



(12)发明专利

(10)授权公告号 CN 109469919 B

(45)授权公告日 2020.07.28

(21)申请号 201811336474.8

G06K 9/62(2006.01)

(22)申请日 2018.11.12

审查员 周勤

(65)同一申请的已公布的文献号

申请公布号 CN 109469919 A

(43)申请公布日 2019.03.15

(73)专利权人 南京工程学院

地址 211167 江苏省南京市江宁科学园弘景大道1号

(72)发明人 顾慧 崔晓波 陈磐

(74)专利代理机构 南京瑞弘专利商标事务所

(普通合伙) 32249

代理人 张耀文

(51)Int.Cl.

G06F 16/906(2019.01)

F23L 15/00(2006.01)

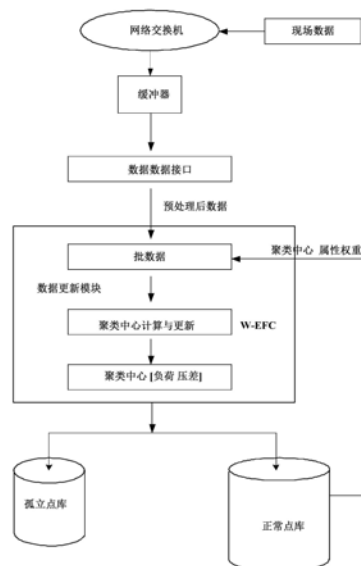
权利要求书1页 说明书5页 附图1页

(54)发明名称

一种基于权重聚类的电站空预器堵灰监测方法

(57)摘要

一种基于权重聚类的电站空预器堵灰监测方法,涉及电站空气预热器工况监测方法的技术领域。本发明在离线聚类EFC算法的基础上,增加权重项w,该权重衡量聚类样本各个属性之间的重要程度,融合权重w和欧式距离加权获得新的聚类相异性表征,弱化孤立点的聚类结果影响,从而达到剔除孤立点的效果。记录聚类获得的各中心,并在处理下一批动态过程数据时,完成相应更新操作,最终获得空预器运行过程中,剔除孤立点后的负荷和压差(空气、烟气)的聚类中心。本发明能够自动更新聚类结果,较为准确地反映空气预热器的负荷和压差之间的关系;可用作空预器堵灰程度的监测,并可以结合现场机组设定的阈值来进一步确定空预器适当的冲洗时间。



1. 一种基于权重聚类的电站空预器堵灰监测方法,其特征在于:包括如下步骤:

(1) 现场数据经由网络交换机输出至数据接口;

(2) 批量采集空预器进口烟气压力 p_{yi} ,空预器出口烟气压力 p_{yo} ,空预器进口空气压力 p_{ki} ,空预器出口空气压力 p_{ko} 和机组负荷 P_{load} 建立样本数据库 $X=[p_{yi}, p_{yo}, p_{ki}, p_{ko}, P_{load}]$, M 为该批数据样本数量, m 为样本属性个数;

(3) 集合权重的聚类算法得到聚类结果 $\{cluster\}$, $cluster$ 代表各簇聚类中心;

(4) 处理完此批次数据后,将获得的聚类类别和聚类中心存入正常点库,并作为下一批数据处理过程中聚类中心更新的依据;

(5) 再次执行步骤(2)、步骤(3)、步骤(4);

(6) 将步骤(5)获得的聚类结果与上一批次样本的聚类结果合并更新,将更新后的聚类结果和聚类中心存入正常点库;

(7) 按照上述步骤将过程数据按批次处理,最终获得当前时间戳下的聚类结果, $cluster$ 的聚类中心属性 $=\{p_{yi}, p_{yo}, p_{ki}, p_{ko}, P_{load}\}$,根据各聚类中心计算出 $[\Delta p_y, \Delta p_k, P_{load}]$,分别为空预器进出口烟气压力差,空预器进出口空气压力差和机组负荷。

2. 根据权利要求1所述的基于权重聚类的电站空预器堵灰监测方法,其特征在于:上述步骤(3)的聚类算法为:

3.1、计算第 i 个样本第 j 属性的权重 w_{ij}

$$\begin{cases} w_{ij} = \frac{1/St_{ij}}{\sum_{j=1}^m 1/St_{ij}} \\ \sum_{i=1}^M w_{ij} = 1 \end{cases}$$

