

# Time Series Two-Dimensional Piecewise Cloud Model for Distribution Network Power Consumption Mode Partition

Haijing Wang<sup>1</sup>, Yiping Cui<sup>1</sup>, Tian Liu<sup>1</sup>, Sijie Tang<sup>2</sup>, Chengjie Pan<sup>2</sup>, Jinmei Chen<sup>2</sup>

<sup>1</sup>Guangzhou Power Supply Bureau Co., Ltd. Electric Power Test & Research Institute, Guangzhou Guangdong

<sup>2</sup>School of Electrical Engineering, Xi'an Jiaotong University, Xi'an Shaanxi

Email: 492604202@qq.com

Received: Apr. 6<sup>th</sup>, 2019; accepted: Apr. 18<sup>th</sup>, 2019; published: Apr. 25<sup>th</sup>, 2019

## Abstract

In order to solve the problem that the length of time series data such as electricity consumption is too large and the similarity measurement is not accurate, a time series analysis method based on two-dimensional piecewise cloud model is proposed in this paper. In this method, the power consumption time series data are first represented by a two-dimensional piecewise cloud model, and then the similarity of different cloud models is measured based on the method of calculating the overlapping area of the expected curve. Finally, these time series are classified by K-nearest neighbor algorithm, and the experimental results are compared with the traditional methods. The experimental results show that this method can effectively improve the accuracy of power consumption pattern classification on the user side of distribution network.

## Keywords

Distribution Network Power Consumption Mode Division, Time Series Classification, Two-Dimensional Piecewise Cloud Model

# 基于时间序列二维分段云模型的配网用电模式划分方法的研究

王海靖<sup>1</sup>, 崔屹平<sup>1</sup>, 刘田<sup>1</sup>, 汤思杰<sup>2</sup>, 潘程杰<sup>2</sup>, 陈金梅<sup>2</sup>

<sup>1</sup>广州供电局电力试验研究院, 广东 广州

<sup>2</sup>西安交通大学电气工程学院, 陕西 西安

Email: 492604202@qq.com

收稿日期: 2019年4月6日; 录用日期: 2019年4月18日; 发布日期: 2019年4月25日

## 摘要

在对配网用电模式进行划分时,为了解决用电量等时间序列数据长度过大及相似性度量不精确的问题,本文提出了一种基于二维分段云模型的时间序列分析方法。该方法首先将用电时间序列数据用二维的分段云模型来表示,然后在基于计算期望曲线重叠面积的方法上对不同云模型的相似性进行度量,最后通过K-最邻近算法对这些用电时间序列进行分类,并将实验结果与传统方法进行比较。实验结果表明:该方法能有效提高对配网用户侧用电模式分类的准确率。

## 关键词

配网用电模式划分, 时间序列分类, 二维分段云模型

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

智能电表的普及使得电网公司能够掌握海量的电力客户用电数据[1],为了从这些用电数据中获得有用的信息,从而更加深刻地了解配网用户侧的用电模式与特点,进而为制定更具个性化的用户服务方案和对策提供支撑[2]。因此,对于配网用户侧用电模式的划分成为了目前科学家们研究的热点。

近年来,不少学者对配网用电模式划分进行了研究。文献[3]提出了 k-means 的并行算法来对用电客户进行分类。但由于这种并行算法往往受制于计算机的硬件,使得计算效率受到严重的影响[4]。文献[5]通过总结电价变动对用户用电情况影响的规律,来为电网公司提供确定电价的标准。文献[6]基于熵权法来对划分好的评价指标进行衡量,从而反映不同用户的价值与潜力。但是该法在确立指标时主观性过大,因此会影响分类结果的客观性。文献[7]基于模糊综合评价的方法来对电力客户用电行为进行量化分级,文献[8][9]为了减少聚类算法的复杂度,对负荷曲线数据采取先降维再聚类的处理过程。但由于该方法是直接对数据进行降维,因此具有一定的盲目性,且易丢失原始数据中的有用信息。此外,以上方法绝大部分是应用于大型的工业和农业用户层面的,在对于配网用户,尤其是居民的用电模式分析上应用甚少,且在应用过程中常常会出现分类繁杂且不准确的结果。而通过采集用户家用智能电表上的时间序列数据进行分析,因数据能够直接反映用户的用电习惯和特点,因此分类结果更加具有客观性和准确性。目前最为常见的时间序列分析方法是 DTW(动态时间规整)算法。该算法的优点是技术上较为简单,资源要求较小。但是由于该算法需要很大的运算量,往往在海量数据的分析中,运行效率和分类的准确率都无法达到要求[10]。

