

DOI: 10.12158/j.2096-3203.2020.02.022

# 基于 AdaBoost-DT 算法的电力市场串谋行为识别研究

张海生<sup>1</sup>, 曹喆<sup>1</sup>, 杨昌海<sup>1</sup>, 骆雲鹏<sup>2</sup>, 华回春<sup>2</sup>

(1. 国网甘肃省电力公司经济技术研究院, 甘肃 兰州 730050;

2. 新能源电力系统国家重点实验室(华北电力大学), 河北 保定 071003)

**摘要:**针对电力市场中购电商串谋的识别方法定性分析居多, 实时性不高的问题, 文中提出基于 AdaBoost-DT 算法的串谋行为智能识别方法, 将 AdaBoost-DT 集成分类算法用于串谋识别中, 解决了串谋行为难以量化识别的问题。从串谋机理出发, 设计了一套基于任意 2 个购电商之间的串谋识别指标体系。面对数据不均衡问题, 采用过采样对训练数据集进行增广, 利用 AdaBoost-DT 分类算法训练串谋行为智能识别模型。最后, 以月度交易数据为支撑进行算例分析, 采用接收者操作特性曲线 (ROC) 和接收者操作特性曲线下的面积 (AUC 值) 评价模型的识别效果。实验结果表明, 该串谋行为识别方法的准确率较高且实时性较好, 充分验证了算法的有效性。

**关键词:** 电力市场; 串谋; 过采样; 决策树; AdaBoost-DT

中图分类号: TM744

文献标志码: A

文章编号: 2096-3203(2020)02-0152-07

## 0 引言

2015年, 随着《中共中央国务院关于进一步深化电力体制改革的若干意见》的公布, 新一轮电力市场体制改革拉开大幕<sup>[1]</sup>。在电力体制改革的迅速推进下, 售电侧被逐渐解放, 电力市场多元化的竞争格局逐步形成。然而新电改下的电力市场交易过程中“滥用市场力”、“恶意串谋”等违规行为较多, 严重危害电力市场的稳定发展。

现在, 我国的电力市场交易制度还不完善, 如果电力市场主体为了扩大利益, 采用串谋的方式进行违规报价, 将会影响市场价格, 甚至对整个电力市场的信用和诚信交易造成巨大影响。因此, 对于串谋这种潜在的危害行为, 建立一种实时、高效的识别方法迫在眉睫<sup>[2-3]</sup>。通常, 串谋主要有 2 种方式, 一种是通过协议进行直接联盟; 另一种就是默契串谋。由于反垄断法的制约, 默契串谋被企业采用的较多。针对默契串谋, 平行定价行为方法被认为是较有效的识别方法之一。如果销售同一种产品的企业在同一时间或者同一容量段改变价格的比例基本相同, 即产生了平行定价行为, 有理由认为其进行了串谋<sup>[4]</sup>。目前, 关于电力市场串谋行为的研究, 主要以机理、监测指标等定性分析为主。文献[5-7]系统介绍了电力市场交易过程中的违规行为以及识别指标。文献[8]针对电力市场潜在的危害行为给出基于云模型和模糊 Petri 网的识别

方法。文献[9-10]详细介绍了串谋溢价行为规制方法以及神经网络在市场主体信用评价方面的应用。在这些研究中, 针对串谋行为的识别主要集中在定性分析和指标研究等方面, 受主观影响较大且操作复杂, 效率性和实时性有待考证。

现在, 机器学习算法被广泛应用在工程监测和识别方面, 可以有效发掘数据集内部潜在联系, 在实时性和效率方面有较好表现。文献[11-14]详细介绍了深度学习算法和分类算法在处理不同监测问题方面的有效性和优良性能。电力市场串谋识别问题, 实质上是一个二分类问题, 基于决策树的集成分类算法 AdaBoost-DT 在解决二分类问题上有着极好的效果<sup>[15-17]</sup>。

文中在平行定价行为理论的基础上, 提出一套改进的串谋监测指标体系, 并将 AdaBoost-DT 分类算法应用到识别过程, 构建一种串谋行为智能识别方法。阐述了串谋行为智能识别具体流程, 并通过购电商报价数据进行算例分析, 验证了 AdaBoost-DT 智能识别方法在串谋行为识别方面的有效性。

## 1 改进的串谋识别指标体系

电力市场主体的串谋行为可以通过其报价策略反映出来。文中对购电商的报价策略和交易过程中串谋行为的监测指标<sup>[5-6]</sup>进行了详细研究, 结合电力市场的实际情况, 对相关的指标进行改进, 构建一种新的串谋行为监测指标体系。