其中标准差 $St_{ij} = \sqrt{\frac{1}{M} \sum_{l=1, l \neq i}^M (x_{ij} - x_{l,j})^2}$,其中 $l=1, 2, \dots, m$,且 $l \neq i$,将 St_i 大于阈值 α 的样本记为孤立点并从总样本集中标出, M 为数据样本数量, m 为样本属性个数;

3.2、计算每两个数据样本之间的相似度, $S_{ij} = e^{-\alpha D_{ij}}$, D_{ij} 表示数据样本 x_i 和 x_j 的权重欧氏距离, $D_{ij} = \sqrt{\sum_{l=1}^m (w_{il} \times x_{il} - w_{jl} \times x_{jl})^2}$, $x_i, x_j \in X$,式中 $\alpha = -\ln 0.5 / \tilde{D}$, \tilde{D} 为数据集内权重欧式距离的平均值;

3.3、对各样本 x_i 计算信息熵 E_i , $E_i = -\sum_{j \in X}^{j \neq i} (S_{ij} \log_2 S_{ij} + (1 - S_{ij}) \log_2 (1 - S_{ij}))$;

3.4、选择各数据样本熵中的最小值, $E_{\min, iter}$,其中下标 $iter$ 为当前迭代次数标记, $E_{\min, iter} = E_k = \min \{E_i\}$, $i, k=1, 2, \dots, M$,并标记具有最小信息熵 $E_{\min, iter}$ 的样本 x_k ;

3.5、筛选与样本 x_k 的相似度集合 $\{S_{ij}\}$, $j=1, 2, \dots, M', j \neq i$,从原数据样本中剔除样本 x_k 和与 x_k 的相似度大于阈值的样本,记为更新后的 X , M' 为更新后数据样本数量;

3.6、此次迭代被踢除的样本记为这一代获得的聚类类别 $cluster_{iter}$;

3.7、判断样本集 X 是否为空,若不为空,跳至步骤3.1;

3.8、若样本集 X 为空,获得聚类结果 $\{cluster\}$ 。

一种基于权重聚类的电站空预器堵灰监测方法

技术领域

[0001] 本发明涉及一种电站空气预热器工况监测方法,兼有综合聚类和模糊理论的工况划分模型属于数据挖掘建模领域。

背景技术

[0002] 在锅炉风烟系统中,当烟气进入空气预热器时,烟气中的硫酸蒸汽将会在金属壁上产生凝结现象,对壁面产生酸腐蚀,该现象称为低温腐蚀。低温腐蚀造成空气预热器受热面金属的破裂穿孔,使得空气大量漏到烟气中,加重空气预热器的积灰堵灰现象,造成引风机出力不足,使得燃烧室微负压难以维持。另一方面,如果空气漏至烟气中,将致使送风不足,炉内燃烧恶化,严重影响锅炉的安全经济运行。在运行参数上主要反映为相同负荷下,空预器进出口烟气压差、空气压差的非正常变化。

[0003] 在此类工业的过程数据中,存在一些无法复现、反映系统真是性能的孤立点,这些孤立点的存在加大设备建模的难度,因此需要将其识别并剔除。在火电机组的空预器监测过程中,需要对现场实时数据进行分析,从这些动态数据获取设备性能状况。过程数据存在的孤立点对监测空预器堵灰程度有比较大的影响。

发明内容

[0004] 本发明目的是提供一种能够自动更新聚类结果,较为准确地反映空气预热器的负荷和压差之间的关系的基于权重聚类的电站空预器堵灰监测方法。

[0005] 一种基于权重聚类的电站空预器堵灰监测方法,其特征在于:包括如下步骤:

[0006] 1、一种基于权重聚类的电站空预器堵灰监测方法,其特征在于:包括如下步骤:

[0007] (1) 现场数据经由网络交换机输出至数据接口;

[0008] (2) 批量采集空预器进口烟气压强 p_{yi} ,空预器出口烟气压强 p_{yo} ,空预器进口空气压强 p_{ki} ,空预器出口空气压强 p_{ko} 和机组负荷 P_{load} 建立样本数据库 $X = [p_{yi}, p_{yo}, p_{ki}, p_{ko}, P_{load}]$, M 为该批数据样本数量, m 为样本属性个数;

[0009] (3) 集合权重的聚类算法得到聚类结果 $\{cluster\}$, $cluster$ 代表各簇聚类中心;

