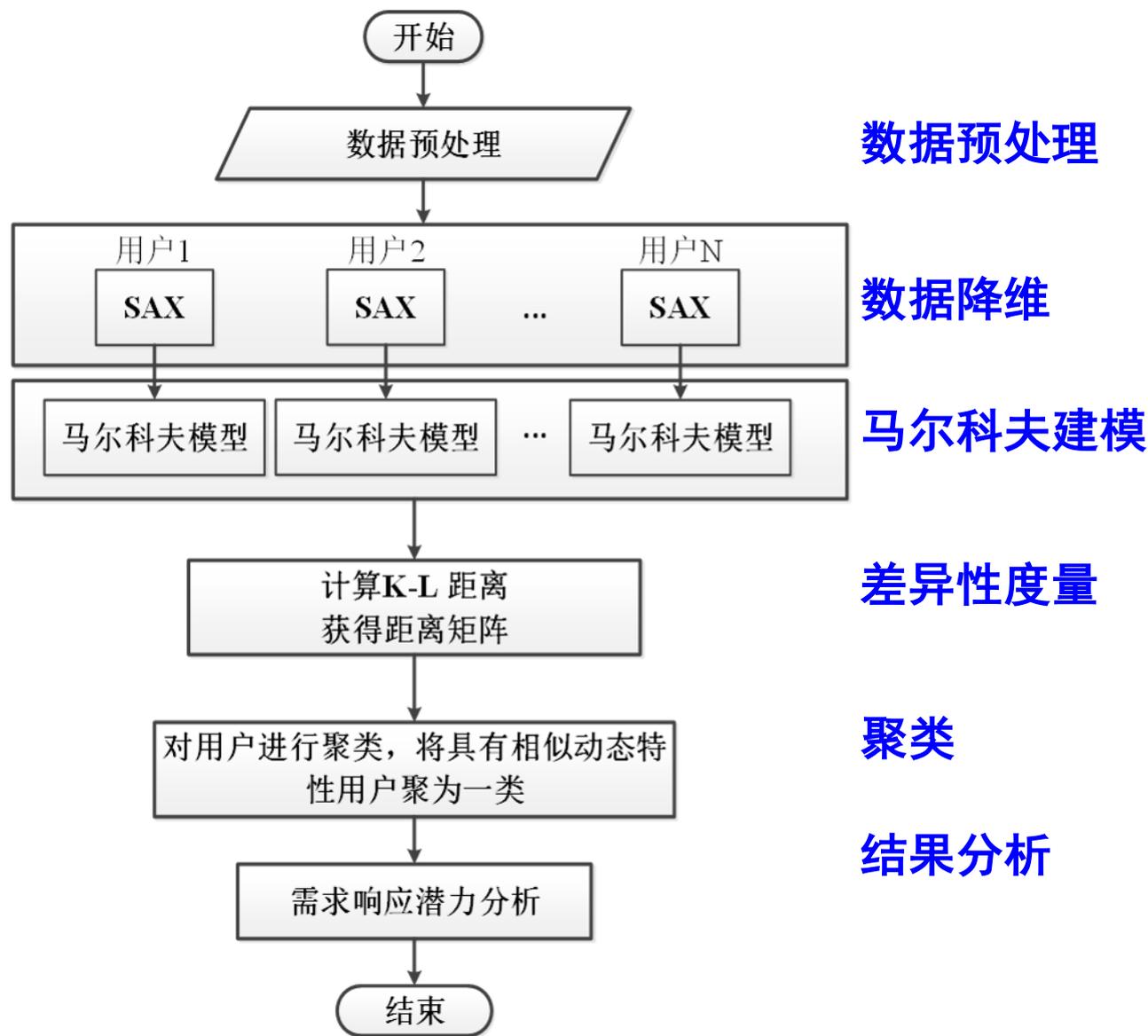
- 分类高效的电力用户管理；
- 需求响应潜力辨识与电价制定；

- ◆ 用户用电特性：1) 波动性与多样性；2) 高维特性；3) 海量用户；
- ◆ 对用户集群的划分应该是面向应用的，不拘泥于曲线形状本身的。

6. 考虑用电行为动态特性的海量用户集群辨识

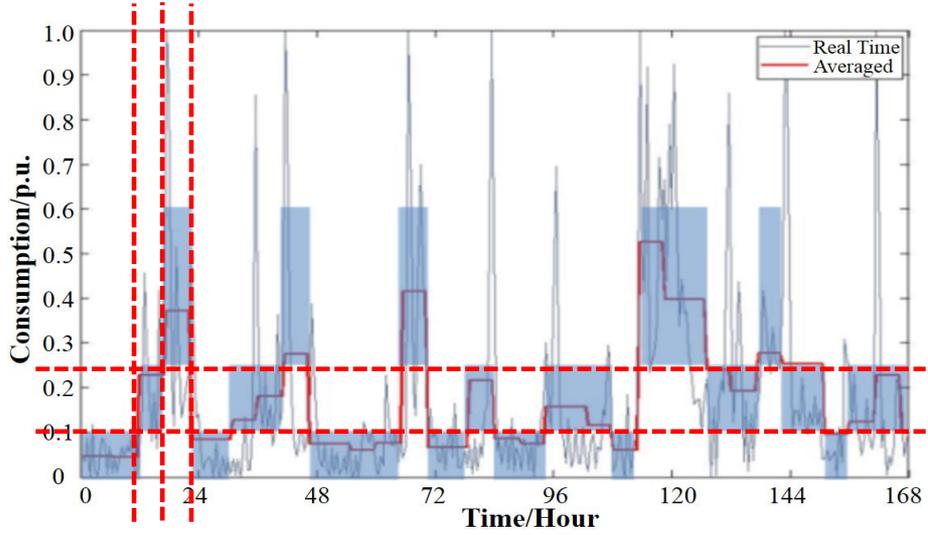
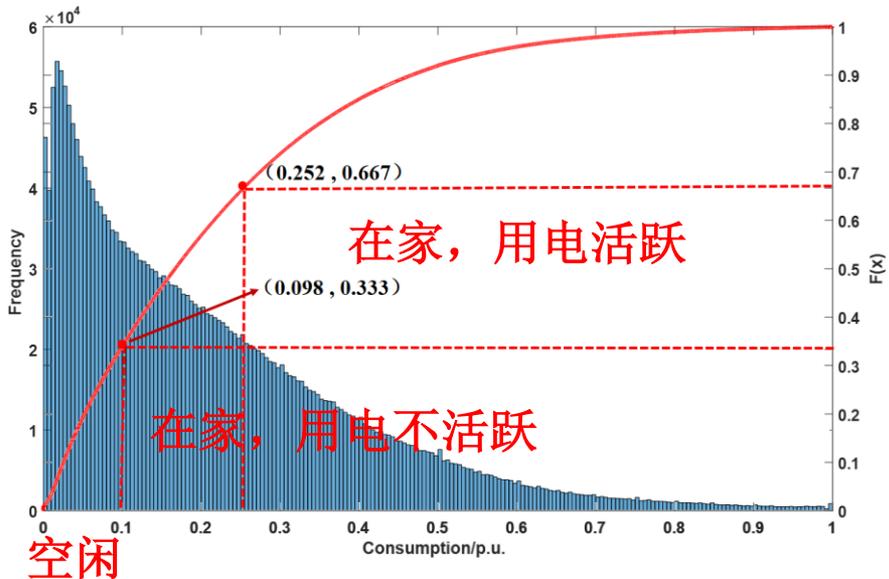
背景：需求响应是为电力系统提供灵活性和消纳可再生能源的重要手段，实施需求响应的第一步是找到合适的用户，能否通过用户用电规律寻找合适的用户参与需求响应？