因此本文提出了基于二维分段云模型的时间序列分析法,通过对从用户智能电表上采集的时间序列数据进行处理,从而实现了对配网用户用电模式的划分。最终实验结果表明:该方法能够有效提高对配网用户侧的用电模式分类的准确率。

## 2. 云模型的基本概念

设  $X$  为一个普通的集合,  $X = \{x\}$ , 为论域, 如果对于论域里的任意一个元素  $(x_1, x_2)$ , 都存在有且仅有一个与它相对应的数  $\mu_{\tilde{A}}(x_1, x_2) \in [0, 1]$ , 则将  $\mu_{\tilde{A}}(x_1, x_2)$  组成的集合  $\tilde{A}$  称为  $X$  上的模糊集合, 将

$\mu_{\tilde{A}}(x_1, x_2)$  称为  $X$  中元素对  $\tilde{A}$  的隶属度。如果  $X$  中的元素分布是简单而且有序的, 那么可以将  $X$  看作是基本变量, 隶属度  $\mu_{\tilde{A}}(x)$  在  $X$  上的分布我们称其为隶属云; 而如果  $X$  之中的元素分布不是简单而且有序的, 但根据某个具体的规则  $f$ , 我们可以把  $X$  映射到另外一个有序的论域  $X'$ , 并且  $X'$  之中有且只有一个  $(x'_1, x'_2)$  和  $X$  中的  $(x_1, x_2)$  对应, 则我们称  $X'$  为基本变量, 隶属度  $\mu_{\tilde{A}}(x)$  在  $X'$  上的分布我们称之为隶属云 [11], 即上文提到的二维云模型。

论域中的特定的点到概念的隶属度并不是固定的, 总是存在着较小的差异, 而实际上这种变化是具有稳定倾向的, 这对云模型特征没有特别明显的影响, 云滴的这种分布特性集中体现了定性概念具有模糊性与随机性的特点。

二维云模型用期望  $(Ex_1, Ex_2)$ , 熵  $(En_1, En_2)$  和超熵  $(H_1, H_2)$  来表示其数学性质:

期望  $(Ex_1, Ex_2)$ : 二维云模型云滴的空间分布在  $x_1Ox_2$  平面上投影的中心, 是云滴在该平面上的分布最有代表性的点。

熵  $(En_1, En_2)$ : 二维云模型云滴在  $x_1Oy$  和  $x_2Oy$  平面上的分布的期望曲线的熵, 反映了二维云模型云滴在  $x_1Oy$  和  $x_2Oy$  平面上的分布的不确定度。

超熵  $(H_1, H_2)$ : 二维云模型云滴在  $x_1Oy$  和  $x_2Oy$  平面上的分布的期望曲线的超熵, 反映了二维云模型云滴在  $x_1Oy$  和  $x_2Oy$  平面上的分布的熵的不确定度 [12]。

### 3. 二维分段云模型

本节将介绍对于原始的用电时间序列数据分段策略以及两个二维云模型之间相似性的度量方法。

#### 3.1. 时间序列的分段策略

为了解决在对原始的用户用电时间序列数据分段时存在的无法平均分段以及容易丢失大量有效信息的问题, 本文提出了一种灵活的重叠分割策略(OPS)。它允许两个相邻的片段相互重叠并且具有  $M$  个公共点。这样, 当  $M$  足够大时, 我们保持联系紧密的点之间的连接, 保护有效信息免于丢失。OPS 的实现过程如下:

设从智能电表采集到的用电时间序列为  $T = (t_1, t_2, \dots, t_N)$ , 目标分段数为  $w$ , 分段后时间子序列间的重叠点个数为  $M$ 。