[0010] (4) 处理完此批次数据后,将获得的聚类类别和聚类中心存入正常点库,并作为下一批数据处理过程中聚类中心更新的依据;

[0011] (5) 再次执行步骤(2)、步骤(3)、步骤(4);

[0012] (6) 将步骤(5)获得的聚类结果与上一批次样本的聚类结果合并更新,将更新后的聚类结果和聚类中心存入正常点库;

[0013] (7) 按照上述步骤将过程数据按批次处理,最终获得当前时间戳下的聚类结果, $cluster$ 的聚类中心属性 $= \{p_{yi}, p_{yo}, p_{ki}, p_{ko}, P_{load}\}$,根据各聚类中心计算出 $[\Delta p_y, \Delta p_k, P_{load}]$,分别为空预器进出口烟气压差,空预器进出口空气压力差和机组负荷。

[0014] 2、根据权利要求1所述的基于权重聚类的电站空预器堵灰监测方法,其特征在于:上述步骤(3)的聚类算法为:

[0015] 3.1、计算第*i*个样本第*j*属性的权重 w_{ij}

$$[0016] \begin{cases} w_{ij} = \frac{1/S_{ij}}{\sum_{j=1}^m 1/S_{ij}} \\ \sum_{i=1}^M w_{ij} = 1 \end{cases}$$

[0017] 其中标准差 $S_{ij} = \sqrt{\frac{1}{M} \sum_{l=1, l \neq i}^M (x_{ij} - x_{l,j})^2}$, 其中 $l=1, 2, \dots, m$, 且 $l \neq i$, 将 S_{ij} 大于阈值 α 的样本记为孤立点并从总样本集中标出, M 为数据样本数量, m 为样本属性个数;

[0018] 3.2、计算每两个数据样本之间的相似度, $S_{ij} = e^{-\alpha D_{ij}}$, D_{ij} 表示数据样本 x_i 和 x_j 的权重欧氏距离, $D_{ij} = \sqrt{\sum_{l=1}^m (w_{il} \times x_{il} - w_{jl} \times x_{jl})^2}$, $x_i, x_j \in X$, 式中 $\alpha = -\ln 0.5 / \tilde{D}$, \tilde{D} 为数据集内权重欧式距离的平均值;

[0019] 3.3、对各样本 x_i 计算信息熵 E_i , $E_i = -\sum_{j \in X}^{j \neq i} (S_{ij} \log_2 S_{ij} + (1 - S_{ij}) \log_2 (1 - S_{ij}))$;

[0020] 3.4、选择各数据样本熵中的最小值, $E_{\min, iter}$, 其中下标 $iter$ 为当前迭代次数标记, $E_{\min, iter} = E_k = \min \{E_i\}$, $i, k=1, 2, \dots, M$, 并标记具有最小信息熵 $E_{\min, iter}$ 的样本 x_k ;

[0021] 3.5、筛选与样本 x_k 的相似度集合 $\{S_{ij}\}$, $j=1, 2, \dots, M'$, $j \neq i$, 从原数据样本中剔除样本 x_k 和与 x_k 的相似度大于阈值的样本, 记为更新后的 X , M' 为更新后数据样本数量;

[0022] 3.6、此次迭代被剔除的样本记为这一代获得的聚类类别 $cluster_{iter}$;

[0023] 3.7、判断样本集 X 是否为空, 若不为空, 跳至步骤3.1;

[0024] 3.8、若样本集 X 为空, 获得聚类结果 $\{cluster\}$ 。

[0025] 至此, 空预器堵灰监测的主体部分已经完成。考虑锅炉中空预器设备的堵灰情况在运行时间较长或者中途经过大小修时发生变化, 可根据现场实际情况, 定期或者不定期重复核算, 获得具有标志性的时间戳下的聚类结果。

[0026] 有益效果: 本发明相对于现有技术而言, 具有以下优点:

[0027] (1) 结合权重修正聚类过程中的欧式距离, 并融合信息熵的概念, 完成样本自适应聚类, 将相邻两次批处理的结果合并, 完成过程数据的权重自适应聚类。