难点：如何提取有效反应需求响应潜力的指标？如何应对海量数据带来的“大数据挑战”？



6. 考虑用电行为动态特性的海量用户集群辨识

用电曲线符号聚合近似

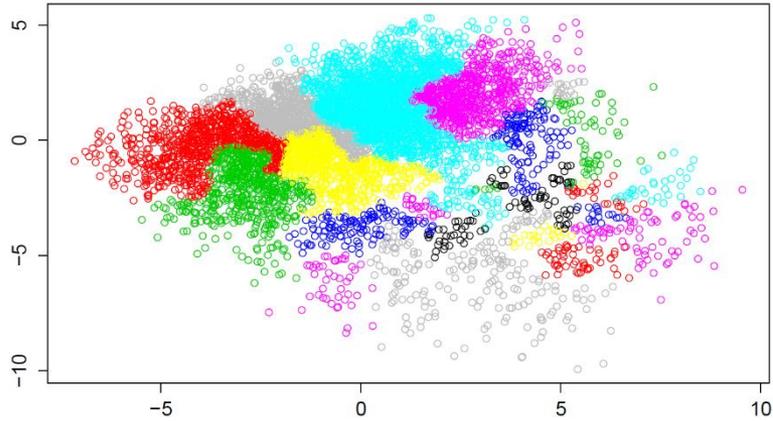
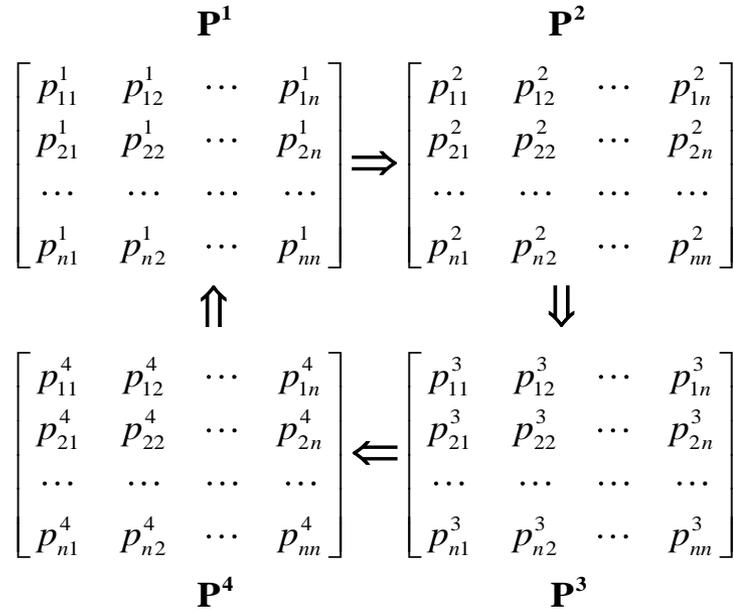


根据用电分布情况，将用户划分为三种用电状态：空闲、在家但用电少，在家且用电活跃

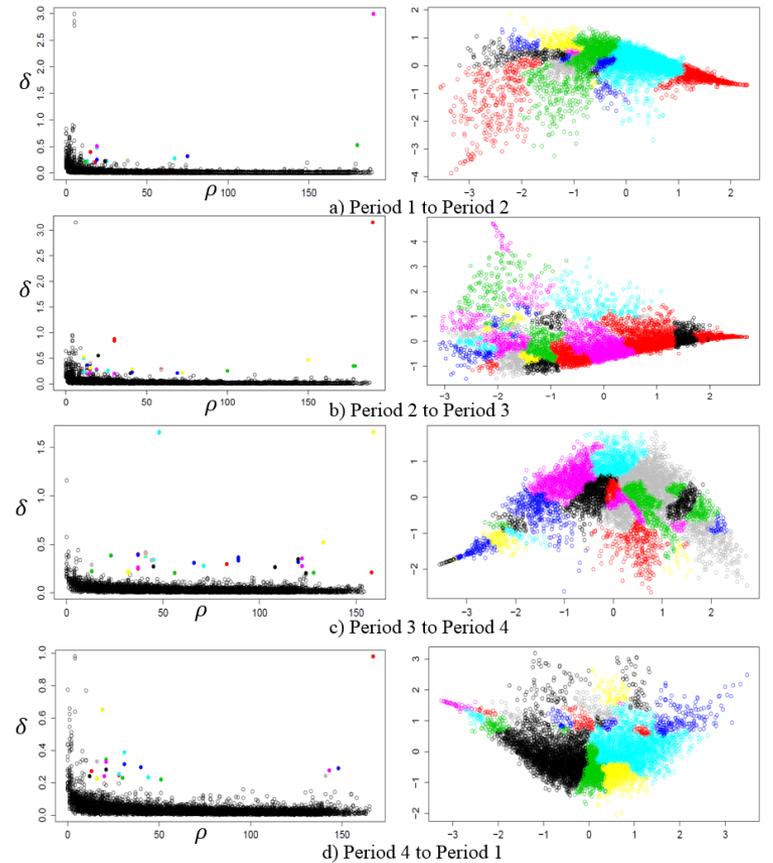
"abcabbcaaacabaabbaccbbbcabb"

6. 考虑用电行为动态特性的海量用户集群辨识

利用马尔科夫模型刻画用户的**动态变化规律**；



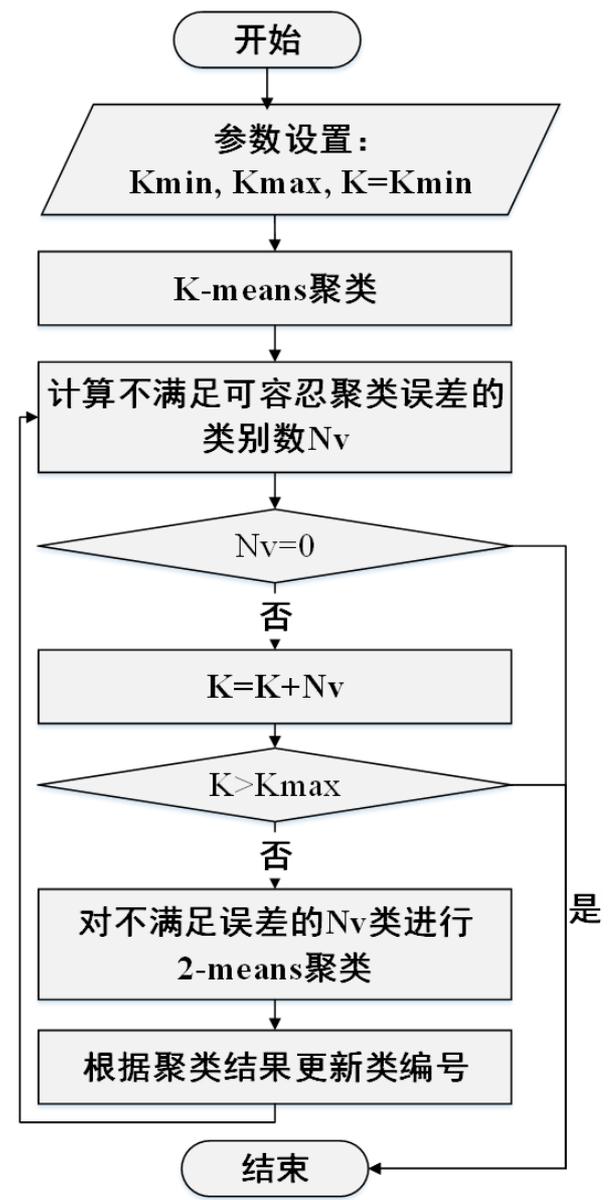
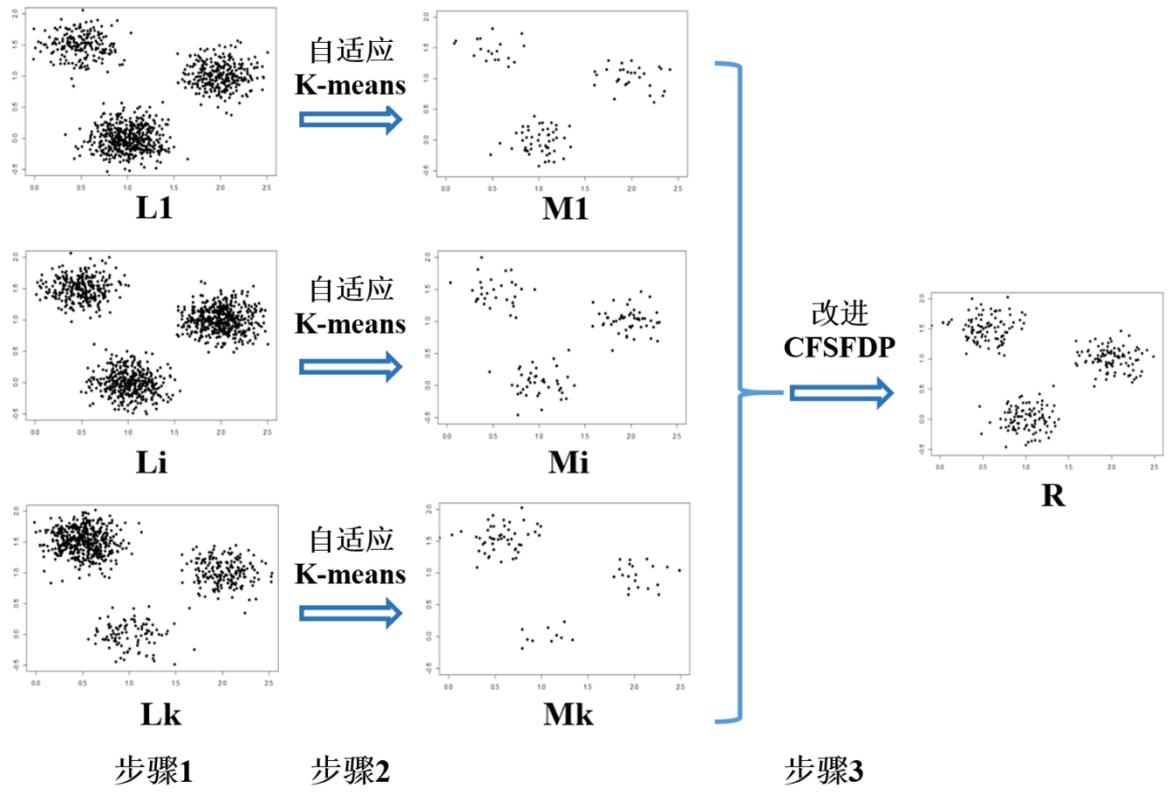
6000多个用户分布情况



