



(12)发明专利

(10)授权公告号 CN 106054104 B

(45)授权公告日 2019.01.11

(21)申请号 201610342042.2

(22)申请日 2016.05.20

(65)同一申请的已公布的文献号
申请公布号 CN 106054104 A

(43)申请公布日 2016.10.26

(73)专利权人 国网新疆电力公司电力科学研究院

地址 830000 新疆维吾尔自治区乌鲁木齐市高新技术产业开发区(新市区)长春中路东二巷66号

专利权人 新疆大学

(72)发明人 李宁 袁铁江 杨金成 蒋平
王刚 董小顺 罗庆 李国军
薛飞 段志尚 山宪武

(74)专利代理机构 北京科迪生专利代理有限公司 11251

代理人 关玲

(51)Int.Cl.
G01R 35/04(2006.01)

(56)对比文件
CN 104021264 A,2014.09.03,
CN 103532745 A,2014.01.22,
CN 103532745 A,2014.01.22,
CN 105182065 A,2015.12.23,
CN 104915898 A,2015.09.16,
CN 104794195 A,2015.07.22,
CN 103559655 A,2014.02.05,
CN 104111920 A,2014.10.22,
US 2015/0112903 A1,2015.04.23,

审查员 赵娟娟

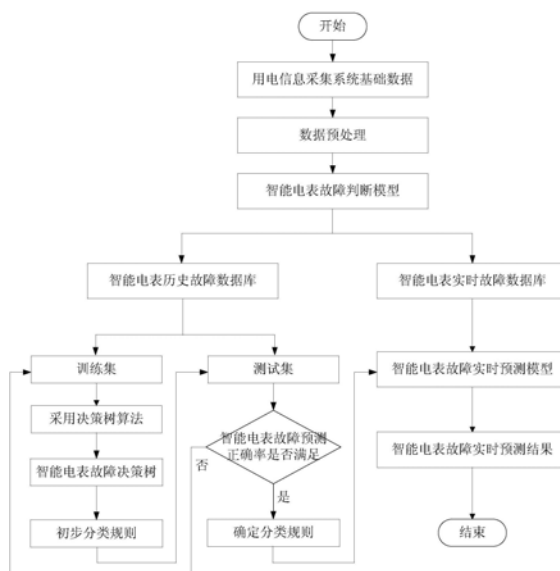
权利要求书3页 说明书9页 附图1页

(54)发明名称

一种基于决策树的智能电表故障实时预测方法

(57)摘要

一种基于决策树的智能电表故障实时预测方法,步骤为:1、对用电信息采集系统的智能电表数据预处理;2、根据智能电表故障判断模型,筛选用电信息采集系统的智能电表有故障的数据至智能电表故障数据库中;3、将智能电表故障数据库中的历史数据分为训练集和测试集,采用决策树算法对训练集进行数据挖掘,形成智能电表故障决策树和初步分类规则;4、通过测试集的数据对初步分类规则进行正确率评估,若正确率满足要求,则确定分类规则,若不满足要求,则返回训练集,重新训练;5、由最终确定的分类规则生成智能电表故障实时预测模型;6、将智能电表实时故障数据库链接至智能电表故障实时预测模型进行实时预测,得到智能电表故障实时预测结果。



CN 106054104 B

1. 一种基于决策树的智能电表故障实时预测方法,其特征在于,所述的实时预测方法的具体步骤如下:

步骤1,对用电信息采集系统的智能电表数据进行数据预处理;

步骤2,根据智能电表故障判断模型,将用电信息采集系统的智能电表有故障的数据筛选至智能电表故障数据库中;

所述智能电表故障判断模型如下:

(1) 当进行电表总电量与各费率电量之和不等的故障数据筛选时,采用下式判断电表总电量与各费率电量之和不等的故障:

$$\left\{ \begin{array}{l} W_{\pm i} > 0 \\ \sum_{i=1}^{e+1} W_{\pm i} > 0 \\ \left| W_{\pm 1} - \sum_{i=2}^{e+1} W_{\pm i} \right| > \rho \end{array} \right. \quad (1)$$

式中:W代表电量,+代表电量的方向为正向,-代表电量的方向为反向,i=1代表总时段,W₊₁代表正向总电量,W₋₁代表反向总电量,i=2代表高峰时段,i=3代表平时段,i=4代表低谷时段,i=5代表尖峰时段,e为电表费率数,ρ为总电量与各费率电量之和不等的判断因子;当e=4时,表示电表为四费率电表,取ρ=0.4;当e=3时,表示电表为三费率电表,取ρ=0.3;当e=2时,表示电表为二费率电表,取ρ=0.2;

所述的四费率电表指支持高峰时段、平时段、低谷时段、尖峰时段4个时段计费的电表;三费率电表指支持仅高峰时段、平时段、低谷时段3个时段计费的电表;二费率电表指仅支持平时段、低谷时段2个时段计费的电表;

判断原则如下:

①电表的正反向总电量、正反向高峰时段电量、正反向平时段电量、正反向低谷时段电量均大于0,且不为空;

②电表正反向各费率电量之和大于0;