串谋行为的发生是建立在 2 个及以上购电商之间的行为, 如果采用只针对单一市场主体的串谋监测指标, 得出的结果不能很好地符合实际需求且不

收稿日期: 2019-10-17; 修回日期: 2019-11-26

基金项目: 国家电网有限公司科技项目 (SGGSJY00PSJS190-0060); 中央高校基本科研业务费项目 (2017MS197)

能直观地判断串谋是哪些市场主体之间的行为,会对后续的串谋行为处理办法的落实造成一定的困难。考虑到上述问题,文中在平行定价行为的基础上,改进以往提出的只针对单个市场主体的可信度较高的监测指标,构建基于任意 2 个市场主体之间的监测指标,并且提出新的监测指标:报价曲线差异面积比率,进而建立了一套基于任意 2 个主体之间的购电商串谋行为识别指标体系,如表 1 所示。

表 1 购电商串谋行为指标体系  
Table 1 Index system of collusive behavior of e-commerce purchasers

市场行为	交易过程	监测指标
串谋	交易中	申报电量市场份额均值
		报价一致性
		申报电量一致性
		报价曲线差异面积比率
		报价安全度均值
		报价相对比均值

指标体系中指标的产生都是通过通过分析购电商的报价策略,研究以往串谋行为监测指标的表现形式和数据特点,并对这些指标进行改进,将针对单个购电商的串谋监测指标进行扩展、改进,使之适应于 2 个购电商之间的串谋识别。在此基础上,文中充分研究了平行定价行为理论,创新性地提出一个新的串谋监测指标:报价曲线差异面积比率,以此建立起串谋监测指标体系,实现将任意 2 个市场主体联系起来进行监测的目的。改进后指标的具体计算方法如下。

(1) 申报电量市场份额均值:

$$x_{1ij} = (S_i + S_j) / \left( 2 \sum_{i=1}^n S_i \right) \times 100\% \quad (1)$$

式中:  $S_i$  为市场中第  $i$  个购电商这次竞价的申报电量;  $S_j$  为市场中第  $j$  个购电商这次竞价的申报电量;  $n$  为这次竞价中参与的购电商的个数。

申报电量市场份额均值越高的 2 个购电商操控市场价格的能力越强,串谋的可能性越高。

(2) 报价一致性:

$$x_{2ij} = \frac{\sum_{a=1}^3 (p_{ia} - p_{ja})^2}{\sum_{a=1}^3 (p_{ia} - \bar{p}_a)(p_{ja} - \bar{p}_a)} \quad (2)$$

式中:  $p_{ia}$ ,  $p_{ja}$  分别为第  $i$  个和第  $j$  个购电商在这次竞价中的第  $a$  段报价;  $\bar{p}_a$  为这次竞价中所有购电商第  $a$  段报价的均值。

当购电商出现一致性报价时,很有可能是购电

商串谋产生的结果。

(3) 申报量一致性:

$$x_{3ij} = \frac{\sum_{a=1}^3 (S_{ia} - S_{ja})^2}{\sum_{a=1}^3 (S_{ia} - \bar{S}_a)(S_{ja} - \bar{S}_a)} \quad (3)$$

式中:  $S_{ia}$ ,  $S_{ja}$  分别为第  $i$  个和第  $j$  个购电商在这次竞价中的第  $a$  段申报电量;  $\bar{S}_a$  为这次竞价中所有购电商第  $a$  段申报电量的均值。

当购电商出现申报电量一致性时,很有可能是购电商串谋产生的结果。

(4) 报价曲线差异面积比率:

$$x_{4ij} = \frac{\int_0^{S'} |f_i - f_j| ds}{S'} \quad (4)$$

式中:  $f_i$ ,  $f_j$  分别为第  $i$  个和第  $j$  个购电商在这次竞价中的报价曲线函数;  $S'$  为 2 个购电商申报电量中的较小值。

假设购电商 A 和购电商 B 同时参加了集中竞价,图 1 为购电商 A、B 的报价曲线,结合式(4),可知图 1 中阴影部分的面积即为式(4)的分子部分,也就是报价曲线的差异面积,而分母上的  $S'$  值可通过观察横坐标值得出。

