

基于扩展面板大数据的电力经济特征提取新方法

张秋雁¹, 宋强¹, 张俊玮¹, 张亚茹², 赵鹏程³, 王波⁴, 马恒瑞⁵

(1. 贵州电网有限责任公司 电力科学研究院, 贵州 贵阳 550002; 2. 国网冀北电力有限公司 秦皇岛供电公司, 河北 秦皇岛 066000; 3. 贵州电网有限责任公司 都匀供电局, 贵州 都匀 558000; 4. 武汉大学 电气与自动化学院, 湖北 武汉 430072; 5. 青海大学 启迪新能源学院, 青海 西宁 810016)

A Novel Feature Extraction Method in Power Economic Assessment Research Based on Extraction Panel Data

ZHANG Qiuyan¹, SONG Qiang¹, ZHANG Junwei¹, ZHANG Yaru², ZHAO Pengcheng³,
WANG Bo⁴, MA Hengrui⁵

(1. Electric Power Research Institute of Guizhou Power Grid Co., Ltd., Guiyang 550002, Guizhou, China; 2. Qinghuangdao Electric Power Supply Company of State Grid Jibei Electric Power Co., Ltd., Qinghuangdao 066000, Hebei, China; 3. Duyun Power Supply Bureau, Guizhou Power Grid Co., Ltd., Duyun 558000, Guizhou, China; 4. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, Hubei, China; 5. Tus-Institute for Renewable Energy, Qinghai University, Xining 810016, Qinghai, China)

ABSTRACT: As electricity serves as an economic “oscilloscope”, feature extraction and intelligent parameter estimation of massive electricity consumption big data are the key steps of power economy evaluation. In this paper, a modeling method and economic related feature extraction method suitable for massive power economic big data are proposed. First, the extended panel data model is constructed according to the spatiotemporal characteristics of the binary big data of electric power economy, and the stationarity and cointegration are tested. Second, taking the power consumption as the dependent variable, the weight factors of other power economic characteristics are determined by constructing regression equation. Finally, the grey relational clustering is used to extract the features, and the weight factor is used as the criterion to select the clustering center, so as to obtain the optimal feature subset. The simulation and comparative analysis of the actual power consumption data in a province shows that the proposed method can greatly eliminate redundancy on the premise of preserving the physical meaning of feature subset, meet the

needs of economic evaluation, and have a certain generality.

KEY WORDS: feature extraction; power economy evaluation; extraction panel data; grey relation; grey clustering

摘要: 电力作为经济的“示波器”, 对海量电力用电大数据进行特征提取和智能参数估计是电力经济评估的关键步骤。提出了一种适用于海量电力经济大数据的建模方法和经济相关特征提取方法。首先针对电力经济二元大数据的时空特征构造扩展面板数据模型, 并进行平稳性和协整性检验; 然后以用电量为因变量, 通过构造回归方程确定与其他电力经济特征量的权重因子; 最后采用灰色关联聚类进行特征提取, 并以权重因子为判据进行聚类中心选择, 从而获取最优特征子集。通过对某省实际用电数据的仿真对比分析, 验证所提方法能够在保存特征子集物理意义的前提下, 极大消除冗余, 满足了经济评估的需要, 并具有一定的通用性。

关键词: 特征提取; 电力经济评估; 扩展面板数据; 灰色关联; 灰色聚类

在过去的 30 多年时间里, 中国经历了高速增长, 自改革开放以来, 实现了年均 9.7% 的经济增长速度^[1], 成为世界第二大经济体, 经济总量仅次于美国。电力作为重要能源支柱, 直接关系到社会经济健康发展, 同时也是经济运行的先行指标^[2-6]。地区区域经济发展水平与其电力消费水平具有正比例关系, 所以电力消费指标可以用来体现经济增长情况,

基金项目: 国家自然科学基金项目(51777142&51907096); 青海省自然科学基金项目(2019-ZJ-950Q)。

Project Supported by the National Natural Science Foundation of China (51777142&51907096); the Natural Science Foundation of Qinghai Province (2019-ZJ-950Q).