③电表总电量与各费率电量之和差值的绝对值大于某个阈值,阈值规则如下:如果是四费率电表,按0.4判断,三费率电表按0.3判断,二费率电表按0.2判断;

④若同时满足判断原则①②③,则判断为严重故障;

(2) 当进行电能表飞走和突变故障数据筛选时,判断电能表飞走和突变的故障程序如下:

a. 首先计算当天最大电量W_f:

$$W_f = \begin{cases} \frac{220 \times I_{\max} \times 12}{1000} & \text{月份} = 1, 2, 7, 8, 9, 12 \\ \frac{220 \times 3 \times I_b \times 8}{1000} & \text{月份} = 3, 4, 5, 6, 10, 11 \end{cases} \quad (2)$$

式中:W_f为当天最大电量;I_{max}为最大电流;I_b为基本电流;

b. 然后计算电表飞走和突变因子K:

$$K = \frac{W_f}{W_t} \quad (3)$$

式中:K为电表飞走和突变因子, W_t 为当日电量;

判断原则如下:

①针对居民表,在1、2、7、8、9、12月份取最大电流 I_{\max} ,按12小时计算电表飞走和突变因子K;在3、4、5、6、10、11月份取3倍基本电流 I_b ,按8小时计算电表飞走和突变因子K;

②当进行电能表飞走和突变故障数据筛选时,采用公式(4)判断电能表飞走和突变故障:

$$K \geq 1 \quad (4)$$

③若满足判据②,则判断为严重故障;

(3)当进行电表反向有功示值大于零故障数据筛选时,采用下式判断电表反向有功示值大于零的故障:

$$\begin{cases} P_{-总} > 0 \\ Q_{-总} > 0 \end{cases} \quad (5)$$

式中:P代表有功功率,Q代表无功功率, $P_{-总}$ 代表电表反向有功总功率, $Q_{-总}$ 代表电表反向无功总功率;

判断原则如下:

①电能表出现反向有功总功率或反向无功总功率大于0;

②若满足判据①,则判断为严重故障;

(4)当进行电能表倒走故障数据筛选时,电能表倒走判断的前提为排除抄读电表正反向总电量总为空的记录,采用下式判断电能表倒走的故障:

$$\begin{cases} W_{+1} < W_{+1y} \\ W_{-1} < W_{-1y} \end{cases} \quad (6)$$

式中: W_{+1y} 为前一天的正向总电量, W_{-1y} 为前一天的反向总电量;

判断原则如下:

根据日电量统计表判断,针对低压居民和单相工商业户,只判断前一天正向总电量 W_{+1y} 和反向总电量 W_{-1y} 是否大于当天的抄表示值,即当天正向总电量 W_{+1} 和反向总电量 W_{-1} ,如果成立,则判断为严重故障;

(5)当进行电表时钟不对故障数据筛选时,电表时钟不对的故障判断程序如下:

首先判断对时错误次数m,如果对时错误次数大于3次,则直接判定为严重故障,即:

$$m > 3 \text{ 严重} \quad (7)$$

式中:m为对时错误次数;

若m不大于3,由在线监测查询是否满足 Δt 的要求,并按 Δt 的标准生成电能表时钟超差等级,则采用公式(8)判断:

$$\begin{cases} 5 \leq \Delta t < 15 & \text{一般} \\ 15 \leq \Delta t < 30 & \text{重要} \\ \Delta t > 30 & \text{严重} \end{cases} \quad (8)$$

式中: Δt 为终端与电能表时钟之差,计算方法见式(9):

$$\Delta t = |t_{\text{终端}} - t_{\text{电表}}| \quad (9)$$

式中： $t_{\text{终端}}$ 表示终端时钟， $t_{\text{电表}}$ 表示电能表时钟；

判断原则如下：

- ①如果对时错误次数 m 超过3次，直接判断为严重故障；
- ②如终端与电能表时钟之差 Δt 满足 $5\text{min} \leq \Delta t < 15\text{min}$ ，判断为一般故障；若 $15\text{min} \leq \Delta t < 30\text{min}$ 判断为重要故障；若 $\Delta t > 30\text{min}$ 判断为严重故障；

(6) 当进行电表电能费率设置异常故障数据筛选时，采用下式判断电表电能费率设置异常的故障：

$$W_{+5} \neq 0 \text{ 或 } W_{-5} \neq 0 \quad (10)$$

式中， W_{+5} 为尖峰时段的正向电量， W_{-5} 为尖峰时段的反向电量；

判断原则如下：

- ①只判断DLT-2007规约的电能表；
- ②判断是否存在尖峰时段的正向电量 W_{+5} 或尖峰时段的反向电量 W_{-5} ，若存在，则判断为严重故障；

(7) 当进行电能表正向潜动故障数据筛选时，采用下式判断电能表正向潜动的故障：

$$0.98 \times W_{+3} \leq \frac{14}{24} \times W_{+1} \leq 1.02 \times W_{+3} \text{ 或 } 0.98 \times W_{+4} \leq \frac{10}{24} \times W_{+1} \leq 1.02 \times W_{+4} \quad (11)$$