效果：能够有效识别不同分布形态的用户，其中变化较多的适合基于电价的DR，灵活性强；变化较少的适合基于激励的DR，便于控制！

6. 考虑用电行为动态特性的海量用户集群辨识

设计分布式聚类框架

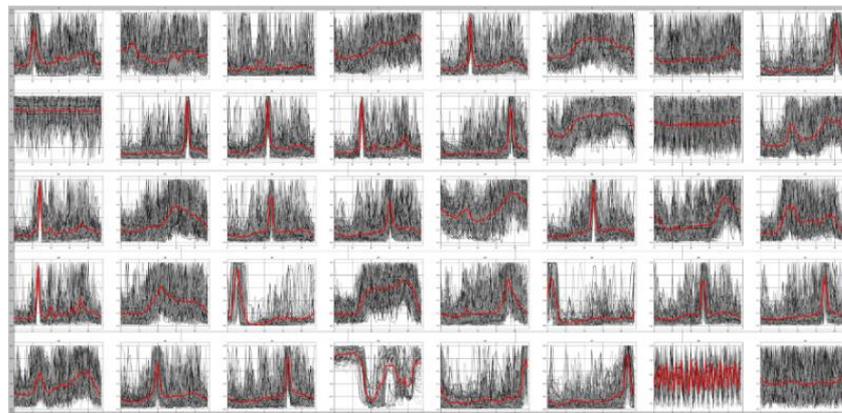
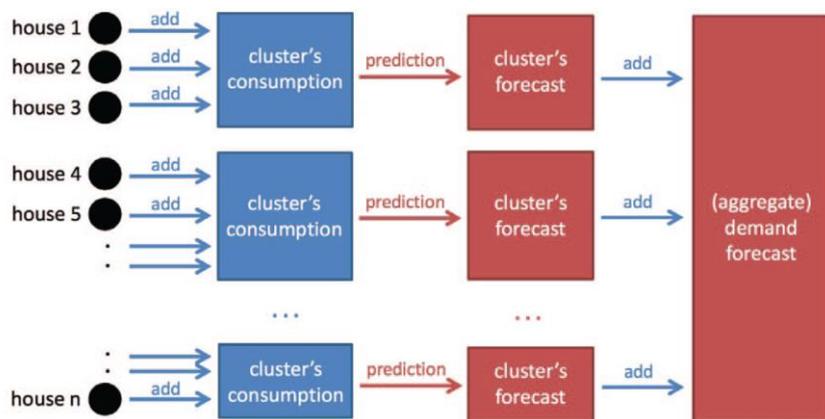


6. 考虑用电行为动态特性的海量用户集群辨识

用户集群划分改善集成负荷预测精度

传统的系统、母线负荷预测主要利用其本身的历史数据和相关因素，而细粒度的用户用电数据提供了更多的“微观信息”。

如何有效组织海量的用户用电数据，为集成负荷预测提供支撑？



相对用户进行**先聚类再分别预测**，最后加和的预测精度有所提升！将用电行为相似的用户“**打包**”分析，更好利用海量用户中的固定用电模式，削弱预测不确定性。

6. 考虑用电行为动态特性的海量用户集群辨识



- 聚类数应该如何选取？
- 不同数据集最优聚类数不一样？

改变聚类数据能够得到不同预测结果，从而可以进一步**开展集成预测**，进一步提高预测精度。

真实负荷 \rightarrow $L_{en,t}$ 第n个预测结果 \rightarrow $\hat{L}_{en,t}$

$$\hat{\omega} = \arg \min_{\omega} \sum_{t=1}^T \frac{1}{T} \frac{|L_{en,t} - \hat{L}_{en,t}|}{L_{en,t}} \quad \rightarrow \text{最小化MAPE}$$
$$s.t. \hat{L}_{en,t} = \sum_{n=1}^N \omega_n \hat{L}_{en,n,t}, \quad \sum_{n=1}^N \omega_n = 1, \quad \omega_n \geq 0.$$