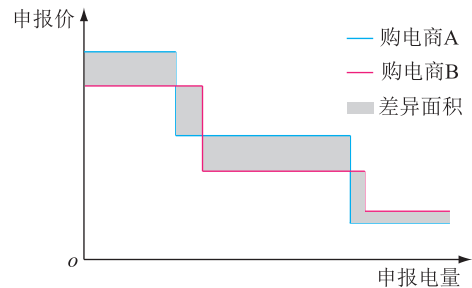


图 1 集中竞价的报价曲线图

Fig.1 Quotation curve of centralized bidding

报价曲线差异面积比率在结合电力市场集中竞价规则的基础上,综合考虑了购电商的报价策略。可以很好地将购电商在集中竞价过程中的分段式报价的价格和申报容量段通过报价曲线差异面积比率的方式进行结合,使之可以综合考虑购电商之间报价的相似程度。报价差异面积比率的值越小,说明报价的相似程度越高,那么串谋的可能性越大。

(5) 报价安全度均值:

$$x_{5ij} = (\bar{p}_i + \bar{p}_j - 2E) / 2E \times 100\% \quad (5)$$

式中:  $\bar{p}_i$ ,  $\bar{p}_j$  分别为第  $i$  个和第  $j$  个购电商在这次竞价中的加权平均申报价格;  $E$  为市场边际价格的期望值,可以用历史交易的边际价格计算得到。

报价安全度均值是衡量企业之间报价与历史出清价格的偏离程度的指标,如果偏离程度较大,说明串谋嫌疑较大。

(6) 报价相对比均值:

$$x_{6ij} = n(\bar{p}_i + \bar{p}_j) / \left( 2 \sum_{i=1}^n \bar{p}_i \right) \times 100\% \quad (6)$$

报价相对比均值反映了 2 个购电商的报价与这次竞价的平均价格的区别,如果该值远离 100%,则说明发生串谋的可能性较大。

通常,串谋行为不是通过上述指标体系中的某一个指标就能精确地监测出来,而是通过指标体系的综合考虑得出结论。

## 2 AdaBoost-DT 集成学习算法

### 2.1 AdaBoost

AdaBoost 算法是一种经典的集成分类算法,通过对同一个训练样本训练若干弱分类器,基于分类器的错误率分配权重参数,最后通过累加加权的方式将这些弱分类器组合成一个强分类器,从而达到较好的分类效果。

AdaBoost 算法的基本思想为:开始训练数据的每一个样本时,对每一个样本都赋予一个相同的权重,根据此样本训练出一个弱分类器,并且计算该分类器的错误率,在下次迭代训练时,根据上一次分类器的错误率,调整训练样本的权重,降低分类错误的样本权重,增大分类正确的样本权重,基于新的权重训练得到新的一个弱分类器。当通过  $L$  次迭代或者训练错误率  $e$  小于误差阈值  $\varepsilon$  时,停止迭代,此时产生  $L$  个弱分类器,根据这些弱分类器的错误率,给每个分类器分配不同的权重,错误率高的分类器权重系数更低,错误率低的分类器权重系数更高,最后对  $L$  个弱分类器按照权重系数进行线性组合,构成一个训练效果更好的强分类器。

AdaBoost 算法对基分类器的要求不高,而且可以根据迭代过程中分类错误率不断调整训练样本的权重,使得最终组合成的强分类器在分类效果上具备较高的准确率。

### 2.2 AdaBoost-DT 算法介绍

考虑到文中分类问题实质上是二分类问题,综合考虑计算复杂度和分类准确率的因素,采用决策树作为 AdaBoost 集成分类算法的基分类器<sup>[18-19]</sup>。在集成算法每次的迭代过程中产生一个决策树分类器,最后根据决策树的权重分配将这些基分类器组合成一个强分类器,实现对数据的预测分类。

在 AdaBoost-DT 算法中,迭代次数  $L$  会随着样

本集的规模而有所变化,文中通过交叉验证的方式进行确定。交叉验证是指在得到样本数据集之后,将数据集分为 2 个部分:训练集和验证集,通过设置分类精度,观察不同迭代次数下分类模型在验证集上的分类效果,从中选择分类效果较好的迭代次数的分类模型,确定迭代次数  $L$ 。

图 2 为 AdaBoost-DT 集成分类算法流程,具体步骤如下。

(1) 初始化输入数据集和样本权重:假设数据集的样本个数为  $N$ ,将每个样本的初始权重设置为  $1/N$ ;

(2) 用训练样本集训练得到基分类器  $G_l$ ,计算得到  $G_l$  的训练误差率,误差率为所有分类错误的样本的权值之和;

(3) 根据  $G_l$  的误差率,得到  $G_l$  的权重;