式中， W_{+3} 为平时时段的正向电量， W_{+4} 为低谷时段的正向电量， W_{+1} 为正向总电量；

判断原则如下：

若满足公式(11) 3次，则判断为重要故障，若满足公式(11) 5次，则判断为严重故障；

(8) 当进行电能表反向潜动故障数据筛选时，采用下式判断电能表反向潜动的故障：

$$\begin{cases} W_{+1} > 0.1 \\ W_{-1} > 0.1 \end{cases} \quad (12)$$

式中， W_{+1} 代表正向总电量， W_{-1} 代表反向总电量；

判断原则如下：

针对结算类的电能表正向总电量和反向总电量同时存在并大于0.1的情况，则判断为严重故障；

步骤3，选取智能电表故障数据库中的历史数据，将其分为训练集和测试集，采用决策树算法对训练集进行数据挖掘形成智能电表故障决策树，然后形成初步分类规则；

步骤4，通过测试集的数据对初步分类规则进行正确率评估，若正确率满足要求，则确定分类规则，若正确率不满足要求，则返回至训练集，重新进行训练；

步骤5，由最终确定的分类规则生成智能电表故障实时预测模型；

步骤6，将智能电表实时故障数据库链接至智能电表故障实时预测模型进行实时预测，得到智能电表故障实时预测结果。

2. 根据权利要求1所述的基于决策树的智能电表故障实时预测方法，其特征在于，所述的步骤1对用电信息采集系统的智能电表数据进行数据预处理的步骤为：首先去除与智能电表故障无关的属性数据、有明显错误的数据和属性重复的数据，然后将各个属性连续数据离散化。

一种基于决策树的智能电表故障实时预测方法

技术领域

[0001] 本发明涉及一种智能电表故障预测方法,特别涉及一种基于决策树的智能电表故障实时预测方法。

背景技术

[0002] 随着国家电网公司“三集五大”体系建设的深入推进,“大营销”推广的用电信息采集系统建设已具规模,特别是智能电能表已获得大范围的推广应用。如此庞大数量的智能电表一旦出现故障,直接影响到用户的切身利益和社会的安全稳定,因此及时预测出智能电表故障并对其维修或更换是十分重要的。

[0003] 随着智能电表的广泛应用,运行故障越来越多地呈现出突发性、多面性、复杂性的特征,传统电表的故障预测方法已不再适用。目前国内外学者主要研究的神经网络、灰色预测和专家系统等预测方法已在各领域得到了应用,并取得了良好的效果,但仍存在不足之处。用电信息采集系统提供的海量数据恰好给了数据挖掘技术一个用武之地,而在智能电表故障预测领域,数据挖掘技术方面的相关研究尚未深入展开。

发明内容

[0004] 本发明的目的是针对上述问题,提出一种基于决策树的智能电表故障实时预测方法,以实现在预测正确率可靠的情况下,对智能电表故障进行实时预测。

[0005] 为达到上述目的,本发明采用技术方案是:

[0006] 步骤1,对用电信息采集系统的智能电表数据进行数据预处理;

[0007] 步骤2,根据智能电表故障判断模型,将用电信息采集系统的智能电表有故障的数据筛选至智能电表故障数据库中;

[0008] 所述的智能电表故障判断模型的作用是判断智能电表是否发生故障。

[0009] 步骤3,选取智能电表故障数据库中的历史数据,将其分为训练集和测试集,采用决策树算法对训练集进行数据挖掘,形成智能电表故障决策树,然后形成初步分类规则;

[0010] 步骤4,通过测试集的数据对初步分类规则进行正确率评估,若正确率满足要求,则确定分类规则,若正确率不满足要求,则返回至训练集,重新进行训练;

[0011] 步骤5,由最终确定的分类规则生成智能电表故障实时预测模型;

[0012] 步骤6,将智能电表实时故障数据库链接至智能电表故障实时预测模型进行实时预测,得到智能电表故障实时预测结果。

[0013] 所述的步骤1对用电信息采集系统的智能电表数据进行数据预处理的步骤为,首先去除与智能电表故障无关的属性数据、有明显错误的数据和属性重复的数据,然后将各个属性连续数据离散化。

[0014] 所述的步骤2根据智能电表故障判断模型将用电信息采集系统的智能电表有故障的数据筛选至智能电表故障数据库中的方法如下:

[0015] (1) 当进行电表总电量与各费率电量之和不等故障数据筛选时,采用下式判断

电表总电量与各费率电量之和不等故障:

$$[0016] \quad \left\{ \begin{array}{l} W_{\pm i} > 0 \\ \sum_{i=1}^{e+1} W_{\pm i} > 0 \\ \left| W_{\pm 1} - \sum_{i=2}^{e+1} W_{\pm i} \right| > \rho \end{array} \right. \quad (1)$$