并且电力供应已成为国民经济发展的制约重要性因素。因此,研究经济增长与电力消费的关系,尤其是分省区研究各省、直辖市、自治区的经济增长与电力消费的关系,对于剖析各地区经济增长对电力的依赖程度,制定有效的地区经济发展政策有着重要的现实意义。据相关文献统计数据,1952—2006年间,社会用电量和中国的GDP之间的相关度达到0.993,与第二产业和第三产业的相关度更是分别达到0.994和0.995。同时由于电力生产、传输和消费在瞬间完成及电力不可能大规模存储的特性,使得依据电力使用量判断社会经济走势的方法具有极高的准确度^[7-8]。用电需求在一定程度上成为中国经济增长快慢的“晴雨表”。针对电力需求与经济发展关系的研究一直为国内外学者所关注^[9-12]。因此,挖掘电能消费和社会经济增长之间的关系和相互影响的状况将变得愈发重要。

面板数据是在时间和截面空间上可以同时取得二维数据的一种数据类型^[13],由于其可以明显地显示出多个变量序列数据在空间和时间这两个不同维度上的相互关系,因此可将其应用于经济分析领域。常见的分析手段有:1)将经济数据进行时间序列平稳性的单位根检验,防止伪回归;2)通过协整检验来考察同阶单整的变量之间长期均衡关系;3)对面板数据进行回归和参数估计,目的是说明和体现变量之间可能存在的误差分析以及因果关系。

该类方法虽然具备多个变量在相同时间区间的动态演进过程,易于刻画时空特性的复杂数据等优点,但是电力系统量测数据采集频率高,导致其对数据源的预处理依赖性大,必须重点研究如何提取尽可能少的特征子集来描述电力系统的本质,从而快速降维以提高稳定分析的速度。文献[7]采用模糊算法对原始特征量进行压缩,但是该文献中提到的输入特征量数目相对较少,缺乏普遍意义;文献[8-10]分别采用Tabu搜索算法、二进粒子群优化算法和遗传算法进行特征提取,这3种算法本质上都属于寻优搜索算法,在实际寻优过程中计算量大、耗时长,不利于在线应用和推广;文献[11]采用了主成分分析法进行特征压缩,能够极大的消除冗余,但是也破坏了原有特征量的物理意义,不便于进行机理分析;文献[12]使用两阶段的特征提取法,以最后暂态分析结果为导向,采用逐步消去各个特征的试探,确定对分类结果影响力最大的特征组

合,该方法计算量大,而且可能对分类算法存在一定的依赖性而不具备普遍性;文献[13]采用灰色理论进行特征压缩,但在灰度聚类之后没有详细解释每个聚类中选择哪个特征作为该聚类的代表。

因此,本文提出了一种基于电力-经济时空特性的扩展面板数据的建模和分析方法。首先,通过平稳性和协整性的检验避免出现伪回归现象,并利用回归方程对用电量进行处理后获得权重因子;然后,使用聚类和灰色关联分析方法从原始特征量中提取得到有效特征量;最后,引入前面获得的权重因子对有效特征量进行筛选得到最优特征子集。本文提出的方法考虑到了以电力经济为输入源数据的时空维度特性,在保留各个特征量物理含义的同时,最大程度降低了由于原始特征量冗余信息带来的负面影响。并且,本文所提出的方法不依赖于具体分类算法,具有良好的普适性和操作性。

1 扩展面板数据模型的预处理

1.1 基于电力经济二元数据的扩展面板建模

面板数据是一种时间序列截面数据^[14]。面板数据能够提供更多的信息量,展现更多的变化,特别适合展示和分析不同区域、多个变量在相同时间区间的动态演进过程,即有利于刻画时空特性的复杂数据。目前,面板数据已被普遍应用到统计学和经济学等学科领域,也已被相关学者应用到电力市场领域进行城市居民用电消费情况^[15]的分析。

面板数据模型的一般形式如式(1)所示。

$$y_{it} = + \sum_{k=1}^K \beta_{kit} x_{kit} + \mu_{it} \quad (1)$$

式中: y_{it} 为被解释变量; x_{kit} 为解释变量; $i=1,2,3,\dots,N$,表示 N 个个体; $t=1,2,3,\dots,T$,表示 T 个取值时间点; k 为解释变量的数目; β_{kit} 为相对解释变量的待估计系数; μ_{it} 为随机误差项,并且相互独立。

在电力经济分析中,经济侧统计年鉴和电力侧计量量测用电数据分别具有不同的采集频率和量值,而且特征量数量大。以上特征量可以反映企业在不同层次上的空间特性,并且这些特征量是在时间轴上顺序取样的数据,是典型的时序数据,在将它们相互结合后,可以得到如表1所示的电力经济二元扩展面板数据。