(4) 结合误差率,更新训练集中样本的权重,作为下一次迭代的样本权重;

(5) 将基分类器线性组合成强分类器  $G$ ;

(6) 根据总体误差率或者迭代次数判断迭代是否终止:若误差率  $e < \varepsilon$  或者迭代次数  $l > L$ ,则迭代终止,得到最终的强分类器  $G$ ;否则,  $l = l + 1$ ,返回第(2)步。

迭代结束得到的集成分类器  $G$  即为最终的串谋识别模型。

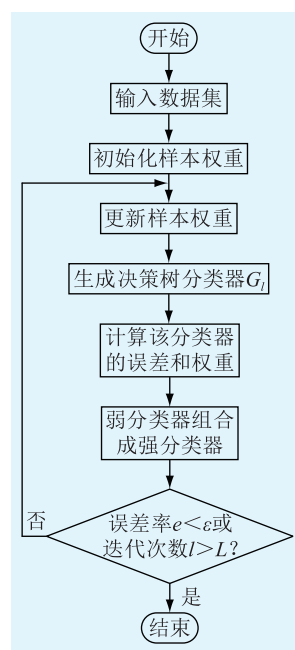


图 2 算法流程

Fig.2 Flow chart of algorithm

## 3 串谋行为智能识别方法

### 3.1 样本数据集

为充分发掘集中竞价过程中报价数据之间的

联系,实现串谋行为的机理分析和数据信息的有机结合,文中以串谋识别的指标体系作为 AdaBoost-DT 算法的样本数据集,以提高智能识别模型的准确度,降低模型的训练复杂度。

在电力市场的集中竞价过程中,利用购电商的历史报价数据和出清电价数据,计算串谋识别指标体系中的指标数据,以此作为智能识别模型的数据集。对于针对购电商的串谋识别,样本数据集的构成应为串谋识别指标体系数据:申报电量市场份额均值  $x_{1ij}$ , 报价一致性  $x_{2ij}$ , 申报电量一致性  $x_{3ij}$ , 报价曲线差异面积比率  $x_{4ij}$ , 报价安全度均值  $x_{5ij}$ , 报价相对比均值  $x_{6ij}$  以及串谋标签  $\lambda_{ij}$ 。

故第  $i$  个和第  $j$  个购电商的数据集序列为:

$$A_{ij} = \{x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij}, x_{5ij}, x_{6ij}, \lambda_{ij}\} \quad (7)$$

式中:  $\lambda_{ij}$  为  $i, j$  2 个购电商的串谋标签,  $\lambda_{ij} \in \{-1, 1\}$ ,  $-1$  为未串谋,  $1$  为串谋。

### 3.2 数据预处理

#### 3.2.1 数据增广

通过购电商的历史报价数据和出清电价,计算得到的购电商串谋识别数据集为:  $A_{ij} = \{x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij}, x_{5ij}, x_{6ij}, \lambda_{ij}\}$ 。

在实际数据集中,关于串谋的数据样本较少,属于典型的数据不均衡问题,可能导致分类器学习的串谋数据较少,造成分类器分类效果不好。所以,文中对训练数据采用简单过采样的方法来增加小类的数据样本个数<sup>[20-23]</sup>,即对样本数据集中的串谋样本进行随机抽样,通过衍生的方式对抽样串谋样本进行增广,使得最终的训练样本中不同类别的样本数量基本一致,进而提高分类器的准确度。

#### 3.2.2 数据归一化处理

由于训练的数据在不同维度上的量纲差距较大,为了提高分类器的训练效率,需要对每一个维度上的所有样本数据进行归一化处理,使得归一化后的数据处于同一尺度,降低量纲差距较大对分类效果的影响。归一化公式如下:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (8)$$

式中:  $\max(x)$ ,  $\min(x)$  分别为同一维度上样本数据的最大值和最小值。

### 3.3 智能识别方法

以往的购电商串谋行为的识别方法主要是人为判断和定性分析,在识别的实时性和效率上会有一定的延时。文中在平行定价行为理论的基础上,将设计的串谋行为识别指标体系和 AdaBoost-DT 分类算法结合,构建购电商串谋行为的智能识别方