[0017] 式中:W代表电量,+代表电量的方向为正向,-代表电量的方向为反向,i=1代表总时段,W₊₁代表正向总电量,W₋₁代表反向总电量,i=2代表高峰时段,i=3代表平时段,i=4代表低谷时段,i=5代表尖峰时段,e为电表费率数,ρ为总电量与各费率电量之和不等判断因子。当e=4时,表示电表为四费率电表,取ρ=0.4;当e=3时,表示电表为三费率电表,取ρ=0.3;当e=2时,表示电表为二费率电表,取ρ=0.2;

[0018] 所述的四费率电表指支持高峰时段、平时段、低谷时段、尖峰时段4个时段计费的电表;三费率电表指支持仅高峰时段、平时段、低谷时段3个时段计费的电表;二费率电表指仅支持平时段、低谷时段2个时段计费的电表;

[0019] 判断原则如下:

[0020] ①电表的正反向总电量、正反向高峰时段电量、正反向平时段电量、正反向低谷时段电量均大于0,且不为空;

[0021] ②电表正反向各费率电量之和大于0;

[0022] ③电表总电量与各费率电量之和差值的绝对值大于某个阈值,阈值规则如下:如果是四费率电表,按0.4判断,三费率电表按0.3判断,二费率电表按0.2判断;

[0023] ④若同时满足判断原则①②③,则判断为严重故障;

[0024] (2)当进行电能表飞走和突变故障数据筛选时,判断电能表飞走和突变的故障程序如下:

[0025] a.首先计算当天最大电量W_f:

$$[0026] \quad W_f = \begin{cases} \frac{220 \times I_{max} \times 12}{1000} & \text{月份} = 1,2,7,8,9,12 \\ \frac{220 \times 3 \times I_b \times 8}{1000} & \text{月份} = 3,4,5,6,10,11 \end{cases} \quad (2)$$

[0027] 式中:W_f为当天最大电量;I_{max}为最大电流;I_b为基本电流;

[0028] b.然后计算电表飞走和突变因子K:

$$[0029] \quad K = \frac{W_f}{W_t} \quad (3)$$

[0030] 式中:K为电表飞走和突变因子,W_t为当日电量;

[0031] 判断原则如下:

[0032] ①针对居民表,在1、2、7、8、9、12月份取最大电流I_{max},按12小时计算电表飞走和突变因子K;在3、4、5、6、10、11月份取3倍基本电流I_b,按8小时计算电表飞走和突变因子K;

[0033] ③当进行电能表飞走和突变故障数据筛选时,采用公式(4)判断电能表飞走和突变故障:

[0034] $K \geq 1$ (4)

[0035] ③若满足判据②,则判断为严重故障;

[0036] (3)当进行电表反向有功示值大于零故障数据筛选时,采用下式判断电表反向有功示值大于零的故障:

$$[0037] \begin{cases} P_{-总} > 0 \\ Q_{-总} > 0 \end{cases} \quad (5)$$

[0038] 式中:P代表有功功率,Q代表无功功率, $P_{-总}$ 代表电表反向有功总功率, $Q_{-总}$ 代表电表反向无功总功率;

[0039] 判断原则如下:

[0040] ①电能表出现反向有功总功率或反向无功总功率大于0;

[0041] ②若满足判据①,则判断为严重故障;

[0042] (4)当进行电能表倒走故障数据筛选时,电能表倒走判断的前提为排除抄读电表正反总电量总为空的记录,采用下式判断电能表倒走的故障:

$$[0043] \begin{cases} W_{+1} < W_{+1y} \\ W_{-1} < W_{-1y} \end{cases} \quad (6)$$

[0044] 式中: W_{+1y} 为前一天的正向总电量, W_{-1y} 为前一天的反向总电量;

[0045] 判断原则如下:

[0046] 根据日电量统计表判断,针对低压居民和单相工商业户,只判断前一天正向总电量和反向总电量是否大于当天的抄表示值,如果成立,则判断为严重故障;

[0047] (5)当进行电表时钟不对故障数据筛选时,电表时钟不对的故障判断程序如下:

[0048] 首先判断对时错误次数m,如果对时错误次数大于3次,则直接判定为严重故障,即:

$$[0049] m > 3 \text{ 严重} \quad (7)$$

[0050] 式中:m为对时错误次数;

[0051] 若m不大于3,由在线监测查询是否满足 Δt 的要求,并按 Δt 的标准生成电能表时钟超差等级,则采用公式(8)判断:

$$[0052] \begin{cases} 5 \leq \Delta t < 15 & \text{一般} \\ 15 \leq \Delta t < 30 & \text{重要} \\ \Delta t > 30 & \text{严重} \end{cases} \quad (8)$$

[0053] 式中: Δt 为终端与电能表时钟之差,计算方法见式(9):

$$[0054] \Delta t = |t_{\text{终端}} - t_{\text{电表}}| \quad (9)$$

[0055] 式中: $t_{\text{终端}}$ 表示终端时钟, $t_{\text{电表}}$ 表示电能表时钟;

[0056] 判断原则如下:

[0057] ①如果对时错误次数m超过3次,直接判断为严重故障;

[0058] ②如终端与电能表时钟之差 Δt 满足 $5\text{min} \leq \Delta t < 15\text{min}$,判断为一般故障;