表1 电力经济二元扩展面板数据

Table 1 The binary extended panel data of power economy

层次	指标	时刻1	时刻2	...	时刻T
产业	出口率	$X_{11}(1)$	$X_{11}(2)$...	$X_{11}(T)$
	出厂价格指数	$X_{12}(1)$	$X_{12}(2)$...	$X_{12}(T)$
	专利申请数
	⋮	⋮	⋮	⋮	⋮
	购进价格指数	$X_{1m}(1)$	$X_{1m}(2)$...	$X_{1m}(T)$
⋮	⋮	⋮	⋮	⋮	
电力	特征量1	$X_{11}(1)$	$X_{11}(2)$...	$X_{11}(T)$
	特征量2	$X_{12}(1)$	$X_{12}(2)$...	$X_{12}(T)$
	⋮	⋮	⋮	⋮	⋮
	特征量m	$X_{1m}(1)$	$X_{1m}(2)$...	$X_{1m}(T)$

其中,时间长度为 T , $X_{is}(t)$ 表示第 i 个电源管理单元(power management unit, PMU)节点第 s 个特征在 t 时间的数值。

1.2 基于面板数据的协整性以及平稳性检验

面板数据常见的分析步骤:对经济数据进行时间序列平稳性的单位根检验,防止伪回归;通过协整检验来考察同阶单整的变量之间长期均衡关系;对面板数据进行回归和参数估计,目的是说明和体现变量之间可能存在的误差分析以及因果关系。关于平稳性检验和协整性检验,相应数学工具的应用已经很成熟。早期的平稳性检验有 DF 或 ADF 单位根检验法^[16]、PP 单位根检验法^[17],目前比较常用的有 LLC 检验算法^[18]和 IPS 检验算法^[19]。需要特别指出的是,面板数据中的部分数据为非平稳序列,但是可以通过对其进行线性组合处理使其转化为平稳序列数据。且这一特点说明了面板数据中的变量存在相互影响但又相互独立的协整性特点。在数据分析中,Kao 检验^[20]和 Pedroni 检验算法^[21]因适用于较多的场景而广泛使用。

1.3 基于最小二乘法的电力经济回归方程

通过单位根检验和协整检验之后的面板数据可以进行基于最小二乘法的回归方程的参数估计。以电力系统电力经济中的用电量稳定为研究对象,回归方程以用电量为因变量、其他电力经济量为自变量,构造回归方程,如式(2)所示。

$$\delta_t = \sum_1^m C(i) \frac{dx_{it}}{dt} + \sum_m^n C(i)x_{it} + e_{it} \quad (2)$$

式中: δ_t 为用电量; x_{it} 为其他电力经济经济量,求

导关系表明单位根检验非平稳,一阶差分后平稳; $C(i)$ 为权重因子,表征该变量对用电量的影响力大小; e_{it} 为误差参数。

2 基于改进的灰色关联特征提取方法

2.1 基于电力经济的原始特征集合构建

电力经济二元数据包括电力侧数据和经济侧数据,具体的用电量、营销电量、一/二/三产业生产总值、一/二/三产业价格指数、一/二/三产业能源弹性系数、利润总额、一/二/三产业生产增加值、一/二/三产业产业构成比。本文以电力及经济二元数据为基础,综合分析文献[7-13],总结归纳了 8 维特征量作为原始特征量,如表 2 所示。其中, X1~X2 时间尺度为天, X3~X18 时间尺度为月。

表2 原始特征集

Table 2 Original feature sets

特征	特征描述
X1	日负荷冻结用电量
X2	日营销电量
X3	第一产业生产总值
X4	第二产业生产总值
X5	第三产业生产总值
X6	第一产业价格指数
X7	第二产业价格指数
X8	第三产业价格指数
X9	第一产业能源弹性系数
X10	第二产业能源弹性系数
X11	第三产业能源弹性系数
X12	用户利润总额
X13	第一产业生产增加值
X14	第二产业生产增加值
X15	第三产业生产增加值
X16	第一产业构成比
X17	第二产业构成比
X18	第三产业构成比

由 X1~X18 构成的特征值集合数量众多,并且不同类型特征值之间具有相互稳定而又相互影响的协整现象,这将影响计算速度。因此,本文利用基于灰色关联算法对以上特征值进行筛选降维,降低冗余特征值对计算带来的负面影响。

2.2 灰色关联分析

灰色关联度理论应用到电力系统电力经济评估的特征量选择的具体操作方法是^[22]:

根据多个特征量曲线的几何形状的相似程度来判断这些特征量的相互联系是否紧密。相似度分析是灰色聚类的基础,相似度越大,越有可能划分成同一类。将各个特征量进行描述,如定义1所示。

定义1:设 $X_0=(x_0(1),x_0(2),x_0(3),\cdots,x_0(n))$ 为系统特征序列,且相关序列表示如式(3)所示。

$$\begin{cases} X_1=(x_1(1),x_1(2),x_1(3),\cdots,x_1(n)) \\ \vdots \\ X_i=(x_i(1),x_i(2),x_i(3),\cdots,x_i(n)) \\ \vdots \\ X_m=(x_m(1),x_m(2),x_m(3),\cdots,x_m(n)) \end{cases} \quad (3)$$