上述模型能够通过引入辅助变量**转换为线性规划（LP）问题**，非常方便求解！

需要确定不同的预测结果的权重

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

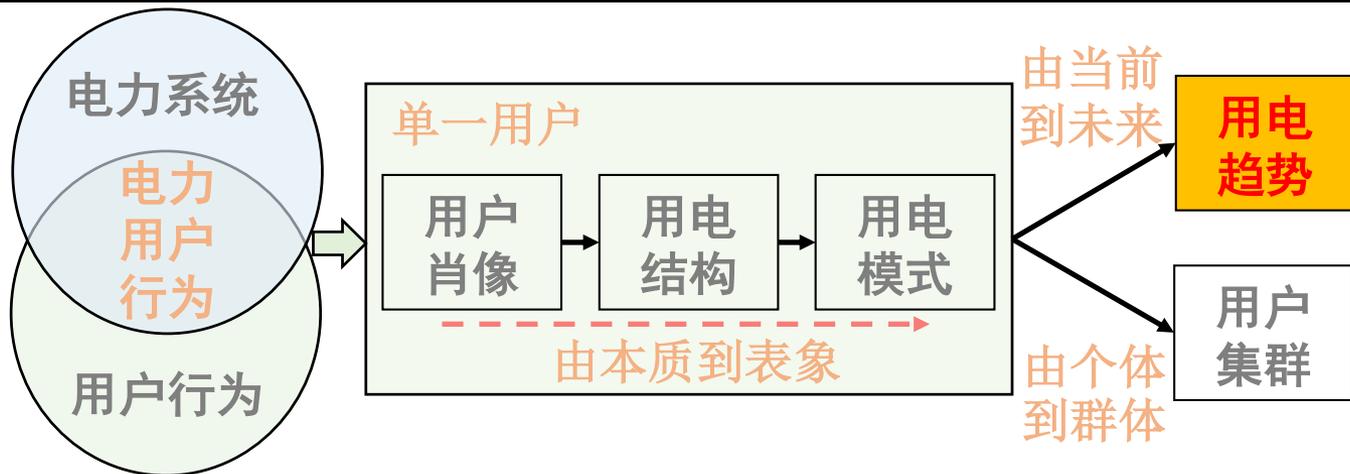
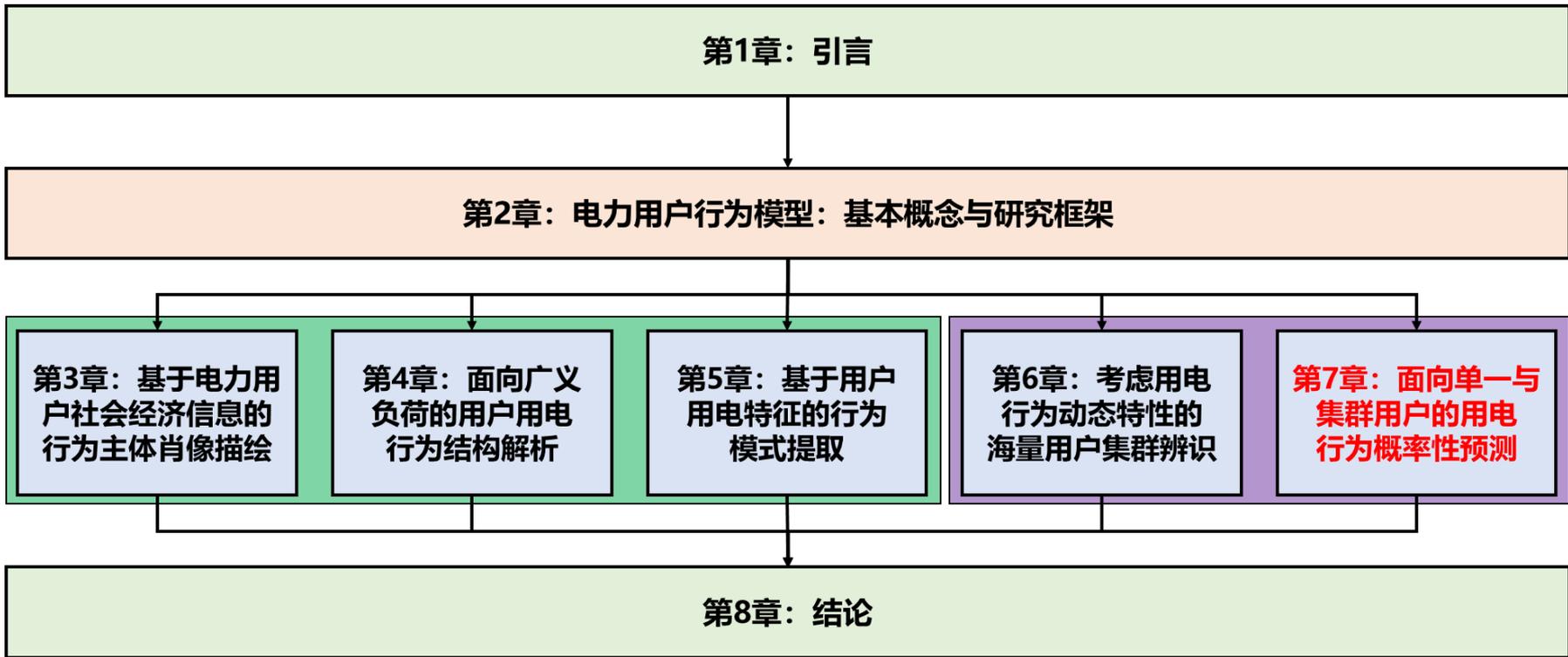
效果：改变聚类数据能够得到不同预测结果的集成能够从单一预测精度的4.25%提升到4.05%。

$$\begin{aligned} & \max_{S_1, S_2, \dots, S_N} \prod_{n=1}^N \prod_{i \in S_n} \text{Prob}(F_i^t | i \in S_n) \\ & \text{s.t. } S_1 \cup S_2 \cup \dots \cup S_N = I \\ & \quad S_1 \cap S_2 \cap \dots \cap S_N = \emptyset \end{aligned}$$

$$\rightarrow S_n = f(P_{i,t}, F_i^t)$$

用户集群 用电曲线 动态特性

- 提出了一种考虑用户动态特性的用户集群划分方法；
- 提出了面向海量用电大数据的分布式聚类方法，有效提升聚类效率；
- 基于用户集群划分和最优综合模型能够改善集成负荷预测精度。