[0059] 若 $15\text{min} \leq \Delta t < 30\text{min}$ 判断为重要故障;若 $\Delta t > 30\text{min}$ 判断为严重故障;

[0060] (6)当进行电表电能费率设置异常故障数据筛选时,采用下式判断电表电能费率设置异常的故障:

[0061] $W_{+5} \neq 0$ 或 $W_{-5} \neq 0$ (10)

[0062] 判断原则如下:

[0063] ①只判断DLT-2007规约的电表;

[0064] ②判断是否存在正向尖峰时段电量或反向尖峰时段电量,若存在,则判断为严重故障;

[0065] (7)当进行电表正向潜动故障数据筛选时,采用下式判断电表正向潜动的故障:

$$[0066] \quad 0.98 \times W_{+3} \leq \frac{14}{24} \times W_{+1} \leq 1.02 \times W_{+3} \text{ 或 } 0.98 \times W_{+4} \leq \frac{10}{24} \times W_{+1} \leq 1.02 \times W_{+4} \quad (11)$$

[0067] 判断原则如下:

[0068] 若满足公式(11)3次,则判断为重要故障,若满足公式(11)5次,则判断为严重故障;

[0069] (8)当进行电表反向潜动故障数据筛选时,采用下式判断电表反向潜动的故障:

$$[0070] \quad \begin{cases} W_{+1} > 0.1 \\ W_{-1} > 0.1 \end{cases} \quad (12)$$

[0071] 判断原则如下:

[0072] 针对结算类的电表正向总电量和反向总电量同时存在并大于0.1的情况,则判断为严重故障。

[0073] 所述的步骤3选取智能电表故障数据库中的历史数据,将其分为训练集和测试集,采用决策树算法对训练集进行数据挖掘,形成智能电表故障决策树,然后形成初步分类规则,其步骤如下:

[0074] 设R是智能电表故障数据库的历史数据,并将R的60%数据分给训练集S,R的40%数据分给测试集T。以采用C5.0算法为例对训练集S进行数据挖掘。

[0075] 设具有n个值的属性X将S分成n个子集 S_1, S_2, \dots, S_n ,设S中样本的总数为 $|S|$, $\text{freq}(C_i, S)$ 是S中属于类别 C_i ($i=1, 2, \dots, N$)的样本个数,S中某个样本属于类别 C_i 的概率为 $\frac{\text{freq}(C_i, S)}{|S|}$,它所传达的信息为: $\frac{\log_2(\text{freq}(C_i, S))}{|S|}$;

[0076] 训练集S的熵 $\text{info}(S)$ 由式(13)表示;

$$[0077] \quad \text{info}(S) = \sum_{i=1}^N \frac{\text{freq}(C_i, S)}{|S|} \log_2 \left(\frac{\text{freq}(C_i, S)}{|S|} \right) \quad (13)$$

[0078] 将训练集S根据属性X分割成n个子集后,计算各个子集的信息熵,则集合S的期望信息 $\text{info}_X(S)$ 用公式(14)来表示;

$$[0079] \quad \text{info}_X(S) = \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{info}(S_i) \quad (14)$$

[0080] 为了度量按照属性检验X进行分区的S所得到的信息,使用增益标准 $\text{gain}(X)$,它选择使其值能最大化,即此标准选择的是具有最高信息增益的属性来进行每次分区;

$$[0081] \quad \text{gain}(X) = \text{info}(S) - \text{info}_X(S) \quad (15)$$

[0082] 根据属性X的n个不同取值将S分割成 S_1, S_2, \dots, S_n 共n个子集后产生的潜在信息 $split_info(X)$ 用式(16)所示:

$$[0083] \quad split_info(X) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (16)$$

[0084] 公式(17)为X对S进行划分信息增益率 $gain_ratio(X)$;

$$[0085] \quad gain_ratio(X) = \frac{gain(X)}{split_info(X)} \quad (17)$$

[0086] 首次计算采用选择最高信息增益率的属性作为决策树的根节点,然后决策树的每一个节点用此方法计算剩余属性的信息增益率,仍选择最高信息增益率的属性作为决策树的当前节点,直到最终形成整个决策树;

[0087] 由该决策树形成初步分类规则。

[0088] 所述的步骤4通过测试集的数据对初步分类规则进行正确率评估,若正确率满足要求,则确定分类规则,若正确率不满足要求,则返回至训练集,重新进行训练,其详细步骤如下:

[0089] (1) 正确率评估公式为:

$$[0090] \quad Z = \frac{A}{B} \times 100\% \quad (18)$$

[0091] 式中Z为评估正确率,A为初步分类规则在测试集T中预测正确的个数;B为测试集T的总数据个数;

[0092] (2) 设 Z' 为设定的正确率;若 $Z \geq Z'$,则确定分类规则;若 $Z < Z'$,则返回训练集重新训练决策树。

[0093] 本发明智能电表故障实时预测方法可准确地反映智能电表实时故障情况。

附图说明

[0094] 图1是本发明方法的总体实施流程框图。

具体实施方式

[0095] 下面结合附图及具体实施方式对本发明做进一步说明。

[0096] 下面以一批型号为DDZY102-Z智能电表在新疆某地区2年的运行数据统计为例对本发明作进一步说明。

[0097] 如图1所示,本发明基于决策树的智能电表故障实时预测方法包括以下步骤:

[0098] 步骤1,对用电信息采集系统的DDZY102-Z智能电表数据进行数据预处理,步骤如下:

[0099] (1) 去除与智能电表故障无关的属性数据、有明显错误的数据和属性重复的数据;

[0100] (2) 将各个属性连续数据离散化。

[0101] 步骤2,根据DDZY102-Z智能电表故障判断模型,将用电信息采集系统的智能电表有故障的数据筛选至智能电表故障数据库中,步骤如下:

[0102] 所述的智能电表故障判断模型的作用是判断智能电表是否发生故障。

[0103] DDZY102-Z智能电表故障判断模型如下:

[0104] (1) 当进行电表总电量与各费率电量之和不等的故障数据筛选时,因DDZY102-Z智能电表为四费率电表, $e=4$,总电量与各费率电量之和不等的判断因子 ρ 取0.4。采用下式判断电表总电量与各费率电量之和不等的故障:

$$[0105] \begin{cases} W_{\pm i} > 0 \\ \sum_{i=1}^5 W_{\pm i} > 0 \\ \left| W_{\pm 1} - \sum_{i=2}^5 W_{\pm i} \right| > 0.4 \end{cases} \quad (1)$$

[0106] 判断原则如下:

[0107] ①电表的正反向总电量、正反向高峰时段电量、正反向平时段电量、正反向低谷时段电量均大于0,且不为空。

[0108] ②电表正反向各费率电量之和大于0;

[0109] ③电表总电量与各费率电量之和差值的绝对值大于0.4。

[0110] ④若同时满足判断原则①②③,则判断为严重故障。

[0111] (2) 当进行电能表飞走和突变故障数据筛选时,判断电能表飞走和突变的故障程序如下:

[0112] a. 首先计算当天最大电量 W_f :

$$[0113] W_f = \begin{cases} \frac{220 \times I_{\max} \times 12}{1000} & \text{月份} = 1, 2, 7, 8, 9, 12 \\ \frac{220 \times 3 \times I_b \times 8}{1000} & \text{月份} = 3, 4, 5, 6, 10, 11 \end{cases} \quad (2)$$

[0114] 式中: W_f 为当天最大电量; I_{\max} 为最大电流; I_b 为基本电流。

[0115] b. 然后计算电表飞走和突变因子 K :

$$[0116] K = \frac{W_f}{W_t} \quad (3)$$

[0117] 式中: K 为电表飞走和突变因子; W_t 为当日电量。

[0118] 判断原则如下:

[0119] ①针对DDZY102-Z智能电表的两年数据中,在1、2、7、8、9、12月份取最大电流 I_{\max} ,按12小时计算电表飞走和突变因子 K ;在3、4、5、6、10、11月份取3倍基本电流 I_b ,按8小时计算电表飞走和突变因子 K 。

[0120] ②当进行电能表飞走和突变故障数据筛选时,采用公式(4)判断电能表飞走和突变故障:

$$[0121] K \geq 1 \quad (4)$$

[0122] ③若满足判据②,则判断为严重故障。

[0123] (3) 当进行电表反向有功示值大于零故障数据筛选时,采用下式判断电表反向有功示值大于零的故障:

$$[0124] \begin{cases} P_{-总} > 0 \\ Q_{-总} > 0 \end{cases} \quad (5)$$

[0125] 式中:P代表有功功率,Q代表无功功率, $P_{-总}$ 代表电表反向有功总功率, $Q_{-总}$ 代表电表反向无功总功率。

[0126] 判断原则如下:

[0127] ①电能表出现反向有功总功率或反向无功总功率大于0。

[0128] ②若满足判据①,则判断为严重故障。

[0129] (4)当进行电能表倒走故障数据筛选时,电能表倒走判断的前提为排除抄读电能示值正、反向有功总为空的记录,采用下式判断电能表倒走的故障:

$$[0130] \quad \begin{cases} W_{+1} < W_{+1y} \\ W_{-1} < W_{-1y} \end{cases} \quad (6)$$

[0131] 式中: W_{+1y} 为前一天的正向总电量, W_{-1y} 为前一天的反向总电量。

[0132] 判断原则如下:

[0133] 根据日电量统计表判断,针对低压居民和单相工商业户,只判断前一天正向总电量和反向总电量是否大于当天的抄表示值,如果成立,则判断为严重故障。

[0134] (5)当进行电表时钟不对故障数据筛选时,电表时钟不对的故障判断程序如下:

[0135] 首先判断对时错误次数 m ,如果对时错误次数大于3次,则直接判定为严重故障,即:

$$[0136] \quad m > 3 \text{ 严重} \quad (7)$$

[0137] 式中: m 为对时错误次数。

[0138] 若 m 不大于3,由在线监测查询是否满足 Δt 的要求,并按 Δt 的标准生成电能表时钟超差等级,则采用公式(8)判断:

$$[0139] \quad \begin{cases} 5 \leq \Delta t < 15 & \text{一般} \\ 15 \leq \Delta t < 30 & \text{重要} \\ \Delta t > 30 & \text{严重} \end{cases} \quad (8)$$