首先, 计算所有子序列的总长度  $L$  以及重叠率  $O$ :

$$L = \sum_{i=1}^w \text{length}(S_i) = N + (W - 1) \times M$$

$$O = (W - 1) \times M / L \times 100\%$$

当  $L$  被  $W$  所除时, 计算商数  $Q$  以及余数  $R$ 。对于前  $R$  段子序列, 其构成为:

$$S_i = (t_{i \times (Q+1-M)+1}, t_{i \times (Q+1-M)+2}, \dots, t_{i \times (Q+1-M)+Q})$$

对于剩余的  $W-R$  段时间子序列, 其构成为:

$$S_{W-j+i} = (t_{N-(j-1) \times (Q-M)-Q+1}, t_{N-(j-1) \times (Q-M)-Q+2}, \dots, t_{N-(j-1) \times (Q-M)})$$

至此, 我们将原始的长度为  $N$  的用电时间序列数据分段为  $W$  条子序列  $S_1, S_2, \dots, S_W$ , 且它们间的重叠率  $O$ 。

#### 3.2. 二维分段云模型间的相似性度量

为了解决不同原始时间序列之间相似性度量不准确的问题, 本节将介绍两个二维分段云模型之间相似性的度量方法。

设两个原始时间序列为：

$$T = (t_1, t_2, \dots, t_n), \bar{T} = (\bar{t}_1, \bar{t}_2, \dots, \bar{t}_n)$$

通过 OPS 分段后可表示为：

$$T' = (t'_1, t'_2, \dots, t'_w), \bar{T}' = (\bar{t}'_1, \bar{t}'_2, \dots, \bar{t}'_w)$$

其中每一段子序列可以表示为：

$$t'_i = (T_i, \Delta T_i) = ((t_1, e_1), \dots, (t_{d_i}, e_{d_i})), \bar{t}'_i = (\bar{T}_i, \Delta \bar{T}_i) = ((\bar{t}_1, \bar{e}_1), \dots, (\bar{t}_{d_i}, \bar{e}_{d_i}))$$

将上述的二维时间子序列分别投影到各自维度，则可表示为两个独立的一维时间子序列，即两个独立的一维列向量  $X$  和  $Y$ ，将投影后的二维时间子序列代入逆向云发生器中，可将时间序列转化为二维云模型。逆向云发生器的实现过程如下：

输入：列向量  $X = (x_1, x_2, \dots, x_n)$ 、列向量  $Y = (y_1, y_2, \dots, y_n)$

步骤 1，计算期望值  $Ex, Ey$ ：

$$Ex = \text{mean}(X) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$Ey = \text{mean}(Y) = \frac{1}{n} \sum_{i=1}^n y_i$$

步骤 2，计算熵  $Enx, Eny$ ：

$$Enx = \frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{i=1}^n |x_i - Ex|$$

$$Eny = \frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{i=1}^n |y_i - Ey|$$

步骤 3，计算超熵  $Hex, Hey$ ：

$$Hex = \sqrt{\text{var}(X) - Enx^2}$$

$$Hey = \sqrt{\text{var}(Y) - Eny^2}$$

输出：以参数  $(Ex, Ey)$ 、 $(Enx, Eny)$ 、 $(Hex, Hey)$  为特征的二维云模型，用符号  $C$  表示。

将所有的时间子序列分别代入逆行云发生器中，则两个二维分段云模型分别可表示为：

$$\begin{cases} C_1(Ex_1, Ey_1, Enx_1, Eny_1, Hex_1, Hey_1) \\ C_2(Ex_2, Ey_2, Enx_2, Eny_2, Hex_2, Hey_2) \\ \vdots \\ C_w(Ex_w, Ey_w, Enx_w, Eny_w, Hex_w, Hey_w) \\ \bar{C}_1(\bar{Ex}_1, \bar{Ey}_1, \bar{Enx}_1, \bar{Eny}_1, \bar{Hex}_1, \bar{Hey}_1) \\ \bar{C}_2(\bar{Ex}_2, \bar{Ey}_2, \bar{Enx}_2, \bar{Eny}_2, \bar{Hex}_2, \bar{Hey}_2) \\ \vdots \\ \bar{C}_w(\bar{Ex}_w, \bar{Ey}_w, \bar{Enx}_w, \bar{Eny}_w, \bar{Hex}_w, \bar{Hey}_w) \end{cases}$$