法。该方法在获得购电商中长期或者实时报价的情况下,即可完成对串谋行为的智能识别。该智能识别方法获得报价数据后,通过计算得到串谋识别指标体系的数据集,应用 AdaBoost-DT 分类算法,训练数据集得到串谋行为智能识别模型。训练后的模型可对任意 2 个购电商的串谋行为进行识别。因为该方法建立在平行定价行为理论的指导下,所以得到的任意 2 个购电商串谋行为的识别结果具有传递性,即当超过 2 个以上的购电商形成串谋联盟时,其中任意 2 个购电商都将会被贴上串谋的标签。也就是说,如果若干个购电商之间任意 2 个被贴上串谋的标签,则认为该若干个购电商形成了串谋联盟,以此可以实现 2 个及以上购电商之间的串谋识别。具体步骤如下:

(1) 在得到购电商的报价数据之后,计算串谋识别指标体系中的数据,将其作为训练数据;

(2) 对步骤(1)中得到的训练数据集进行数据增广和归一化处理;

(3) 将预处理后的训练数据输入到分类器中进行训练,得到一个强分类器  $F(X)$ ;

(4) 对测试集的数据进行归一化处理,  $F(X)$  进行测试,计算分类器的分类准确率,根据结果对分类器的参数进行调整,最终形成一个分类效果较好的分类器;

(5) 步骤(4)中最后得到的分类器具备识别任意 2 个购电商之间串谋行为的功能,其输出结果为  $-1$  或  $1$ , 当输出结果为  $1$  时,说明 2 个市场主体串谋;

(6) 在得到任意 2 个购电商的串谋识别结果后,在平行定价行为理论的指导下,如果存在 2 个以上的购电商之中任意 2 个被识别为串谋行为,则认为这几家购电商之间形成了串谋联盟,进而实现 2 个及以上购电商之间的串谋识别。

## 4 算例分析

### 4.1 数据集的划分

针对上述串谋智能识别方法,以文献[7]中江西省电力直接交易中月度集中竞价交易为背景,结合其给出的 25 家购电商的具体交易数据,通过文中建立的指标体系,计算任意 2 个市场主体之间的指标数据,构成 AdaBoost-DT 识别算法的数据集,具体形式如式(7)所示。

对所产生的数据集进行随机排序,降低人为选择训练数据的影响。对数据集进行划分,其中 70% 的数据用于训练分类器,30% 的数据用于测试分类

器的识别准确度和效果,以此形成训练数据和测试数据。

### 4.2 识别结果

利用产生的数据集,通过交叉验证,对不同数量的弱分类器,即不同的迭代次数进行训练,得到分类准确率随迭代次数变化的曲线,即图3。

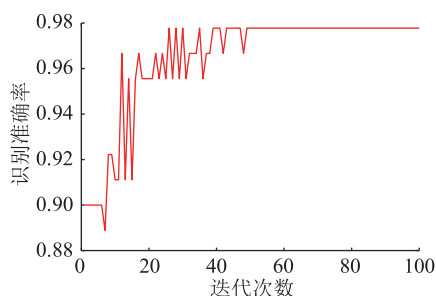


图3 分类效果随迭代次数变化曲线

Fig.3 Curve chart of classification effect changing with iteration times

由图3可以看出,分类识别的准确度随迭代次数,即基分类器个数的增加而增加,当迭代次数达到60之后,分类识别的准确度趋于平稳,以此确定迭代次数 $L$ 。不同迭代次数下的运行时间如图4所示。

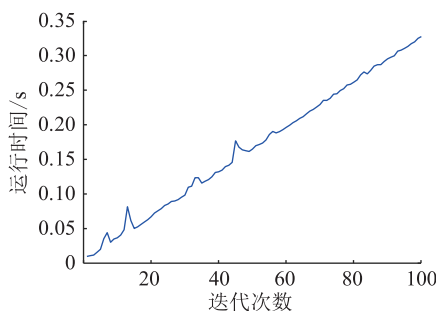


图4 运行时间随迭代次数变化曲线

Fig.4 Running time changing with iteration times

由图4可以看出,运行时间随迭代次数的增加呈上升趋势,但在得到样本数据集之后,分类识别的运行时间整体上均较短。相较于人工判断的方式,具有较好的实时性。

通过图3和图4,综合考虑该算例中的时间成本和分类准确率,将迭代次数 $L$ 定为60。通过样本数据集,对迭代60次的智能识别分类器进行训练,并对分类器的识别准确度进行测算,运行结果如图5所示。