7. 面向单一与集群用户的用电行为概率性预测

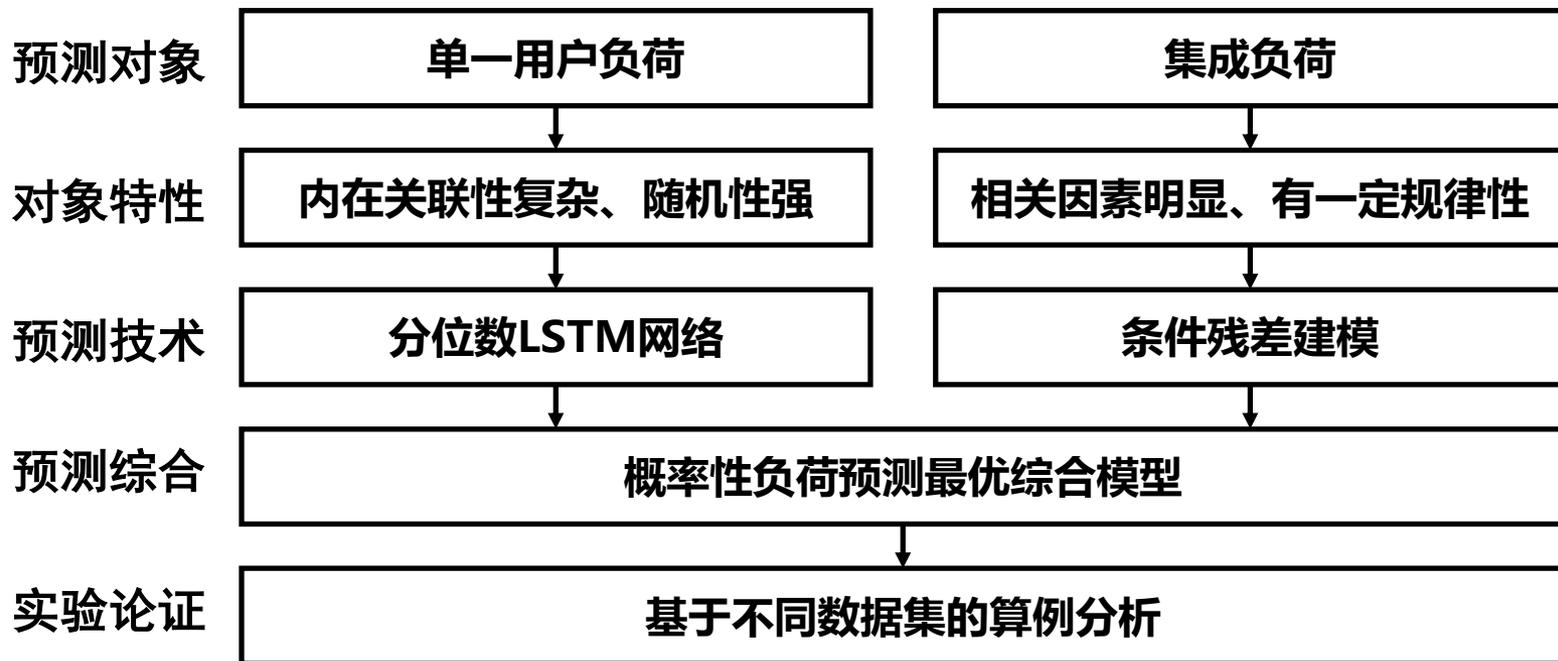


观测不确定性 = 内生不确定性 + 量测不确定性



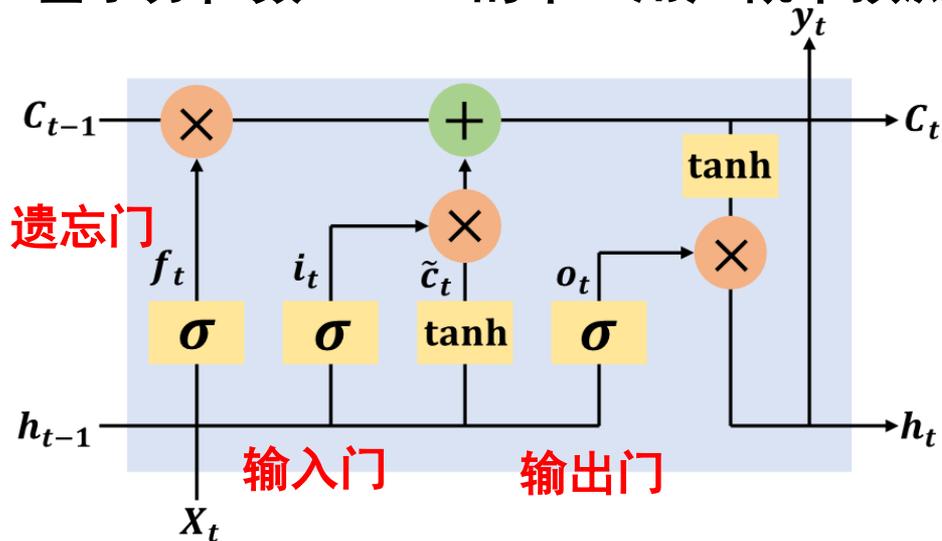
改变用户行为

尽可能减少!



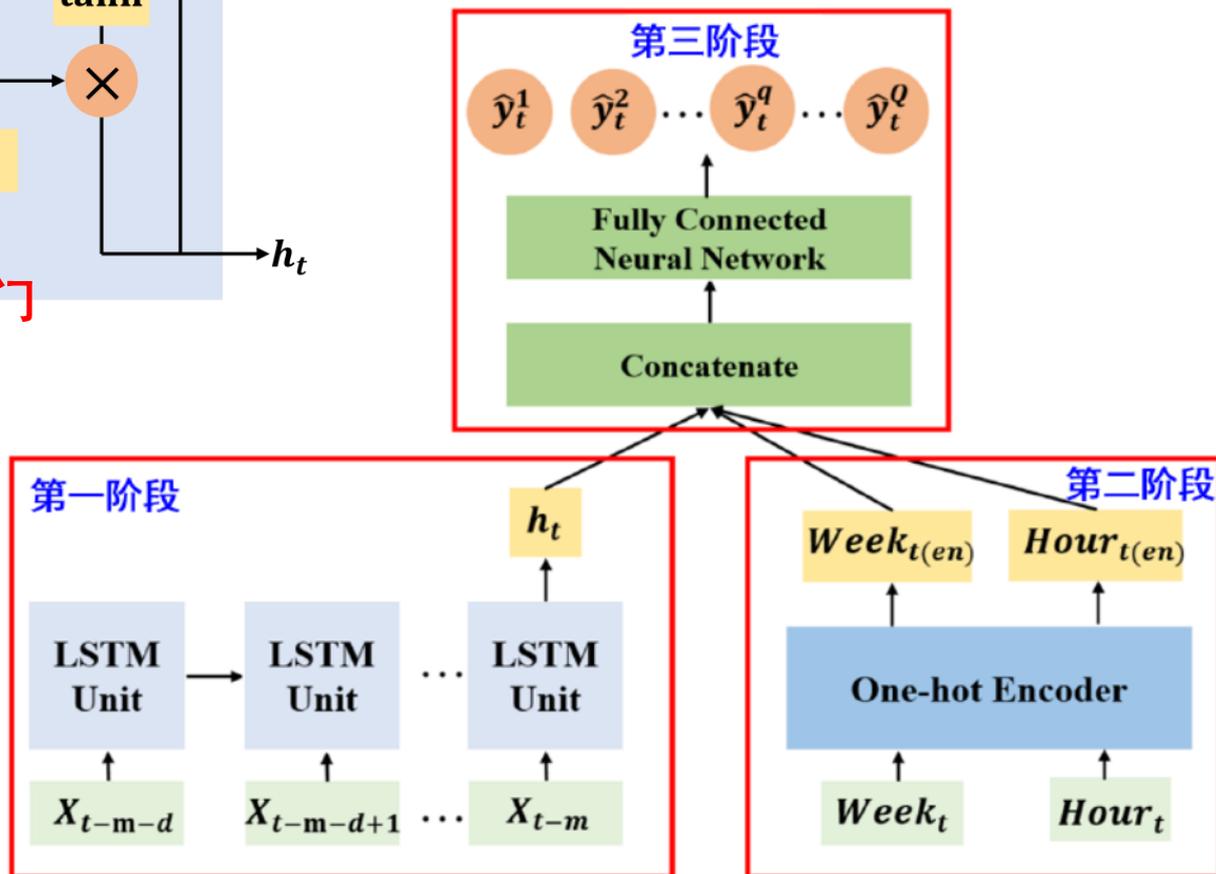
7. 面向单一与集群用户的用电行为概率性预测

基于分位数LSTM 的单一用户概率预测



h 和 c 用于保存短期信息和长期信息

- 1、分位数LSTM 是LSTM 和分位数损失函数的结合；
- 2、利用独热编码器对日类型等进行建模；
- 3、损失函数为所有的分位数损失。



7. 面向单一与集群用户的用电行为概率性预测

基于分位数LSTM 的单一用户概率预测

分位数损失的修正，以便于训练

$$H(y_t, \hat{y}_t^q) = \begin{cases} \frac{(\hat{y}_t^q - y_t)^2}{2\epsilon} & 0 \leq |\hat{y}_t^q - y_t| \leq \epsilon \\ |\hat{y}_t^q - y_t| - \frac{\epsilon}{2} & |\hat{y}_t^q - y_t| > \epsilon, \end{cases} \quad \Rightarrow \quad L_{q,t}(y_t, \hat{y}_t^q) = \begin{cases} (1-q)H(y_t, \hat{y}_t^q) & \hat{y}_t^q \geq y_t \\ qH(y_t, \hat{y}_t^q) & \hat{y}_t^q < y_t. \end{cases}$$

不同方法对100个居民用户预测的平均分位数损失(kW)

	QLSTM	QRNN	I_QRNN	QGBRT	I_QGBRT	LSTM+E	I_LSTM+E
30 分钟	0.0837	0.0867	3.46%	0.0886	5.50%	0.0905	7.52%
1 小时	0.0963	0.0990	2.76%	0.1030	6.48%	0.1020	5.62%
2 小时	0.1018	0.1040	2.18%	0.1077	5.50%	0.1061	4.13%
4 小时	0.1031	0.1054	2.19%	0.1090	5.40%	0.1077	4.27%

不同方法对100个工商业用户预测的平均分位数损失(kW)