给定实数 $\gamma(x_0(k),x_i(k))=\frac{1}{n}\sum_{k=1}^n\gamma(x_0(k),x_i(k))$ 满足如下灰色关联三公理。

1) 规范性:

$$0 < \gamma(X_0, X_i) \leq 1, \text{ 且 } \gamma(X_0, X_i) = 1 (X_0 = X_i)。$$

2) 整体性:

对于 $X_i, X_j \in X \{X_i | s = 0, 1, 2, \cdots, m, m \geq 2\}$, 有

$$\gamma(X_i, X_j) \neq \gamma(X_j, X_i), i \neq j。$$

3) 偶对称性:

对于 $X_i, X_j \in X$, 有 $\gamma(X_i, X_j) = \gamma(X_j, X_i) \Leftrightarrow X = \{X_i, X_j\}$ 。

4) 接近性:

$|x_0(k) - x_i(k)|$ 越小, $\gamma(x_0(k), x_i(k))$ 越大, 则称为 $\gamma(x_0(k), x_i(k))$ 与 $\gamma(x_0(k), x_i(k))$ 的灰色关联度, $\gamma(x_0(k), x_i(k))$ 为 $\gamma(x_0(k), x_i(k))$ 与 $\gamma(x_0(k), x_i(k))$ 在 $\gamma(x_0(k), x_i(k))$ 点的关联系数, 且 $\gamma(x_0(k), x_i(k))$ 的计算公式如式(4)所示。其中 $X_i=(x_i(1),x_i(2),x_i(3),\cdots,x_i(n))$ 是一个 $(0, 1)$ 的分布系数, 常取 $X_i=(x_i(1),x_i(2),x_i(3),\cdots,x_i(n))$ 。

$$\frac{\min_k \min_i |z| + \varepsilon \max_k \max_i |z|}{|z| + \varepsilon \max_k \max_i |z|} \quad (4)$$

$$x_0(k) - x_i(k) = z$$

特征量序列式(3)的折线记做 $s_i = \int_1^n (X_i - x_i(1))dt$, 令

$$s_i = \int_1^n (X_i - x_i(1))dt \quad (5)$$

当 X_0 和 X_i 序列长度相等时, 其灰色关联度如式(6)所示。

$$\varepsilon_{0i} = \frac{1 + |s_0| + |s_i|}{1 + |s_0| + |s_i| + |s_i - s_0|} \quad (6)$$

2.3 基于权重因子灰色聚类分析

灰色聚类将目标观测序列依据灰色关联矩阵分割为若干个类别, 各个类别独立观测时内部元素

具有较高的相似度; 而在相互对比时会发现各个类别之间差异较大。当有 n 个观测目标序列, 每个目标序列有 m 个特征值时, 可以将 X_i 和 X_j 的灰色绝对关联度转化为如式(7)所示的矩阵表现形式。式(7)中的元素可由式(6)求得。

$$C_{m \times m} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ 0 & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \varepsilon_{mm} \end{bmatrix} \quad (7)$$

$C_{m \times m}$ 是一个 $m \times m$ 的上三角方阵, 当其中元素 $\varepsilon_{ij} > r (i \neq j)$ 时, 可以将 X_i 和 X_j 看作同一类的特征, 其中 $r \in [0, 1]$, 一般情况下 $r > 0.5$ 。 r 的大小决定了聚类的严苛程度, 当 r 越大时, 证明聚类时的分类越多, 每个分类中的特征越少, 则分类越严格; 当 r 越小时, 证明聚类时的分类越少, 每个分类中的特征越多, 则分类要求越低。

设原始特征集 $\{X_1, X_2, \cdots, X_{N_1}\}$, $\{X_{N_1+1}, X_{N_1+2}, \cdots, X_{N_1+N_2}\}$ 经过灰色关联分析和灰色聚类分析之后, 能够聚为 N 类: $\{X_1, X_2, \cdots, X_{N_1}\}$, $\{X_{N_1+1}, X_{N_1+2}, \cdots, X_{N_1+N_2}\}$, $\{X_{N_1+N_2+1}, X_{N_1+N_2+2}, \cdots, X_M\}$ 。 在每个聚类的原始特征集选取特征值后的聚类选择约束是: 假设某一特征量与该聚类中其余特征量相比, 对电力经济的“影响因子”最大, 则提取其特征量。