由图5可以看出,在90个测试样本中,只有1个样本被分类错误,训练的集成分类器对测试数据的分类准确率为98.889%,具有较好的训练效果,训练准确度较高。

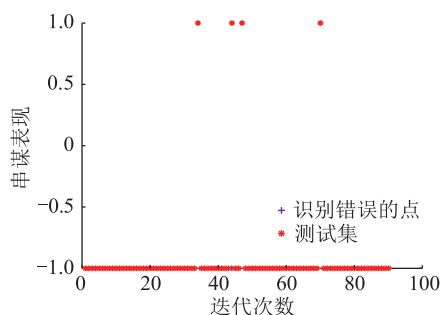


图5 测试集分类效果

Fig.5 Chart of test set classification effect

### 4.3 利用接收者操作特性曲线和接收者操作特性曲线下的面积值评价模型

在分类器模型的评价方法中,接收者操作特性曲线(receiver operating characteristic curve, ROC)的形状及接收者操作特性曲线下的面积值(area under curve, AUC)通常用作评价模型分类效果优劣的指标。如果ROC曲线越靠近坐标轴的左上角,表明该模型的识别效果越好;而AUC值指的是ROC曲线下的面积,取值范围为(0.5,1.0)。在AUC值大于0.5的情况下,如果AUC的值越接近于1,则说明分类效果越好。算例测算中ROC曲线和AUC值见图6。

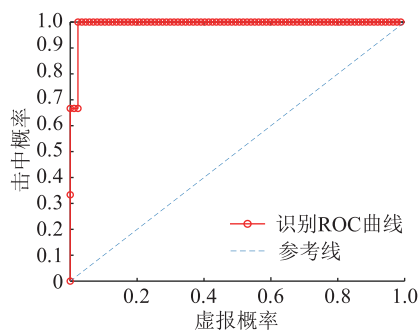


图6 ROC曲线

Fig.6 ROC diagram

图6中红色线为该次模型测算中的ROC曲线。可以看出,红色点集中在虚线的左上角,且非常接近图窗的左上角,可见模型分类效果较好。而AUC值为0.9738,非常接近1,同样说明该申谋行为识别模型在测算中的识别效果较为精确。

## 5 结论

文中从集中竞价过程的报价数据出发,给出一套改进后的基于任意2个电商之间的申谋识别指标体系,运用AdaBoost-DT集成分类算法,构建了基于该算法的购电商申谋行为识别方法,可以对2个及以上购电商的申谋行为进行识别。结论如下:

(1) 提出的报价曲线差异面积比率指标是基于平行定价行为理论的。该指标只适用于对购电侧或者售电侧同种类型的市场主体进行串谋行为识别,其综合考虑了报价价格和申报容量段因素,可以视为对报价一致性和申报量一致性的有力补充,能够较好地反映报价的相似性;

(2) 国内电力市场的串谋行为识别方法以指标法和定性分析居多,存在主观因素影响较大、效率不高的问题,文中方法主要用于交易的集中竞价过程,通过以往交易的报价信息对模型训练,在之后的交易过程中只要给出报价信息即可进行串谋行为的识别,其实时性,识别准确度、可操作性等较好;

(3) 文中方法利用月度竞价数据进行测算得到的串谋识别结果和实际的结果基本一致,验证了方法的有效性。在实际应用中,如果要判断 2 个以上购电商的串谋行为,需要根据智能识别结果,结合平行定价行为的传递性,实现串谋识别。

文中指标体系数据集可以根据电力市场的发展不断更新,以便提升识别效果,但主要针对购电商之间的串谋识别,并不适用于识别发电侧和售电侧之间的串谋识别,针对这一点仍需更深入的研究。