至此得到了原始时间序列数据的二维分段云模型，二维云模型同时关注了原始时间序列的分布和变化特征，因此保留了更多的原始有效信息，为接下来的相似性度量提供了准确性的保证。

首先我们对同一对应段的二维云进行相似性的度量。以第  $i$  段二维云为例, 将第  $i$  段二维云在两个维度上分别进行投影, 则  $C_i(Ex_i, Ey_i, Enx_i, Eny_i, Hex_i, Hey_i)$  被分解为  $Cx_i(Ex_i, Enx_i, Hex_i)$  和  $Cy_i(Ey_i, Eny_i, Hey_i)$  两个独立的一维云。同理, 对  $\bar{C}_i(\bar{E}x_i, \bar{E}y_i, \bar{E}nx_i, \bar{E}ny_i, \bar{H}ex_i, \bar{H}ey_i)$  进行投影, 可得  $\bar{C}x_i(\bar{E}x_i, \bar{E}nx_i, \bar{H}ex_i)$  和  $\bar{C}y_i(\bar{E}y_i, \bar{E}ny_i, \bar{H}ey_i)$  两个对应的一维云。然后我们通过对期望曲线重叠面积的计算 [13], 得到同一维度下一维云模型之间的相似度  $OMx_i$  和  $OMy_i$ , 将这两个相似度相乘, 即可得到第  $i$  段二维云之间的相似度  $TCM_i$ 。

用上述方法可以计算出两个二维分段云模型中对应的每一段二维云的相似度, 则两个二维分段云模型之间的相似度计算公式为:

$$SM(T', \bar{T}') = \sqrt{\frac{1}{w} \sum_{i=1}^w TCM_i}$$

最后, 我们采用 KNN (K 最邻近分类) 算法, 以已经计算好的二维云模型之间的相似度为依据, 对采集的用电时间序列数据进行分类。在下一节中, 我们将通过从智能电表中实际采集到的时间序列数据对该方法进行验证。

#### 4. 实验结果与分析

本实验数据来自 UCR 时间序列分类库中的公共数据集 Small Kitchen Appliances, 该数据集记录了英国 251 个家庭在一个月内的用电数据, 每条时间序列长度为 720 (在 24 小时内, 每两分钟记录一次数据), 将用户行为共分为 3 类。其中训练集包含 375 条的时间序列, 测试集包含 375 条时间序列。

为了使不同的时间序列在后续的相似性度量等处理过程中不会因为原始数据的量纲不同而造成巨大影响, 我们需要对从智能电表采集到的原始用电时间序列数据进行归一化处理:

设原始的用电时间序列为  $T = \{t_1, t_2, \dots, t_m\}$ , 则:

$$t_{i-nm} = \frac{t_i - t_{\min}}{t_{\max} - t_{\min}}$$

其中  $t_{\max} = \max(t_1, t_2, \dots, t_m)$ ,  $t_{\min} = \min(t_1, t_2, \dots, t_m)$ 。

归一化后的用电时间序列为:  $T_{nm} = \{t_{1-nm}, t_{2-nm}, \dots, t_{m-nm}\}$ 。

另外, 为了减少平移和缩放对相似性的影响, 我们需要对用电原始时间序列进行标准化处理, 常用的方法是 z 标准化(Z-score)。

设原始的用电时间序列为  $T = \{t_1, t_2, \dots, t_m\}$ , 其平均值  $t_u$  和标准偏差  $t_\sigma$  分别为:

$$t_u = \sum_{i=1}^m t_i$$

$$t_\sigma = \sqrt{\frac{1}{m} \sum_{i=1}^m (t_i - t_u)^2}$$

则由式:

$$t_{i-sd} = \frac{t_i - t_u}{t_\sigma}$$

标准化后的原始用电时间序列为:  $T_{sd} = \{t_{1-sd}, t_{2-sd}, \dots, t_{m-sd}\}$

本实验首先将原始时间序列代入二维分段云模型中, 通过计算测试集序列与训练集序列的二维分段云模型相似度, 采用 K-最邻近算法实现对时间序列的分类。由于二维分段云模型分段长度的选择与 K-最邻近算法中  $K$  的取值都会影响到分类结果, 因此首先对这两个参数进行优化。