2.4 特征提取算法流程

上述整个特征提取方法的流程图如图1所示。首先对原始特征集归一化处理, 然后进行绝对关联度计算和灰色聚类, 聚类后的分类依据面板数据的分析结果(权重因子)进行进一步的细化处理, 采用对用电量影响力最敏感的特征量作为同一聚类的代表, 最终输出分类结果。

3 算例仿真与分析

3.1 数据及其预处理

本文采用某省经济信息和电力量测信息作为数据来源。其中, 用电信息来自全省关口电能量计量系统, 2016年8月至2018年1月, 全省包含第一产业、第二产业和第三产业4万多用户每天一个点用户日冻结数据, 共计2100万; 经济类统计数据主要有国家统计局网站(发布《中国统计年鉴》)及贵州省统计局网站(发布《贵州统计年鉴》); 仿真软件

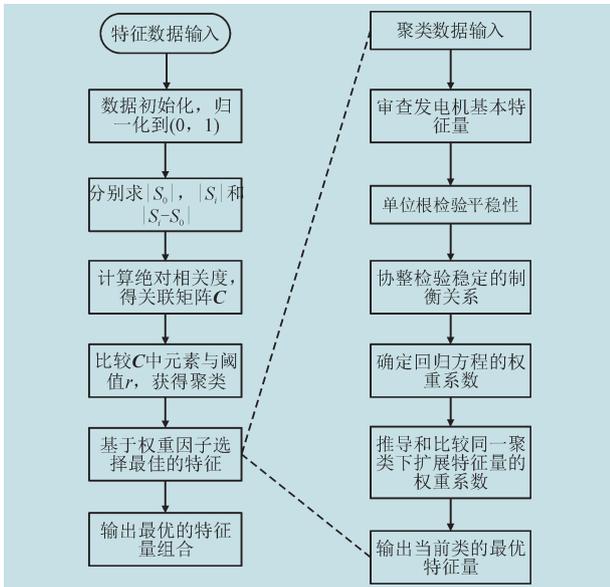


图1 特征提取流程图

Fig.1 Flow chart of the feature extraction

是 EView6.0。

首先由基础特征量构造面板数据,选择用电量、营销电量、生产总值、价格指数、能源弹性系数、利润总额、生产增加值作为基础特征量。选择暂态失稳的数据构造面板数据,然后进行单位根检验,结果如表 3 所示。

表3 单位根检验结果

Table 3 Results of unit root test

项目	水平值		一阶差分	
	LLC 检验	IPS 检验	LLC 检验	IPS 检验
电力-生产总值	1.693 (0.955)	0.051 (0.520)	8.733 (0.000)	-5.55** (0.000)
电力-价格指数	-0.770 (0.221)	1.257 (0.896)	-10.72** (0.000)	-8.74** (0.000)
电力-三产贡献率	-2.440 (0.006)	-4.32** (0.000)	68.591 (0.000)	-8.26** (0.000)
电力-生产增加值	0.280 (0.011)	-4.98** (0.000)	24.799 (0.000)	-13.99** (0.000)
电力-能源弹性系数	-3.57** (0.000)	-1.604 (0.005)	-1.246 (0.106)	-1.511 (0.066)
电力-投入产出完全消费系数	1.044 (0.852)	-5.60** (0.000)	6.19** (0.000)	15.84** (0.000)

注:**表示在 0.01 以下显著水平,括号内表示该统计量原假设成立概率

使用 Eviews6.0 进行的单位根检验,采用 LLC, IPS 检验算法。表 2 中括号内表示该统计量对应的原假设成立的概率值,依据文献[18-19],如果括号

内的概率值小于 0.01,说明原假设不成立,采信备用假设,序列是平稳的。综合分析看,如生产总值和生产增加值是一阶差分平稳的,其它基础特征值的水平值(原始值)是平稳的。然后使用 Kao 算法和 Pedroni 算法进行协整检验,结果如表 4 所示。

表4 协整检验结果

Table 4 Results of the cointegration test

算法	训练样本	统计量值	概率值
Pedroni	Panel v	4.451	0.000 0
	Panel rho	-1.596	0.053 9
	Panel PP	-1.956	0.024 3
	Panel ADF	-6.888	0.000 0
	Group rho	-1.761	0.041 2
	Group PP	-1.593	0.050 0
	Group ADF	-4.102	0.000 0
Kao	ADF	-6.247	0.000 0