[0140] 式中: Δt 为终端与电能表时钟之差,计算方法见式(9):

$$[0141] \quad \Delta t = |t_{\text{终端}} - t_{\text{电表}}| \quad (9)$$

[0142] 式中: $t_{\text{终端}}$ 表示终端时钟, $t_{\text{电表}}$ 表示电能表时钟。

[0143] 判断原则如下:

[0144] ①如果对时错误次数 m 超过3次,直接判断为严重故障;

[0145] ②如终端与电能表时钟之差 Δt 满足 $5\text{min} \leq \Delta t < 15\text{min}$,判断为一般故障;

[0146] 若 $15\text{min} \leq \Delta t < 30\text{min}$ 判断为重要故障;若 $\Delta t > 30\text{min}$ 判断为严重故障;

[0147] (6)当进行电表电能费率设置异常故障数据筛选时,采用下式判断电表电能费率设置异常的故障:

$$[0148] \quad W_{+5} \neq 0 \text{ 或 } W_{-5} \neq 0 \quad (10)$$

[0149] 判断原则如下:

[0150] ①只判断DLT-2007规约的电能表。

[0151] ②判断是否存在正向尖峰时段电量或反向尖峰时段电量,若存在,则判断为严重故障。

[0152] (7)当进行电能表正向潜动故障数据筛选时,采用下式判断电能表正向潜动的故障:

$$[0153] \quad 0.98 \times W_{+3} \leq \frac{14}{24} \times W_{+1} \leq 1.02 \times W_{+3} \text{ 或 } 0.98 \times W_{+4} \leq \frac{10}{24} \times W_{+1} \leq 1.02 \times W_{+4} \quad (11)$$

[0154] 判断原则如下：

[0155] 若满足公式 (11) 3次，则判断为重要故障，若满足公式 (11) 5次，则判断为严重故障。

[0156] (8) 当进行电能表反向潜动故障数据筛选时，采用下式判断电能表反向潜动的故障：

$$[0157] \quad \begin{cases} W_{+1} > 0.1 \\ W_{-1} > 0.1 \end{cases} \quad (12)$$

[0158] 判断原则如下：

[0159] 针对结算类的电能表正向总电量和反向总电量同时存在并大于0.1的情况，则判断为严重故障。

[0160] 将DDZY102-Z智能电表故障判断模型筛选出的所有故障存放至DDZY102-Z智能电表故障数据库中。

[0161] 步骤3，选取DDZY102-Z智能电表故障数据库中的历史数据，将其分为训练集和测试集，采用决策树算法对训练集进行数据挖掘，形成智能电表故障决策树，然后形成初步分类规则，其详细步骤如下：

[0162] 设R是智能电表故障数据库的历史数据，并将R的60%数据分给训练集S，R的40%数据分给测试集T。以采用C5.0算法为例对训练集S进行数据挖掘。

[0163] 设具有n个值的属性X将S分成n个子集 S_1, S_2, \dots, S_n ，设S中样本的总数为 $|S|$ ， $\text{freq}(C_i, S)$ 是S中属于类别 C_i ($i=1, 2, \dots, N$)的样本个数，S中某个样本属于类别 C_i 的概率为 $\frac{\text{freq}(C_i, S)}{|S|}$ ，它所传达的信息为： $\frac{\log_2(\text{freq}(C_i, S))}{|S|}$ ；

[0164] 训练集S的熵 $\text{info}(S)$ 由式 (13) 表示；

$$[0165] \quad \text{info}(S) = \sum_{i=1}^N \frac{\text{freq}(C_i, S)}{|S|} \log_2 \left(\frac{\text{freq}(C_i, S)}{|S|} \right) \quad (13)$$

[0166] 将训练集S根据属性X分割成n个子集后，计算各个子集的信息熵，则集合S的期望信息 $\text{info}_X(S)$ 用公式 (14) 来表示；

$$[0167] \quad \text{info}_X(S) = \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{info}(S_i) \quad (14)$$

[0168] 为了度量按照属性检验X进行分区的S所得到的信息，使用增益标准 $\text{gain}(X)$ ，它选择使其值能最大化，即此标准选择的是具有最高信息增益的属性来进行每次分区；

$$[0169] \quad \text{gain}(X) = \text{info}(S) - \text{info}_X(S) \quad (15)$$

[0170] 根据属性X的n个不同取值将S分割成 S_1, S_2, \dots, S_n 共n个子集后产生的潜在信息 $\text{split_info}(X)$ 用式 (16) 所示；

$$[0171] \quad \text{split_info}(X) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (16)$$

[0172] 公式 (17) 为X对S进行划分信息增益率 $\text{gain_ratio}(X)$ ；

$$[0173] \quad gain_ratio(X) = \frac{gain(X)}{split_info(X)} \quad (17)$$

[0174] 首次计算采用选择最高信息增益率的属性作为决策树的根节点,然后决策树的每一个节点用此方法计算剩余属性的信息增益率,仍选择最高信息增益率的属性作为决策树的当前节点,直到最终形成整个决策树。