[0028] (2) 随着电站空预器设备运行时间、堵灰程度的改变, 自动更新聚类结果, 能够较为准确地反映空气预热器的负荷和压差之间的关系。

[0029] (3) 为电厂监测信息系统高级功能模块(状态监测与故障诊断等)提供可参考模型。

[0030] (4) 静态聚类可以将具有相似属性的离线数据归为同一类, 本发明将空预器运行数据做工况划分, 自适应数据流聚类获得[负荷压差]数据簇, 继而完成空预器堵灰程度的监测, 该过程需要综合考虑聚类结果的准确性。本发明在离线聚类EFC算法的基础上, 增加权重项 w , 该权重衡量聚类样本各个属性之间的重要程度, 融合权重 w 和欧式距离加权获得新的聚类相异性表征, 弱化孤立点的聚类结果影响, 从而达到剔除孤立点的效果。记录聚类获得的各中心, 并在处理下一批动态过程数据时, 完成相应更新操作, 最终获得空预器运行过程中, 剔除孤立点后的负荷和压差(空气、烟气)的聚类中心, 准确反映设备堵灰情况。

附图说明

[0031] 图1是本发明的流程图。

具体实施方式

[0032] 下面结合附图对本发明作更进一步的说明。

[0033] 一种基于权重聚类的电站空预器堵灰监测方法,其特征在于:所述步骤包括如下:

[0034] (1) 现场数据经由网络交换机输出至数据接口;

[0035] (2) 批量采集空预器进口烟气压力 p_{yi} ,空预器出口烟气压力 p_{yo} ,空预器进口空气压力 p_{ki} ,空预器出口空气压力 p_{ko} 和机组负荷 P_{load} 建立样本数据库 $X = [p_{yi}, p_{yo}, p_{ki}, p_{ko}, P_{load}]$, M 为该批数据样本数量, m 为样本属性个数;

[0036] (3) 集合权重的聚类算法得到聚类结果{cluster};

[0037] 3.1、计算第 i 个样本第 j 属性的权重 w_{ij}

$$[0038] \begin{cases} w_{ij} = \frac{1/St_{ij}}{\sum_{j=1}^m 1/St_{ij}} \\ \sum_{i=1}^M w_{ij} = 1 \end{cases}$$

[0039] 其中标准差 $St_{ij} = \sqrt{\frac{1}{M} \sum_{l=1, l \neq i}^M (x_{lj} - x_{lj})^2}$,其中 $l=1, 2, \dots, m$,且 $l \neq i$,将 St_i 大于阈值 α 的样本记为孤立点并从总样本集中标出, M 为数据样本数量, m 为样本属性个数;

[0040] 3.2、计算每两个数据样本之间的相似度, $S_{ij} = e^{-\alpha D_{ij}}$, D_{ij} 表示数据样本 x_i 和 x_j 的权重欧氏距离, $D_{ij} = \sqrt{\sum_{l=1}^m (w_{il} \times x_{il} - w_{jl} \times x_{jl})^2}$, $x_i, x_j \in X$,式中 $\alpha = -\ln 0.5 / \tilde{D}$, \tilde{D} 为数据集内权重欧式距离的平均值;

[0041] 3.3、对各样本 x_i 计算信息熵 E_i , $E_i = -\sum_{j \in X}^{j \neq i} (S_{ij} \log_2 S_{ij} + (1 - S_{ij}) \log_2 (1 - S_{ij}))$;

[0042] 3.4、选择各数据样本熵中的最小值, $E_{\min, iter} = E_k = \min \{E_i\}, i, k=1, 2, \dots, M$,并标记具有最小信息熵的样本 x_k ;

[0043] 3.5、筛选与样本 x_k 的相似度集合 $\{S_{ij}\}, j=1, 2, \dots, M', j \neq i$,从原数据样本中剔除样本 x_k 和与 x_k 的相似度大于相似度阈值0.7的样本,记为更新后的 X , M' 为更新后数据样本数量;

[0044] 3.6、此次迭代被踢除的样本记为这一代获得的聚类类别 $cluster_{iter}$;

[0045] 3.7、判断样本集 X 是否为空,若不为空,跳至步骤3.1;

[0046] 3.8、若样本集 X 为空,获得聚类结果{cluster};

[0047] (4) 处理完此批次数据后,将获得的聚类类别和聚类中心存入正常点库,并作为下一批数据处理过程中类心更新的依据;

[0048] (5) 再次执行步骤(2)、步骤(3)、步骤(4);

[0049] (6) 将步骤(5)获得的聚类结果与上一批次样本的聚类结果合并更新,将更新后的聚类结果和中心存入正常点库;