考虑到二维分段云模型每一段子序列的长度对于分类的合理性与准确率有较大的影响，我们取二维云模型分段数目  $w$  为 1~20 进行实验，测试结果如图 1 所示，当分段数  $w$  取 7 和 12 时，即每段子序列的分段长度为 103 和 60 时，模型的分类效果最好。而当  $w$  取值较小时，二维分段云模型丢失了大量的原始序列信息，因此分类效果不佳；当  $w$  的值较大时，分段后的子序列长度已经很短，因此无法达到减小时间序列长度的效果，同样影响了分类的性能。

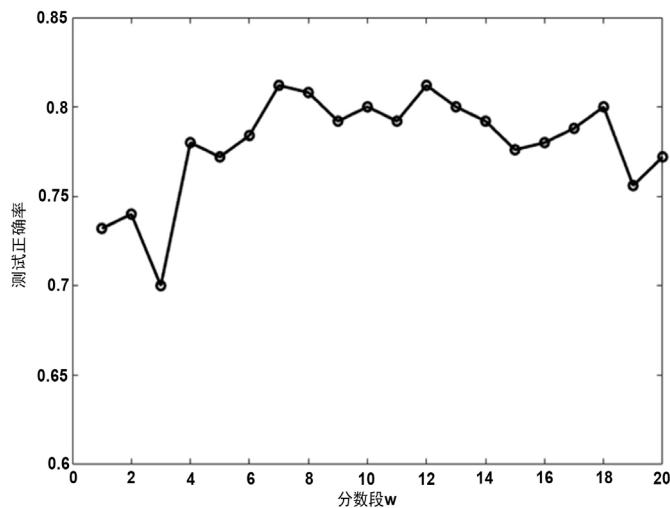


Figure 1. The influence of the number of time series segments on the experimental results

图 1. 时间序列分段数对实验结果的影响

本实验由于采用的 K-最近邻算法进行分类，因此其参数  $K$  的取值对分类的结果存在着很大的影响，为了使  $K$  的取值尽可能的合理，以避免因  $K$  的取值不佳而错误的评估了分段云模型的性能，我们取 10 组数值  $\{1, 2, \dots, 10\}$ ，为了控制变量，我们取分段数  $w = 7$ ，进行实验，其结果如图 2，当  $K = 7$  时，二维分段云模型的分类效果最好，正确率达 83.6%。

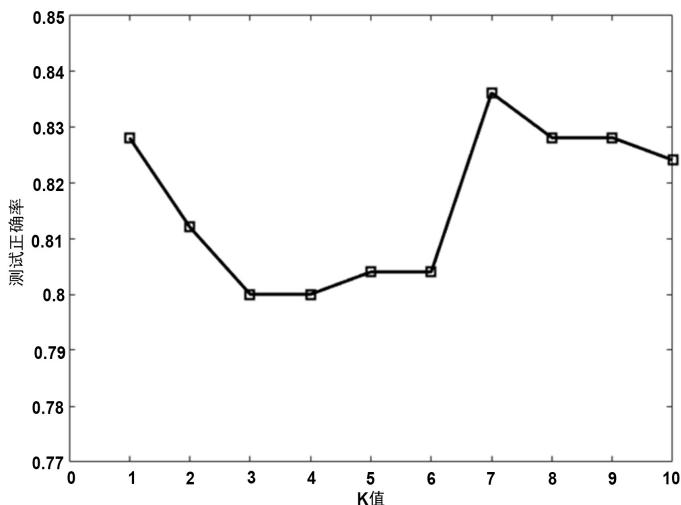


Figure 2. The influence of  $K$  value on the experimental results

图 2.  $K$  取值对实验结果的影响

为了能够更客观地评价二维分段云模型的性能，我们将传统时间序列分类方法 DTW (动态时间归整)与二维分段云模型两种方法进行对比。为了减少因数据样本选择所带来的偶然性，我们将训练集随机抽样三次，每次抽取的新训练集包含 240 条时间序列。将各参数设为最优值，进行实验，分类正确率如表 1 所示，针对本数据集，传统的 DTW (动态时间归整)方法的分类效果对样本的选择十分敏感，且正确率低于本文所提出的二维分段云模型的性能。因此，可以验证二维分段云模型具有较好的分类性能。