由表 4 所示,结合文献[20-21]的评价标准,概率值小于 0.05 说明上述的原假设不成立,采信备用假设,满足协整关系。然后以用电量为因变量,用导数表达,如式(8)所示其他 5 个基础特征量作为自变量,采用最小二乘法进行线性回归,考虑到转速是一阶差分后平稳,用导数表达,如式(8)所示。

$$\delta_{it} = C(1)\frac{d\omega_{it}}{dt} + C(2)U_{it} + C(3)I_{it} + C(4)P_{it} + C(5)P_m + e_{it} \quad (8)$$

式中: δ_{it} 为用电量; ω_{it} 为生产值; U_{it} 为生产增加值; I_{it} 为价格指数; P_{it} 为能源弹性系数; e_{it} 为误差项。通过对式(8)进行拟合后可以求得: $C(1)=3.789$ 、 $C(2)=-4.332$ 、 $C(3)=-1.795$ 、 $C(4)=-0.539$ 、 $C(5)=0.6667$ 。以上系数反映各个经济量对用电量的“敏感程度”,即权重因子。

3.2 特征值提取

使用 2.1 节中的基础特征向量构造 18 维原始向量集合,然后使用 2.2 至 2.4 节阐述的灰色关联聚类的方法进行特征提取。首先计算获得 18 维的绝对关联度矩阵如表 5 所示。在该上三角矩阵中,关联度都是介于 0 和 1 之间的小数,越接近 1,说明两者的绝对关联度越大。然后以不同的关联度值作为门限进行聚类,尽可能不要产生争议,即某一个特征量同时与多个其它特征量的关联度都大于该门限。如果门限取得过高,很少的特征量将被聚为一类,就会失去聚类的意义,综合考虑之后,选择 0.84 作为门限进行聚类。结果为{X1, X3}、{X2, X4,

表5 18维特征向量绝对关联度矩阵

Table 5 Grey relation matrix of 18 original features in the new england system

特征	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18
X1	1	0.63	0.94	0.58	0.63	0.71	0.50	0.50	0.53	0.50	0.78	0.58	0.52	0.51	0.50	0.89	0.64	0.49
X2		1	0.62	0.94	0.87	0.73	0.50	0.50	0.71	0.5	0.73	0.50	0.50	0.53	0.53	0.58	0.54	0.50
X3			1	0.6	0.66	0.75	0.50	0.50	0.55	0.5	0.76	0.50	0.50	0.51	0.51	0.83	0.66	0.50
X4				1	0.83	0.7	0.50	0.50	0.73	0.5	0.7	0.50	0.50	0.53	0.53	0.57	0.53	0.50
X5					1	0.81	0.50	0.50	0.65	0.5	0.81	0.50	0.50	0.52	0.52	0.6	0.55	0.50
X6						1	0.50	0.50	0.59	0.50	0.94	0.50	0.51	0.50	0.50	0.66	0.54	0.50
X7							1	0.59	0.50	0.67	0.51	0.59	0.97	0.55	0.55	0.50	0.51	0.77
X8								1	0.51	0.89	0.50	0.97	0.56	0.50	0.50	0.50	0.51	0.69
X9									1	0.5	0.59	0.5	0.51	0.57	0.57	0.53	0.52	0.5
X10										1	0.5	0.82	0.64	0.52	0.52	0.5	0.5	0.76
X11											1	0.50	0.50	0.51	0.51	0.66	0.62	0.58
X12												1	0.53	0.51	0.51	0.51	0.50	0.66
X13													1	0.57	0.56	0.50	0.51	0.81
X14														1	0.99	0.5	0.5	0.54
X15															1	0.5	0.5	0.53
X16																1	0.75	0.5
X17																	1	0.5
X18																		1

X5}、{X7, X13}、{X8, X12}、{X14, X15}、{X17}、{X9}、{X10}、{X16}、{X18}一共 10 个聚类。

为了在每个聚类中挑选合适的特征量作为该聚类的“代表”，采用 3.1 中计算的权重因子分别代入每个特征值的计算公式中，将对用电量最为敏感的特征提取出来，获得最优的特征集合为：X2, X3, X6, X7, X8, X9, X10, X14, X16, X17, X18。共计 11 维特征向量。

3.3 算例分析

采用人工神经网络(ANN)、隐含马尔科夫(HMM)、支持向量机(SVM)和核心向量机(CVM)4个算法分别在原始特征和最优特征情况下，从训练/评估耗时和准确率 2 方面进行对比，结果如表 6 所示。其中，ANN 算法采用 Matlab 进行仿真分析；SVM 算法采用台湾大学林智仁开发的 LIBSVM 工具包；文献[23]为自行开发的 HMM 算法工具包；文献[24]为 CVM 算法提供了开源的工具包。