[0050] (7) 按照上述步骤将过程数据按批次处理, 最终获得当前时间戳下的聚类结果, cluster的聚类中心属性 = $\{p_{yi}, p_{yo}, p_{ki}, p_{ko}, P_{load}\}$, 根据各聚类中心计算出 $[\Delta p_y, \Delta p_k, P_{load}]$ 。

[0051] 以某电站600MW超临界锅炉(SG-2090/25.4-M975)为例, 采集SIS系统中2014年4月1日0点到2014年10月1日24点的数据, 采集间隔30秒。本发明框架主要有输入数据预处理、数据聚类与更新模块和聚类结果入库模块, 详细流程如图1所示:

[0052] 1) 现场数据经由网络交换机输出至数据接口;

[0053] 2) 以2014年4月1日为第一个时间戳, 以时间间隔30秒采集机组空预器入口烟气压、空预器出口烟气压、空预器入口空气压力、空预器出口空气压力和机组负荷, $[p_{yi}, p_{yo}, p_{ki}, p_{ko}, P_{load}]$, 过滤器去除其中不稳定数据后, 将数据样本分为五个典型负荷工况段 $[400MW, 450MW, 500MW, 550MW, 600MW]$, 各工况段以 $\pm 2MW$ 为邻域。

[0054] 3) 设定批处理数据长度为50, 将最大聚类个数设定为4, 阈值 α 定为0.3, 对第一批数据进行权重聚类分析。

[0055] 3.1对样本分负荷段做差值处理 $(\Delta p_k, \Delta p_y)$, 以450MW段的第一个样本为例, 该计算样本 St_{11} 和 St_{12} , 分别为0.0667和0.0145, 均小于阈值0.3。

[0056] 3.2计算该批450MW数据个样本的熵值, 选择最小熵值0.234对应的样本 $[0.9600.430]$ 。

[0057] 3.3计算其余样本与该样本的相似度, 将大于0.7相似度的样本和该样本合并作为第一类, 共13个样本, 计算其聚类中心, 并从原数据集中删除。

[0058] 3.4更新数据集, 若样本集不为空, 回到步骤3.1。

[0059] 4) 第一批数据完成后, 共分3类, 聚类结果分别入库(孤立点、正常聚类中心)。

[0060] 5) 动态更新下一批处理数据, 依据最近类心距离判断更新后的样本属于哪一类, 样本更新聚类中心, 当到达最后数据时, 停止操作。以450MW负荷段为例, 最终获得3个类别, 其类心分别为 $[0.958 \ 0.431]$ 、 $[1.077 \ 0.470]$ 和 $[1.031 \ 0.451]$ 。

[0061] 6) 以2014年10月1日为第二个时间戳, 比较分析各个时间戳下的聚类结果, 各类别最大最小聚类中心汇总后, 如表1和表2所示。

[0062] 表1和表2数据体现空气预热器运行情况、堵灰程度随着时间的变化情况, 表中给出了负荷、聚类样本、该工况(负荷段)下的类心个数以及各聚类中心的最大值和最小值。空气预热器的进出口空气压差 Δp_k 和进出口烟气压差 Δp_y 越大, 则表明空气预热器的堵灰程度越严重。

[0063] 该算例验证了本发明应用于空气预热器系统堵灰监测的可行性和有效性。以W-EFC聚类批处理过程数据为基础, 通过对相关压差与机组负荷的监测, 能够完成对空预器堵灰程度的监测, 并可以结合现场机组设定的阈值来进一步确定空预器适当的冲洗时间。

[0064] 表1第一个时间戳下的W-EFC聚类结果

[0065]	负荷	样本数	聚类个数	正常中心点 (最小/最大点)		孤立点	
	P_{load}/MW			$\Delta p_k/kPa$	$\Delta p_y/kPa$	$\Delta p_k/kPa$	$\Delta p_y/kPa$
	400	74	3	0.807/0.887	0.398/0.429	0.577	0.250
	450	105	3	0.958/1.077	0.431/0.470	1.272	0.693
	500	217	4	1.063/1.173	0.492/0.530	/	/
	550	130	3	1.126/1.317	0.556/0.608	1.432	0.721
	600	68	2	1.276/1.430	0.610/0.650	1.492	0.791