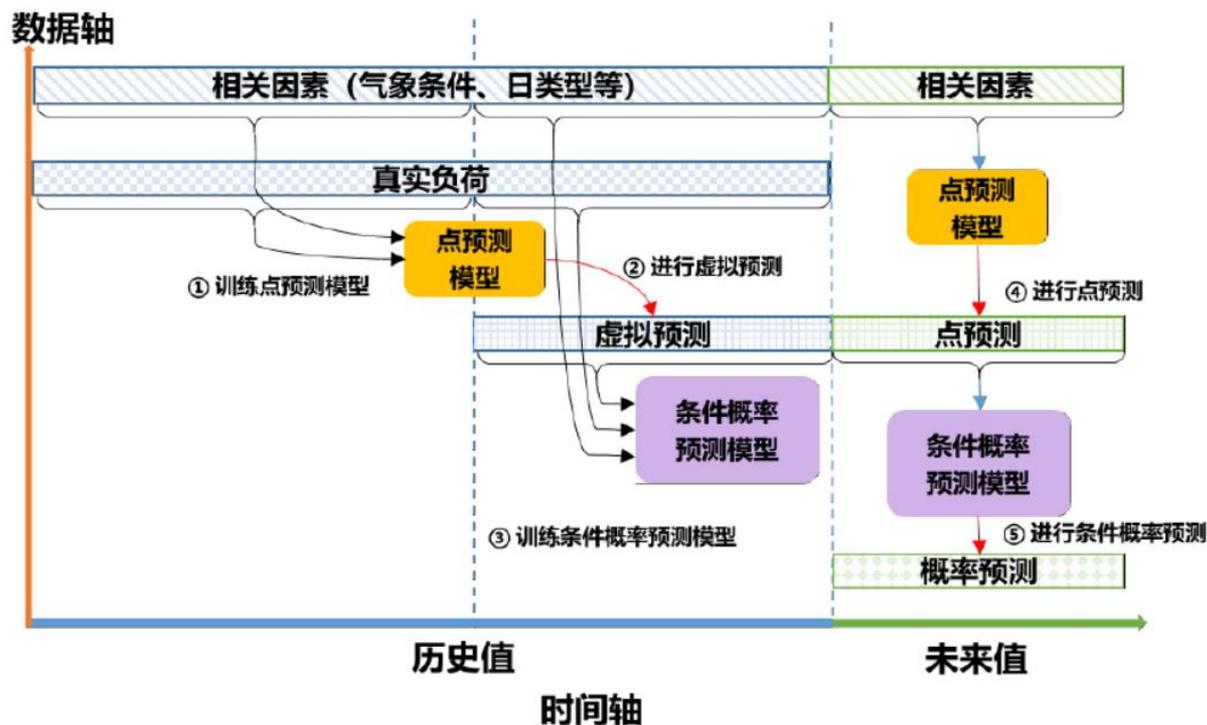
	QLSTM	QRNN	I_QRNN	QGBRT	I_QGBRT	LSTM+E	I_LSTM+E
30 分钟	0.1213	0.1275	4.89%	0.1461	16.98%	0.1391	12.81%
1 小时	0.1552	0.1613	3.79%	0.1975	21.43%	0.1775	12.56%
2 小时	0.1805	0.1883	4.16%	0.2381	24.21%	0.2081	13.29%
4 小时	0.1982	0.2114	6.27%	0.2671	25.80%	0.2252	12.01%

7. 面向单一与集群用户的用电行为概率性预测

基于条件残差建模的集成负荷概率预测

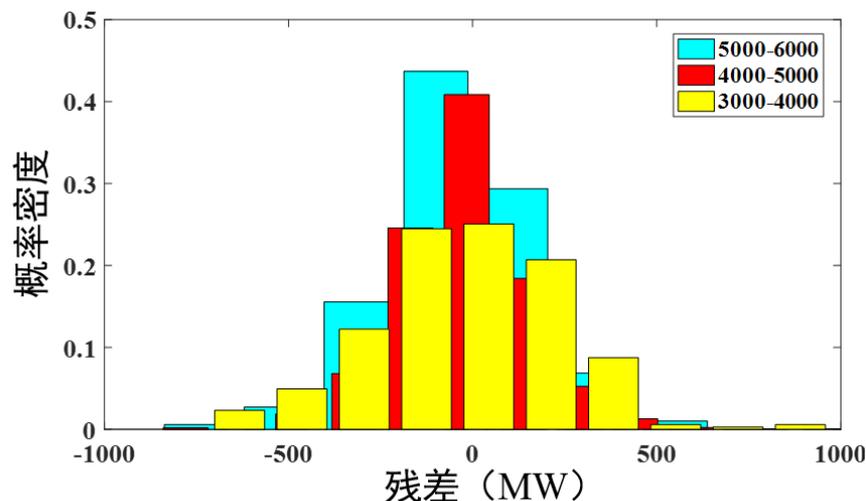
背景：目前已经有很多成熟的点预测方法，概率预测能够更好的利用点预测结果呢？

难点：对点预测的残差进行建模能够得到概率预测结果，但是预测残差不同时段不同，不同负荷水平不同，受到很多条件的影响，不方便直接统计建模。

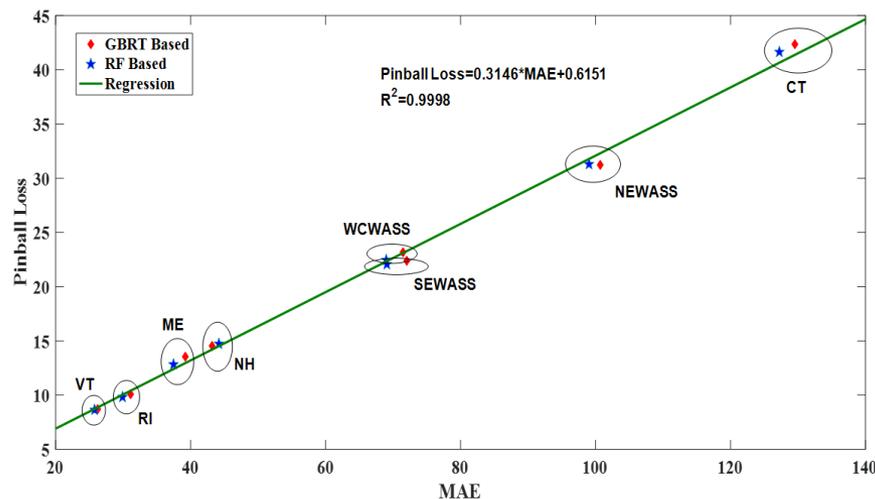


$\epsilon_{t,q} = g_q(\mathbf{W}_q, [\mathbf{X}_t, \hat{y}_t])$ 把点预测结果作为分位数回归模型的输入，残差作为拟合对象，构建残差相对于输入和预测值的条件概率预测模型