**Table 1.** Comparison of the experimental results of the two models

**表 1.** 两种模型实验结果对比

模型	训练集 1	训练集 2	训练集 3	平均正确率
二维分段云模型	82%	80.8%	82%	81.6%
DTW	79.2%	55.9%	51.2%	62.1%

## 5. 结论

本文提出了一种基于二维分段云模型的时间序列分析方法，用于解决在对配网中的用电时间序列数据进行分析时所出现的长度过大和相似性度量不准确的问题。该方法首先将用电时间序列数据用二维的分段云模型来表示，然后在基于计算期望曲线重叠面积的方法上对不同云模型的相似性进行度量，最后通过 K-最邻近算法对用电时间序列进行分类。实验结果表明，针对本文所用数据集，当分段数  $w$  取 7， $K$  值取 7 时，二维分段云模型的性能正确率高达 83.6%。最后与传统的 DTW 动态时间规整方法进行对比，得到二维分段云模型对配网用电模式的分类效果更好。

## 基金项目

中国南方电网有限责任公司科技项目(080037KK52170051 (GZHKJXM20170104))。

## 参考文献

- [1] 袁硕, 陈礼定, 孙国鹏, 林金官. 基于时间序列的电力负荷数据分析[J]. 应用数学进展, 2016, 5(2): 214-224.
- [2] 刘凯悦. 大数据综述[J]. 计算机科学与应用, 2018, 8(10): 1503-1509.
- [3] 张素香, 刘建明, 赵丙镇, 曹津平. 基于云计算的居民用电行为分析模型研究[J]. 电网技术, 2013, 37(6): 1542-1546.
- [4] 吴洁璇, 陈振杰, 张云倩, 等. 多核 CPU 下的 K-means 遥感影像分类并行方法[J]. 计算机应用, 2015, 35(5): 1296-1301.
- [5] 董军, 张晓虎, 李春雪, 等. 自动需求响应背景下考虑用户满意度的分时电价最优制定策略[J]. 电力自动化设备, 2016, 36(7): 67-73.
- [6] 王璨, 冯勤超. 基于价值评价的电力用户分类研究[J]. 价值工程, 2009, 28(5): 64-67.
- [7] 何永秀, 王冰, 熊威, 张婷, 刘洋洋. 基于模糊综合评价的居民智能用电行为分析与互动机制设计[J]. 电网技术, 2012, 36(10): 247-252.
- [8] 张斌, 庄池杰, 胡军, 等. 结合降维技术的电力负荷曲线集成聚类算法[J]. 中国电机工程学报, 2015, 35(15): 3741-3749.
- [9] Wang, Y. and Chen, Q.X., Kang, C.Q., et al. (2016) Sparse and Redundant Representation-Based Smart Meter Data Compression and Pattern Extraction. *IEEE Transactions on Power Systems*, 32, 2142-2151. <https://doi.org/10.1109/tpwrs.2016.2604389>
- [10] 陆薛妹, 胡轶, 方建安. 基于分段极值 DTW 距离的时间序列相似性度量[J]. 微计算机信息, 2007, 23(27):

204-206.

- [11] Fuchs, E., Gruber, T., Nitschke, J. and Sick, B. (2010) Online Segmentation of Time Series Based on Polynomial Least-Squares Approximations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 2232-2245. <https://doi.org/10.1109/tpami.2010.44>
- [12] Markellos, V.V., Black, W. and Moran, P.E. (1974) A Grid Search for Families of Periodic Orbits in the Restricted Problem of Three Bodies. *Celestial Mechanics*, **9**, 507-512. <https://doi.org/10.1007/bf01329331>
- [13] .Li, H.L. and Guo, C.H. (2011) Piecewise Cloud Approximation for Time Series Mining. *Knowledge-Based Systems*, **24**, 492-500. <https://doi.org/10.1016/j.knosys.2010.12.008>

**知网检索的两种方式:**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)