[0066] 表2第二个时间戳下的W-EFC聚类结果

[0067]	负荷	聚类个数	正常中心点 (最小/最大点)		孤立点	
	P_{load}/MW		$\Delta p_k/kPa$	$\Delta p_y/kPa$	$\Delta p_k/kPa$	$\Delta p_y/kPa$
	400	3	1.387/1.542	0.515/0.659	1.732	0.432
	450	4	1.435/1.632	0.703/0.899	1.165	0.313
	500	3	1.476/1.709	0.829/1.144	1.043	0.564
	550	3	1.507/1.811	0.936/1.354	1.202	0.632
	600	3	1.538/1.860	1.111/1.499	1.013	0.695

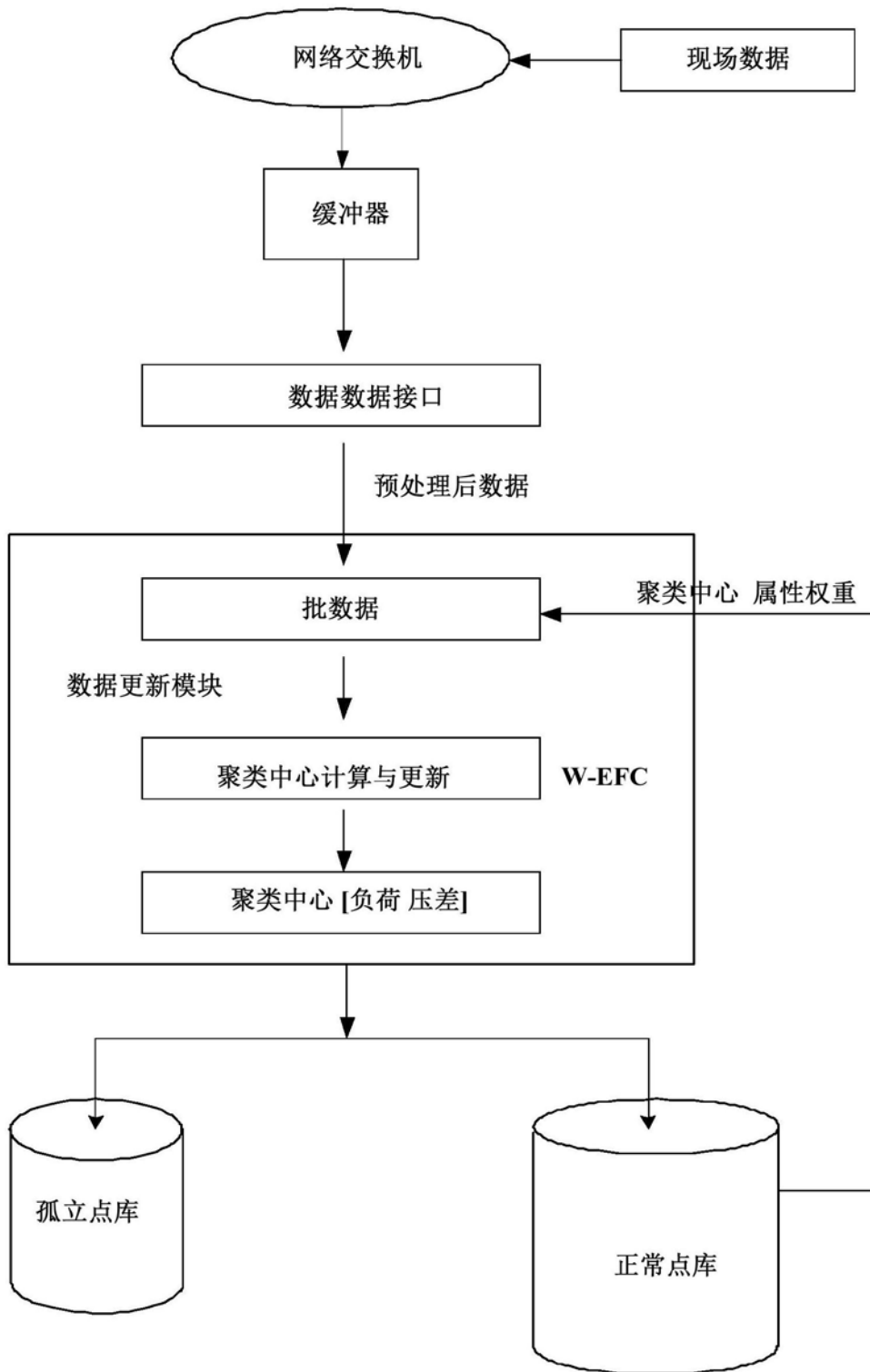


图1