7. 面向单一与集群用户的用电行为概率性预测



预测残差分布



概率预测误差和点预测误差之间的线性关系

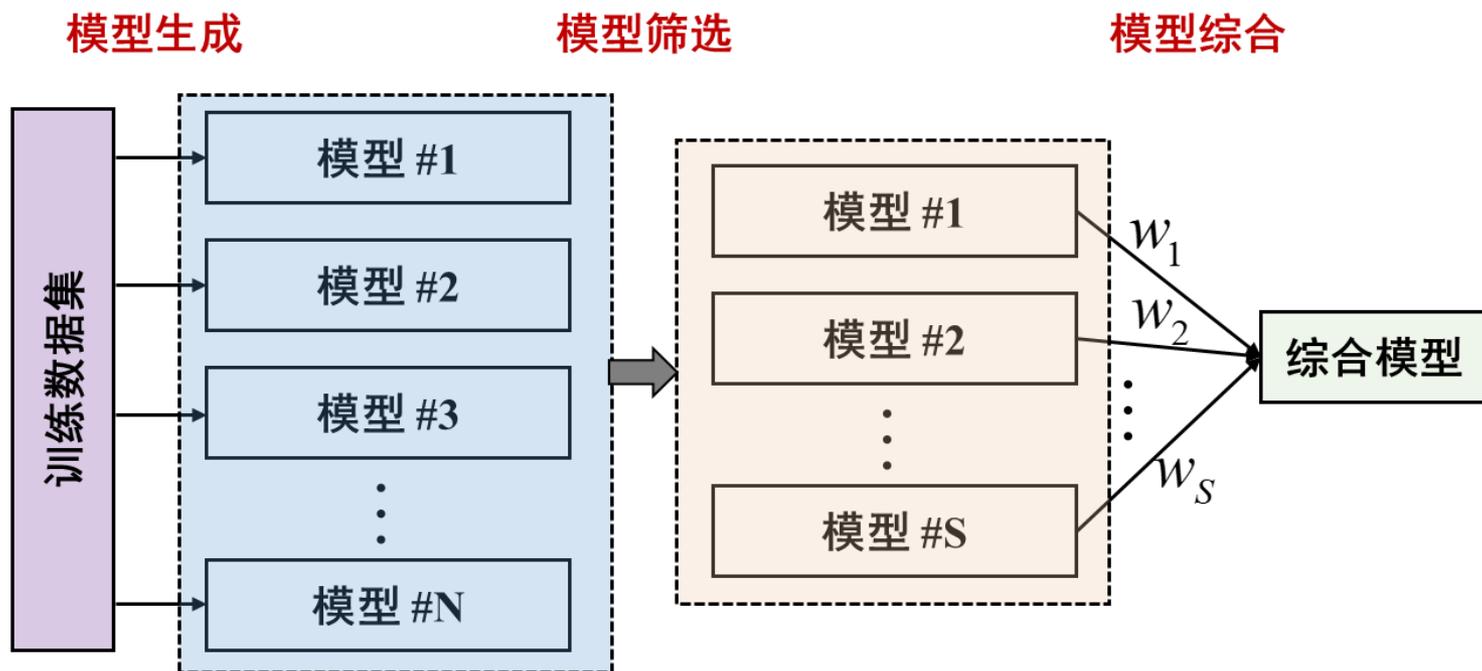
效果：一方面能够有效提高预测精度，相对于直接分位数回归能提高10%左右！
另一方面，可以发现点预测精度很大程度上决定了概率预测结果！

		CT	ME	NH	RI	VT	NE	SE	WC	SYS
直接预测	Pinball	47.02	13.92	16.34	11.15	9.51	35.75	25.40	25.52	164.74
	WS	749.17	213.70	263.15	179.10	150.56	561.60	429.26	405.81	2610.44
条件残差建模	MAPE	3.78	2.95	3.36	3.28	4.21	3.55	4.17	3.68	3.16
	RMSE	171.82	49.06	59.74	40.60	34.81	133.21	94.54	91.85	582.09
	MAE	127.25	37.44	44.22	29.95	25.79	99.01	68.99	69.06	433.44
	Pinball	41.61	12.82	14.69	9.79	8.66	31.27	22.47	22.06	137.85
	WS	666.62	201.03	237.47	155.31	147.04	506.61	370.21	357.49	2261.74
改进 (%)	Pinball	11.50	7.86	10.11	12.23	8.95	12.55	11.52	13.55	16.32
	WS	11.02	5.93	9.76	13.28	2.34	9.79	13.76	11.91	13.36

7. 面向单一与集群用户的用电行为概率性预测

背景： 目前以后较多研究将点预测结果“集成”起来进一步提升点预测精度。那么多个概率预测模型怎么“集成”进一步提升概率预测精度呢？

难点： 点预测集成是一个“一维”问题，而概率预测结果是“高维”的，如何有效且高效整合概率预测结果是个难点。



- 1、生成多种概率预测方法；
- 2、去掉冗余或无用的预测模型；
- 3、实现多种模型的最优综合。

7. 面向单一与集群用户的用电行为概率性预测

从点预测综合到概率预测综合

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

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

点预测

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

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

分位数预测

以分位数损失函数（Pinball Loss）为目标函数，构建优化模型确定不同分位数结果的权重，能够将其装化成为线性规划问题！

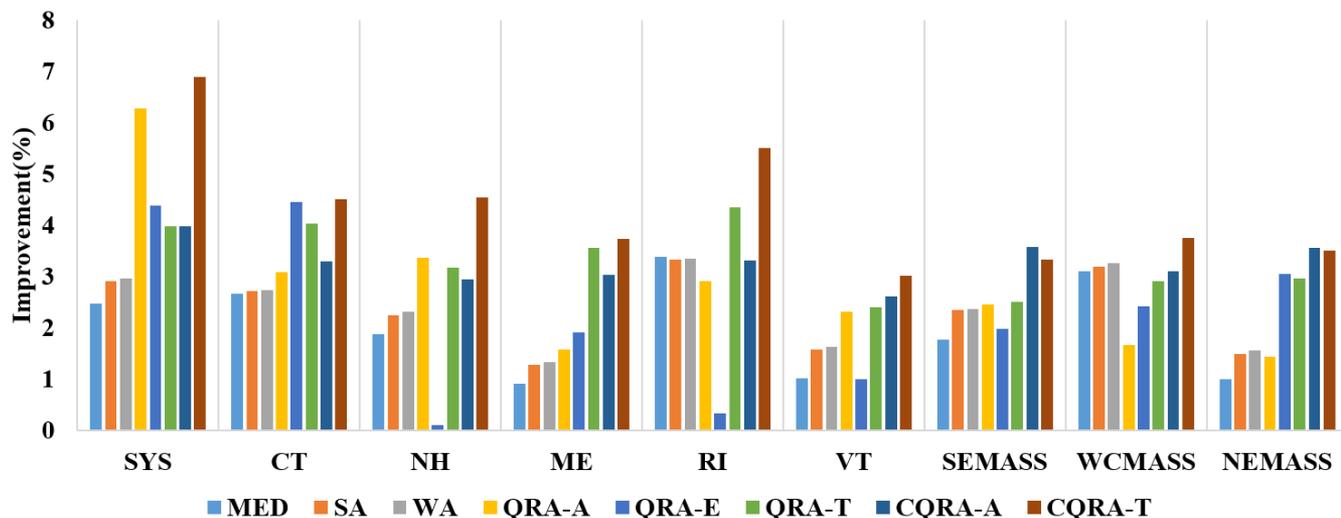
$$\begin{aligned} \hat{\omega}_q &= \arg \min_{\omega_q} \sum_{t \in T} L_{t,q}(\hat{y}_{t,q}, y_t) \\ &= \arg \min_{\omega_q} \sum_{t \in T} \sum_{q \in Q} \max \{ q(y_t - \hat{y}_{t,q}), (1-q)(\hat{y}_{t,q} - y_t) \} \\ \text{s.t.} \quad \hat{y}_{t,q} &= \sum_{n \in N} \omega_{n,q} \hat{y}_{n,t,q}, \quad \sum_{n \in N} \omega_{n,q} = 1, \quad \omega_n \geq 0 \quad \forall n. \end{aligned} \quad \rightarrow \quad \begin{aligned} \hat{\omega}_q &= \arg \min_{\omega_q} \sum_{t \in T} v_{t,q} \\ \text{s.t.} \quad \hat{y}_{t,q} &= \sum_{n \in N} \omega_{n,q} \hat{y}_{n,t,q}, \quad \sum_{n \in N} \omega_{n,q} = 1, \quad \omega_{n,q} \geq 0 \quad \forall n. \\ v_{t,q} &\geq q(y_t - \hat{y}_{t,q}), \quad v_{t,q} \geq (1-q)(\hat{y}_{t,q} - y_t) \\ \{v_{t,q} - q(y_t - \hat{y}_{t,q})\} \{v_{t,q} - (1-q)(\hat{y}_{t,q} - y_t)\} &= 0. \end{aligned}$$

模型线性化、模型选择

7. 面向单一与集群用户的用电行为概率性预测

不同综合预测方法在不同地区的预测效果

方法 \ 区域	SYS	CT	NH	ME	RI	VT	SE	WC	NE
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