在挑选合适的聚类，即选取 11 维特征向量后，计算精度将会略微下降，其中采用 ANN 算法时下降超过 1%，其他算法的下降基本小于 1%。但是训练/评估耗时降低大约 30%，以上结果说明利用本文提出的方法进行特征值提取，可以在对准确率没有较大影响的基础上大幅度降低训练/评估耗时。

表6 不同算法采用 11 维最优特征的性能对比

Table 6 Performance comparison among 4 algorithms based on optimized and original feature sets

算法	核函数	8 维原始特征		6 维最优特征	
		训练/评估耗时	准确率/%	训练/评估耗时	准确率/%
ANN	NA	458 ms / 70 ms	96.4	380 ms / 50 ms	95.3
HMM	NA	827 ms / 97 ms	95.5	319 ms / 42 ms	94.6
SVM	线性	150 ms / 12 ms	93.7	110 ms / 9 ms	93.1
	径向基	990 ms / 12 ms	95.2	750 ms / 9 ms	94.3
CVM	多项式	98 ms / 35 ms	93.7	58 ms / 25 ms	92.5
	径向基	152 ms / 35 ms	94.1	68 ms / 25 ms	93.1

4 结论

电力-经济的关联性评估算法试图通过建立输入(用电量)与输出(经济量)一一对应的关系，来简化整个稳定性评估的复杂程度。在建立电力经济面板数据模型后，为了筛选出合适的特征量用于进一步仿真分析，本文使用改进的灰色关联聚类方法进行了特征提取。在将基于本文提出的方法求得的特征值代入不同类型的稳定评估算法后，得出以下结论：

1) 对电力经济二元数据进行面板数据建模，充分考虑了电力经济二元数据之间的时空关联特性，

为精确提取特征值提供了理论依据。

2) 为了消除特征值中的冗余信息,本文使用了改进的灰色关联聚类算法,这有助于表征电力经济特征的核心信息,在体现特征量自身物理含义的同时提高计算速度,并且适用于多种不同类型的智能算法分类器。

参考文献

- [1] 魏晓明,王蔚. 电力经济运行管理中的经济效益研究[J]. 电网与清洁能源, 2018, 34(4):6-9.
WEI Xiaoming, WANG Wei. Economic benefits in power economic operation management[J]. Power System and Clean Energy, 2018, 34(4):6-9.
- [2] 张琳,何炼成,王俊霞. 电力消费与中国经济增长——基于中国30省市面板数据的协整检验[J]. 山西财经大学学报, 2008,30(12): 18-21.
ZHANG Lin, HE Liancheng, WANG Junxia. Electricity consumption and economic growth in China: based on the 30 provincial panel data co-integration test[J]. Journal of Shanxi Finance and Economics University, 2008, 30(12): 18-21.
- [3] 姜磊,吴玉鸣. 电力消费与经济增长的区域差异研究——基于省级面板数据的单位根、协整与Granger因果检验[J]. 经济前沿, 2009(7): 28-39.
JIANG Lei, WU yuming. Regional differences in power consumption and economic growth: Based on unit roots, cointegration and Granger causality test of provincial panel data[J]. Regional Economy, 2009(7): 28-39.
- [4] 张龙,余敬,乌力雅苏. 中国电力消费与经济增长关系: 基于生产函数和Logistic曲线[J]. 数学的实践与认识, 2016, 46(3): 132-142.
ZHANG Long, YU Jing, WU Liyasu. Research of the relationship between electricity consumption and economic growth in China: based on production function and Logistic curve[J]. Mathematics in Practice and Theory, 2016, 46(3): 132-142.
- [5] GUTIÉRREZ A S, ERAS J J C, HUISINGH D, et al. The current potential of low-carbon economy and biomass-based electricity in Cuba. The case of sugarcane, energy cane and marabu (*Dichrostachys cinerea*) as biomass sources[J]. Journal of cleaner production, 2018, 172: 2108-2122.
- [6] PANG C K, PRABHAKARA F S. Security evaluation in power systems using pattern recognition[J]. IEEE Trans on Power Apparatus and Systems, 1974, 93(3): 969-976.
- [7] ZHANG Qi, HAN Zhenxiang, Tso S K, et al. Input dimension reduction in neural network training for transient stability assessment[J]. Automation of Electric Power Systems, 2001, 25(2): 32-35.
- [8] 谈一鸣,孙伟卿. 基于经济分析的上海电力景气研究[J]. 电网与清洁能源, 2016, 32(7): 45-50, 64.
TAN Yiming, SUN Weiqing. Electric power boom of Shanghai based on economic analysis[J]. Power System and Clean Energy, 2016, 32(7): 45-50, 64.
- [9] OSMAN M, GACHINO G, HOQUE A. Electricity consumption and economic growth in the GCC countries: Panel data analysis[J]. Energy Policy, 2016, 98:318-327.
- [10] RAJAH R, GEOFFREY G. Are foreign firms more productive and export- and technology-intensive than local firms in kenyan manufacturing[J]. Oxford Development Studies. 2005, 33(2): 211-227.
- [11] HOQUE A, BANERJEE R. Does purchasing power parity hold for garment export-oriented developing countries[J]. Procedia - Social and Behavioral Sciences, 2012, 65(3): 8-13.
- [12] NAKAMURA K, OHTSUKI T, MORI H, et al. Novel anti-HIV-1 activity produced by conjugating unsulfated dextran with poly l-lysine[J]. Antiviral Research, 2012, 94(1): 89-97.
- [13] HADJIMICHAEL M, KUCIAUSKAS A P, TAG P M, et al. A meteorological fuzzy expert system incorporating subjective user input[J]. Knowledge and Information Systems, 2002(3): 350-369.
- [14] Choon Seong Leem, BUMYONG Oh. Evaluating information strategic planning: an evaluation system and its application[J]. Journal of Systems Integration, 2001, 10(3): 207-222.
- [15] 王志刚. 面板数据模型及其在经济分析中的应用[M]. 北京: 经济科学出版社, 2008: 1-5.
- [16] 焦艳燕. 基于数据挖掘方法的居民生活用电负荷预测研究[D]. 保定: 华北电力大学, 2009.
- [17] DICKEY A D, FULLER W A. Likelihood ratio statistics for autoregressive time series with unit root[J]. Econometrica, 1981, 49(4): 1057-1072.
- [18] 高铁梅. 计量经济分析方法与建模Eviews应用及实例[M]. 北京: 清华大学出版社, 2005: 331-347.
- [19] LEVIN A, LIN C F, CHU C S J. Unit root tests in panel data: asymptotic and finite-sample properties[J]. Journal of econometrics, 2002, 108(1): 1-24.
- [20] IM K S, PESARAN M H, SHIN Y. Testing for unit roots