#### 参考文献:

- [1] 中共中央国务院. 关于进一步深化电力体制改革的若干意见(中发[2015]9号)[EB/OL]. [2018-05-15]. [http://tgs.ndrc.gov.cn/zyw/201601/t20120160\\_773852.html](http://tgs.ndrc.gov.cn/zyw/201601/t20120160_773852.html).  
State Council of the Central Committee of the Communist Party of China. Some opinions on further deepening the reform of the electric power system (No. 9 of China [2015]) [EB/OL]. [2018-05-15]. [http://tgs.ndrc.gov.cn/zyw/201601/t20120160\\_773852.html](http://tgs.ndrc.gov.cn/zyw/201601/t20120160_773852.html).
- [2] 王洪彬. 电力市场交易行为监管分析及监管系统设计[D]. 北京:北京交通大学,2008.  
WANG Hongbin. Analysis of electricity market transaction behavior supervision and design of supervision system[D]. Beijing: Beijing Jiaotong University, 2008.
- [3] 梁超. 区域电力市场监测与电力稽查的研究[D]. 北京:华北电力大学,2009.  
LIANG Chao. Research on regional electricity market monitoring and electricity inspection[D]. Beijing: North China Electric Power University (Beijing), 2009.
- [4] 王文举,范合君. 企业价格串谋识别的博弈分析及模拟[J]. 商业研究,2010(5):49-52.  
WANG Wenju, FAN Hejun. Game analysis and simulation of enterprise price collusion identification[J]. Business Research, 2010(5):49-52.
- [5] 史述红,刘敦楠,胡会文,等. 市场开放下电力交易全过程违规行为识别探究[J]. 价格理论与实践,2018(8):51-54.  
SHI Shuhong, LIU Dunnan, HU Huiwen, et al. Research on the identification of violations in the whole process of power trading under the open market[J]. Price theory and practice, 2018(8):51-54.
- [6] 杨沫. 电力市场中交易违规行为建模分析及应用[D]. 北京:华北电力大学,2018.  
YANG Mo. Modeling, analysis and application of transaction violations in power market[D]. Beijing: North China Electric Power University, 2018.
- [7] 刘敦楠,李瑞庆,陈雪青,等. 电力市场监管指标及市场评价体系[J]. 电力系统自动化,2004,28(9):16-21.  
LIU Dunnan, LI Ruiqing, CHEN Xueqing, et al. Power market supervision index and market evaluation system[J]. Power system automation, 2004, 28(9):16-21.
- [8] 刘敦楠,张潜,李霄彤,等. 基于云模型与模糊 Petri 网的电力市场潜在危害行为识别[J]. 电力系统自动化,2019,43(2):25-37.  
LIU Dunnan, ZHANG Qian, LI Xiaotong, et al. Identification of potential hazards in power market based on cloud model and fuzzy Petri net[J]. Power system automation, 2019, 43(2):25-37.
- [9] 张粒子,张集,程瑜. 电力市场中的串谋溢价和串谋行为规制[J]. 电网技术,2006,30(24):61-67.  
ZHANG Lizi, ZHANG Ji, CHENG Yu. Regulation of collusion premium and collusion behavior in power market[J]. Power Grid Technology, 2006, 30(24):61-67.
- [10] 李伟,杨照芬,牛东晓. 基于改进神经网络的用电客户信用评价[J]. 华东电力,2007(4):9-12.  
LI Wei, YANG Zhaofen, NIU Dongxiao. Credit evaluation of electricity customers based on improved neural network[J]. East China Electric Power, 2007(4):9-12.
- [11] 谢林枫,钱立军,季聪,等. 基于长短期记忆网络算法的电费回收风险预警[J]. 电力工程技术,2018,37(5):98-103.  
XIE Linfeng, QIAN Lijun, JI Cong, et al. Risk early warning of electricity charge recovery based on long-term and short-term memory network algorithm[J]. Electric Power Engineering Technology, 2018, 37(5):98-103.
- [12] 陈冠宇. 基于 kNN-Smote-LSTM 的信用卡欺诈风险检测网络模型[D]. 杭州:浙江工商大学,2018.  
CHEN Guanyu. Credit card fraud risk detection network model based on kNN-Smote-LSTM[D]. Hangzhou: Zhejiang Industrial and Commercial University, 2018.
- [13] 张双. 基于集成学习的英文文本分类算法研究[D]. 吉林:吉林大学,2018.  
ZHANG Shuang. Research on English text categorization algorithm based on ensemble learning[D]. Jilin: Jilin University, 2018.
- [14] 杜芸芸. 基于数据挖掘的上市公司财务报告违规研究[D]. 成都:西南财经大学,2013.  
DU Yunyun. Research on financial reporting violations of listed companies based on data mining[D]. Chengdu: Southwest University of Finance and Economics, 2013.
- [15] ZHANG D, ABUASSBA A O, LUO X. A heterogeneous AdaBoost ensemble based extreme learning machines for imba-