7. 面向单一与集群用户的用电行为概率性预测

对于总负荷（SYS）预测不同模型在不同分位数下的权重

Quantiles Models	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

对于不同区域负荷预测不同模型在90分位数下的权重

Zones Models	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

$$\begin{aligned}
 \hat{P}_{i,a}^{t+h} &= f_{i,a}(\mathbf{C}_i^t, \mathbf{E}_i^t, \hat{\mathbf{E}}_i^{t+h}, P_{i,a}^t, t) \\
 \hat{P}_i^{t+h} &= f_i(\mathbf{C}_i^t, \mathbf{E}_i^t, \hat{\mathbf{E}}_i^{t+h}, P_i^t, t)
 \end{aligned}
 \quad \rightarrow \quad
 P_i^{t+h} = f(P_{i,t}, \mathbf{E}_i)$$

未来用电
历史用电
气象条件

- 提出了面向单一用户分位数LSTM概率预测模型，在居民用户和中小型商业用户都具有良好的效果；
- 提出了面向集成负荷概率预测的条件残差建模方法，较直接预测方法有较大的性能提高；
- 提出了概率性负荷预测的最优综合模型，以分位数损失作为核心，将最优综合模型转化成为易于求解的线性优化问题，能够得以高效求解。

1

论文研究背景

2

论文研究内容



3

论文研究结论

从**社会学视角**对电力用户行为进行分析，利用**数据驱动方法**开展模型构建，以促进**电力用户与系统的互动**，提出了“物理-信息-社会”耦合的研究范式。

- 提出了社会学视角的**电力用户行为模型**与研究框架；
- 以**用户属性**为核心，提出了面向**行为主体**的电力用户肖像描绘方法，搭建用户社会经济信息与用电曲线的桥梁；
- 以**用电结构**为核心，提出了面向**行为手段**的广义负荷结构解析方法，实现不可见分布式光伏的参数估计和出力重构；
- 以**用电曲线**为核心，提出了面向**行为结果**的用户用电模式提取方法，揭示多样复杂且弱规律的用电曲线基本特征；
- 以**用户集群**为核心，提出了面向**集群行为**的海量用户集群辨识方法，描绘海量用户集群分布和需求响应潜力评估；
- 以**用电趋势**为核心，提出了面向**预见行为**的用户负荷概率预测方法，刻画单一与集成用户的用电行为不确定性。

◆ 各章节对应发表论文

第1章

[1]. **Yi Wang**, Qixin Chen, Tao Hong, and Chongqing Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges” , *IEEE Transactions on Smart Grid*, in press.

第2章

[2]. Chongqing Kang, **Yi Wang**, Yusheng Xue, Gang Mu, and Ruijin Liao, “Big Data Analytics in China's Electric Power Industry: modern Information, Communication Technologies, and Millions of Smart Meters” , *IEEE Power and Energy Magazine*, 2018, 16(3): 54-65.

[3]. **Yi Wang**, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, and Yun Zhao, “Load Profiling and Its Application to Demand Response: A Review” , *Tsinghua Science and Technology*, 2015, 20(2): 117-129.

第3章

[4]. **Yi Wang**, Qixin Chen, Dahua Gan, Jingwei Yang, Daniel Kirschen, and Chongqing Kang, “Deep Learning-Based Socio-demographic Information Identification from Smart Meter Data” , *IEEE Transactions on Smart Grid*, in press.

第4章

[5]. **Yi Wang**, 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 Transactions on Power Systems*, 2018, 33(3): 3255-3264.

[6]. **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 Transactions on Power Systems*, 2018, 33(3): 3255-3264. (**ESI高被引论文**)

◆ 各章节对应发表论文

第5章

- [7]. **Yi Wang**, Qixin Chen, Chongqing Kang, and Qing Xia, “Sparse and Redundant Representations-Based Smart Meter Data Compression and Pattern Extraction”, *IEEE Transactions on Power Systems*, 2017, 32(3): 2142-2151.
- [8]. **Yi Wang**, Qixin Chen, Chongqing Kang, Qing Xia, Yuekai Tan, Zhijian Zeng, Min Luo, “Residential Smart Meter Data Compression and Patterns Extraction via Non-negative K-SVD”, *IEEE Power & Energy Society General Meeting*, July, 2016.

第6章

- [9]. **Yi Wang**, Qixin Chen, Chongqing Kang, and Qing Xia, “Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications”, *IEEE Transactions on Smart Grid*, 2016, 7(5): 2437-2447.
- [10]. **Yi Wang**, Qixin Chen, Mingyang Sun, and Chongqing Kang and Qing Xia, “An Ensemble Forecasting Method for the Aggregated Load with Subprofiles”, *IEEE Transactions on Smart Grid*, 2018, 9(4): 3906-3908.

第7章

- [11]. **Yi Wang**, Qixin Chen, Ning Zhang, and Yishen Wang, “Conditional Residual Modeling for Probabilistic Load Forecasting”, *IEEE Transactions on Power Systems*, in press.
- [12]. **Yi Wang**, Ning Zhang, Yushi Tan, Tao Hong, Daniel Kirschen, and Chongqing Kang, “Combining Probabilistic Load Forecasts”, *IEEE Transactions on Smart Grid*, in press.
- [13]. **Yi Wang**, Dahua Gan, Mingyang Sun, Ning Zhang, and Chongqing Kang, “Probabilistic Individual Load Forecasting Using Pinball Loss Guided LSTM”, *Applied Energy*, 2019, 235: 10-20.

◆ 其他发表论文

- [14]. **Yi Wang**, Ning Zhang, Qixin Chen, Jingwei Yang, Chongqing Kang, Junhui Huang, “Dependent Discrete Convolution Based Probabilistic Load Flow for Active Distribution System”, *IEEE Transactions on Sustainable Energy*, 2017, 8(3): 1000-1009.
- [15]. **Yi Wang**, Ning Zhang, Chongqing Kang, and Qing Xia, “Standardized Matrix Modeling of Multiple Energy Systems”, *IEEE Transactions on Smart Grid*, in press.
- [16]. **Yi Wang**, Jiangnan Chen, Ning Zhang, and Chongqing Kang, “Automatic and Linearized Modeling of Energy Hub and Its Flexibility Analysis”, *Applied Energy*, 2018, 211: 705-714.
- [17]. **Yi Wang**, Ning Zhang, Zhenyu Zhuo, Chongqing Kang, and Daniel Kirschen, “Mixed-Integer Linear Programming Based Optimal Configuration Planning for Energy Hub: Starting from Scratch”, *Applied Energy*, 2018, 210: 1141-1150. (**ESI高被引论文**)
- [18]. **Yi Wang**, Ning Zhang, Hai Li, Jingwei Yang, and Chongqing Kang, “Linear Three-Phase Power Flow for Unbalanced Active Distribution Networks with PV Nodes”, *CSEE Journal of Power and Energy Systems*, 2017, 3(2): 321-324.
- [19]. **Yi Wang**, Ning Zhang, Chongqing Kang, Qian Yao Xu, Hui Li, Jinyu Xiao, Zhidong Wang, Rui Shi, and Shuai Wang, “Ordinal Optimization Theory Based Planning for Clustered Wind Farms Considering the Capacity Credit”, *Journal of Electrical Engineering & Technology*, 2015, 10(5): 1903-1939.
- [20]. **王毅**, 张宁, 康重庆. 能源互联网中能量枢纽的优化规划与运行研究综述及展望. *中国电机工程学报*, 2015, 35(22):5669-5681. (**2018年中国电机工程学会优秀期刊论文二等奖**)
- [21]. **王毅**, 李鼎睿, 康重庆. 低秩矩阵分解在母线坏数据辨识与修复中的应用. *电网技术*, 2017, 41(06): 1972-1979.
- [22]. **Yi Wang**, Qixin Chen, Chongqing Kang, Jianqiang Miao, Jianhua Bai, Miao Miao, “Linear Optimization for Active Distribution Systems Operation Considering Demand Response Mismatch”, *CIGRE Session*, August, 2016.



清华大学

EI Lab
Energy Intelligence Laboratory
智慧能源实验室

清华大学博士学位论文答辩

请各位老师批评指正

专 业：电气工程

博士生：王 毅

导 师：康重庆 教授

2018年12月20日