[0175] 由该决策树形成初步分类规则。

[0176] 步骤4,通过测试集的数据对初步分类规则进行正确率评估,若正确率满足要求,则确定分类规则,若正确率不满足要求,则返回至训练集,重新进行训练,其详细步骤如下:

[0177] (3) 正确率评估公式为:

$$[0178] \quad Z = \frac{A}{B} \times 100\% \quad (18)$$

[0179] 式中Z为评估正确率,A为初步分类规则在测试集T中预测正确的个数;B为测试集T的总数据个数。

[0180] (4) 设Z' 为设定的正确率。若 $Z \geq Z'$,则确定分类规则;若 $Z < Z'$,则返回训练集重新训练决策树。

[0181] 步骤5,由最终确定的分类规则生成DDZY102-Z智能电表故障实时预测模型;

[0182] 步骤6,将DDZY102-Z智能电表实时故障数据库连接至DDZY102-Z智能电表故障实时预测模型进行实时预测,得到DDZY102-Z智能电表故障实时预测结果。

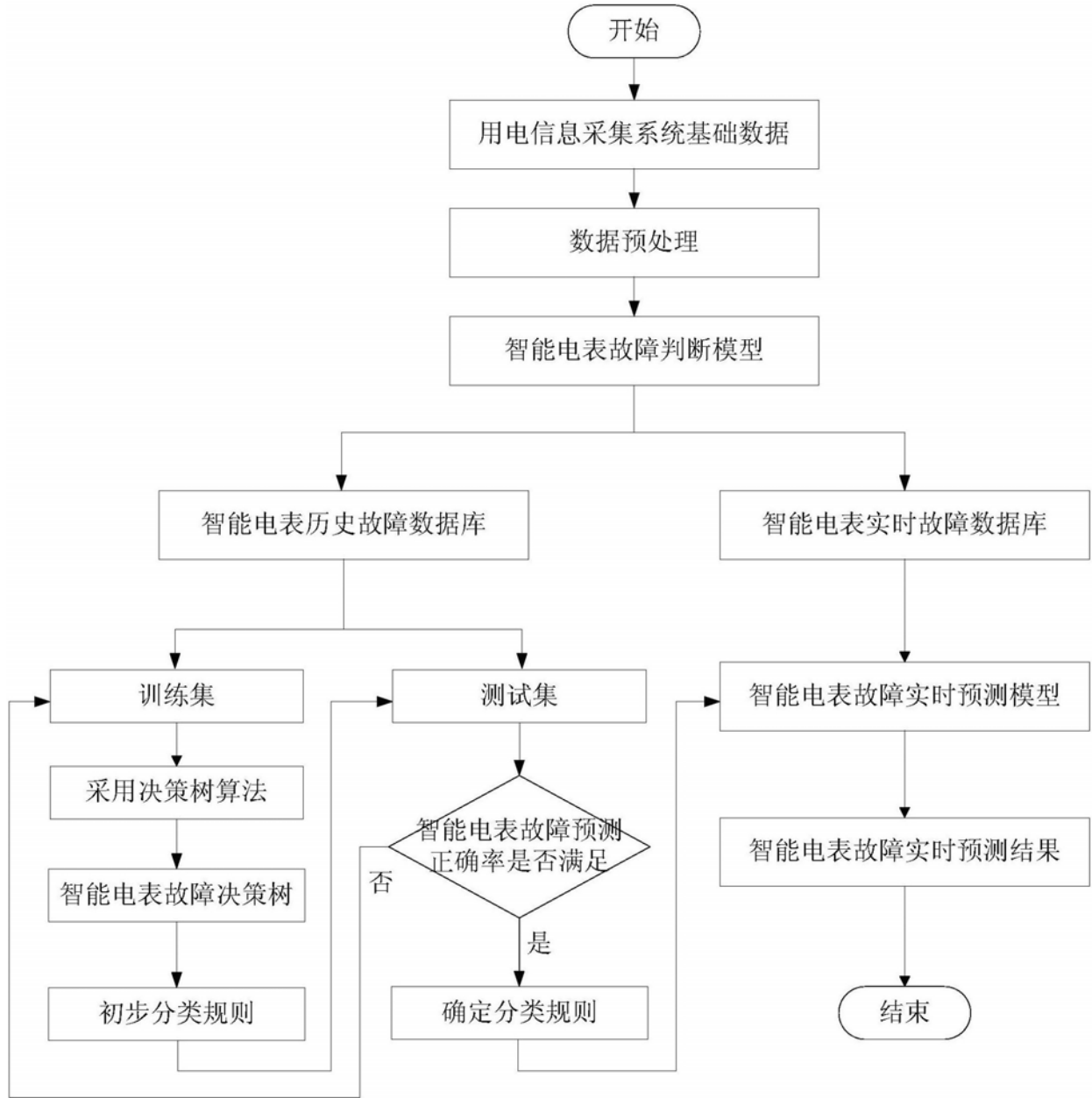


图1