- lanced data[J]. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 2019, 13(3):19-35.
- [16] WALKER K W,JIANG Zhehan. Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction;a machine-learning approach [J]. The Journal of Academic Librarianship, 2019, 45(3):203-212.
- [17] 张宗堂,王森,孙世林. 一种针对不平衡数据分类的集成学习算法[J]. 山东大学学报(工学版), 2019, 49(3):1-7.  
ZHANG Zongtang, WANG Sen, SUN Shilin. An integrated learning algorithm for unbalanced data classification[J]. Journal of Shandong University (Engineering Edition), 2019, 49(3):1-7.
- [18] 韩存鸽,叶球孙. 决策树分类算法中 C4.5 算法的研究与改进[J]. 计算机系统应用, 2019(6):198-202.  
HAN Cunge, YE Qiusun. Research and improvement of C4.5 algorithms in decision tree classification algorithms[J]. Computer System Applications, 2019(6):198-202.
- [19] 尹儒,门昌骞,王文剑. 一种模型决策森林算法[J/OL]. 计算机科学与探索;1-11(2019-06-03)[2019-06-23]. <http://kns.cnki.net/kcms/detail/11.5602.tp.20190531.1722.006.html>.  
YIN Ru, MEN Changqian, WANG Wenjian. A model decision forest algorithm[J/OL]. Computer Science and Exploration;1-11(2019-06-03)[2019-06-23]. <http://kns.cnki.net/kcms/detail/11.5602.tp.20190531.1722.006.html>.
- [20] 陈淑真,朱建平,尤添革,等. 基于 SMOTERF 算法处理客户流失不平衡问题的研究[J]. 数学的实践与认识, 2019, 49(9):204-210.  
CHEN Shuzhen, ZHU Jianping, YOU Tiange, et al. Research on imbalance of customer churn based on SMOTERF algorithm[J]. Practice and understanding of mathematics, 2019, 49(9):204-210.
- [21] 田臣,周丽娟. 基于带多数类权重的少数类过采样技术和随机森林的信用评估方法[J/OL]. 计算机应用;1-8(2019-01-31)[2019-06-23]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20190131.0953.018.html>.  
TIAN Chen, ZHOU Lijuan. Based on the minority over sampling technique with the weight of most classes and the credit evaluation method of random forests[J/OL]. Computer application;1-8(2019-01-31)[2019-06-23]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20190131.0953.018.html>.
- [22] 王凯,王健,刘刚,等. 基于 RetinaNet 和类别平衡采样方法的销钉缺陷检测[J]. 电力工程技术, 2019, 38(4):80-85.  
WANG Kai, WANG Jian, LIU Gang, et al. Pin defect detection based on RetinaNet and category balance sampling method[J]. Electric Power Engineering Technology, 2019, 38(4):80-85.
- [23] 肖连杰,郜梦蕊,苏新宁. 一种基于模糊 C-均值聚类的欠采样集成不平衡数据分类算法[J]. 数据分析与知识发现, 2019, 3(4):90-96.  
XIAO Lianjie, GAO Mengrui, SU Xinning. An under sampling integrated unbalanced data classification algorithm based on fuzzy C-means clustering[J]. Data Analysis and Knowledge Discovery, 2019, 3(4):90-96.

作者简介:



张海生

张海生(1989),男,硕士,工程师,从事电力市场、接入系统设计技术相关工作(E-mail: 869699519@qq.com);

曹喆(1988),男,硕士,工程师,从事电力市场、电力工程技术相关工作;

杨昌海(1982),男,硕士,高级工程师,从事能源经济研究和电网规划相关工作。

### Collusive behavior recognition in electricity market based on AdaBoost-DT algorithm

ZHANG Haisheng<sup>1</sup>, CAO Zhe<sup>1</sup>, YANG Changhai<sup>1</sup>, LUO Yunpeng<sup>2</sup>, HUA Huichun<sup>2</sup>

(1. State Grid Gansu Electric Power Company Economics & Technology Research Institute, Lanzhou 730050, China; 2. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources(North China Electric Power University), Baoding 071003, China)

**Abstract:** In order to solve the problem of qualitative analysis and low real-time performance of collusion identification methods in power market, this paper proposes an intelligent identification method of collusion behavior based on AdaBoost-DT algorithm, which uses AdaBoost-DT integrated classification algorithm to identify collusion behavior, and solves the problem that collusion behavior is difficult to identify quantitatively. Firstly, based on the mechanism of collusion, a set of collusion identification index system is designed. In the face of the problem of data imbalance, the oversampling method is used to expand the training data set, and the AdaBoost-DT classification algorithm is used to train the collusion behavior intelligent identification model. Finally, based on the monthly transaction data, an example is analyzed, and the receiver operating characteristic curve (ROC curve) and the area under the receiver operating characteristic curve (AUC value) are used to evaluate the recognition effect of the model. The experimental results show that the proposed method has good accuracy and real-time performance, which fully verifies the effectiveness of the algorithm.

**Keywords:** electricity market; collusion; oversampling; decision tree; AdaBoost-DT

(编辑 方晶)