(下转第78页)

of chance-constrained programming in rural distribution network[J]. Renewable Energy Resources, 2019, 37(8): 1185-1191.

- [15] 李二霞,亢超群,李玉凌,等.基于设备状态评价和电网损失风险的配电网检修计划优化模型[J].高电压技术, 2018,44(11):3751-3759.

LI Erxia, KANG Chaoqun, LI Yuling, et al. Optimization model of distribution network maintenance plan based on equipment condition evaluation and grid loss risk[J]. High Voltage Engineering, 2018,44(11):3751-3759.

- [16] 傅强,向红伟,刘彪,等.分布式可再生能源接入配电网的运检风险评估[J].可再生能源, 2019, 37(7):1001-1006.

FU Qiang, XIANG Hongwei, LIU Biao, et al. Maintenance operation risk assessment of distributed generation on

distribution network[J]. Renewable Energy Resources, 2019, 37(7):1001-1006.

收稿日期:2020-05-10。

作者简介:

阙波(1961—),男,本科,教授级高工,研究方向为电力系统及其自动化;

黄武浩(1973—),男,硕士,教授级高工,研究方向为电力系统及其自动化;

杨松伟(1976—),男,本科,高级工程师,研究方向为电力系统及其自动化;

钱平(1979—),男,本科,高级工程师,主要从事变电运检专业管理的研究工作;

胡俊华(1981—),男,本科,工程师,研究方向为电气工程及其自动化。

(编辑 沈卢)

(上接第70页)

in heterogeneous panels[J]. Journal of econometrics, 2003, 115(1): 53-74.

- [21] MCCOSKEY S, KAO C. A residual-based test of the null of cointegration in panel data[J]. Econometric Reviews, 1998, 17(1): 57-84.

- [22] PEDRONI P. Fully modified OLS for heterogeneous cointegrated panels and the case of purchasing power parity [J]. Manuscript, Department of Economics, Indiana University, 1996, 5: 1-45.

- [23] 刘思峰,谢乃明.灰色系统理论及其应用:第五版[M].

北京:科学出版社,2008:1-3.

- [24] TSANG I W, KWOK J T, CHEUNG P M. Core vector machines: Fast SVM training on very large data sets [J]. The Journal of Machine Learning Research, 2005, 6(1): 363-392.

收稿日期:2020-07-10。

作者简介:

张秋雁(1965—),男,本科,教授级高级工程师,主要从事电能计量技术和电力大数据研究。

(编辑 沈卢)