文章编号：2095-4980(2024)10-1154-07

基于国产CPU环境的国产数据库历史数据迁移技术

毛　冬[1]，张　辰[1]，陈又咏[3]，刘永清[2]，焦艳斌[2]

(1.国网浙江省电力有限公司 信息通信分公司，浙江 杭州 310007；2.国网信息通信产业集团有限公司，北京 102211；
3.福建亿榕信息技术有限公司，福建 福州 350001)

摘　要：针对目前数据迁移方法存在数据迁移耗时长、存储空间最大占用率较高、迁移学习错误率高和被访问数据在线概率低的问题，开展基于国产CPU环境的国产数据库历史数据迁移技术的研究。首先在国产CPU环境中集群部署系统软硬件，提高历史数据在国产数据库之间的迁移速率。其次建立孤立森林模型，将历史数据输入孤立森林模型中展开趋势预测，剔除国产数据库中存在的异常数据，减少待迁移的数据量。最后，构建数据迁移模型，并采用交替优化策略求取模型最优解，完成国产数据库历史数据的迁移。实验结果表明，该方法的数据迁移时间为18 min，储存空间最大占用率在10%~25%之间，ALC指标值为0.78~0.95，被访问数据在线概率能够始终保持在97%以上，证明该方法数据迁移耗时较短，存储空间最大占用率较低，迁移学习的错误率低，访问效率高，具有较好的应用效果。

关键词：国产CPU；国产数据库；孤立森林模型；交替优化策略；数据迁移技术

中图分类号：TP391　　**文献标志码：**A　　**doi：**10.11805/TKYDA2022250

Historical data migration technology of domestic database based on domestic CPU environment

MAO Dong[1]，ZHANG Chen[1]，CHEN Youyong[3]，LIU Yongqing[2]，JIAO Yanbin[2]

(1.Information Communication Branch of State Grid Zhejiang Electric Power Co.，Ltd，Hangzhou Zhejiang 310007；
2.State Grid Information & Telecommunication Group Co.，Ltd，Beijing 102211，China；
3.Fujian Yirong Information Technology Co.，Ltd，Fuzhou Fujian 350001，China)

Abstract：In view of the problems of long data migration, high maximum occupancy rate of storage space, high error rate of transfer learning and low online probability of visited data, the historical data migration technology of domestic database based on domestic Central Processing Unit(CPU) environment is studied. Firstly, the system software and hardware are clustered and deployed in the domestic CPU environment to improve the migration rate of historical data between domestic databases. Secondly, an isolation forest model is established, and the historical data is input into the isolation forest model for trend prediction, thereby eliminating the abnormal data in the domestic database, and reducing the amount of data to be migrated. Finally, a data migration model is constructed, and an alternating optimization strategy is adopted to find the optimal solution of the model, thus completing the migration of historical data in domestic databases. The experimental results show that the data migration time of this method is 18 minutes, and the maximum occupancy rate of storage space is between 10% and 25%, the ALC(Area under the Learning Curve) index value is 0.78~0.95, and the online probability of the accessed data can always be maintained at more than 97%, proving that this method has a short data migration time, a low maximum occupancy rate of storage space, a low error rate of migration learning, and high access efficiency, demonstrating good application effects.

Keywords：domestic CPU；domestic database；isolation forest model；alternating optimization strategy；data migration technology

收稿日期：2022-12-28；**修回日期：**2023-05-12

基金项目：国网科技基金资助项目(5700–202219187A–1–1–ZN)

随着大数据时代的发展，海量数据信息的储存成为现阶段数据安全领域研究人员重点关注的问题。数据库管理系统的存储能力和读取能力可通过迁移历史数据得以提升，以数据储存的时间作为评价指标，来衡量系统的储存性能，储存时间越短，说明系统的运行能力越强。为了缩短海量数据在储存过程中的访问时间，该领域的学者研究出较多的数据缓存技术，旨在解决因网络速度不达标导致的数据库存储过程中传输速度下降的问题，但是在此过程中依旧会出现数据丢失的情况[1]。比如传输错误、存储介质损坏、操作失误、安全漏洞、迁移配置错误、应用程序错误、迁移过程中断、服务器故障等，都会导致数据丢失，对数据安全造成较大的影响。为解决这一问题，提出数据迁移技术，该技术可通过镜像数据恢复系统业务。同时数据库在系统中的服务器性能过差也会降低系统性能，通过数据库历史数据迁移技术可提高系统的整体性能以及用户的访问速度[2]。蒲勇霖[3]等根据内存和CPU利用资源约束算法确定需要迁移的数据，在Storm平台中设计数据迁移节能策略，降低节点之间通讯所需的开销，实现数据迁移。但是该方法迁移时间较长，并且完成数据迁移后，被访问数据的在线概率较低，数据迁移效果较差。郭辉[4]等根据服务器处理能力、网络带宽和延时建立多参数的马尔科夫决策过程收益函数，设计数据迁移策略，在数据迁移过程中计算历史数据对应的权重，根据计算结果更新数据，完成数据迁移。但是该方法的存储空间最大占用率较高，且主动学习评价指标(ALC)值较低，存在迁移学习错误率高的问题。为了解决上述方法中存在的问题，提出基于国产CPU环境的国产数据库历史数据迁移技术。

1　国产数据库历史数据迁移方法设计

1.1　国产CPU环境下集成部署技术

开发国产CPU环境时，所需的关键软硬件适配技术包括数据库和操作系统等，研究国产CPU的重点内容是基础软硬件的优选整合，选取适配的产品型号为国产数据库历史数据迁移提供技术支持。

基于国产CPU环境的集群部署技术如图1所示。

研究集群部署的解析技术和表示技术，在国产CPU环境下，采用轻量级数据交换格式等形式化语言，统一表示软件路径信息、集群实例、部署软件和信息配置等内容，利用上述信息构建规划文件。服务器端接收引擎传输的软件介质和信息，进行解析后执行相应命令。同时，引擎负责解析并校验规划文件，以完成软件部署过程。在分布式协调等基础架构的支持下，实现软件配置的集中管理。一旦配置发生变更，系统会将变动信息推送至集群中的各个节点，确保配置的统一管理与及时更新。

在国产CPU环境下，通过部署集群的方式，可以解决国产数据库在历史数据迁移过程中遇到的视图和数据表差异性问题，从而提升数据迁移的效率。

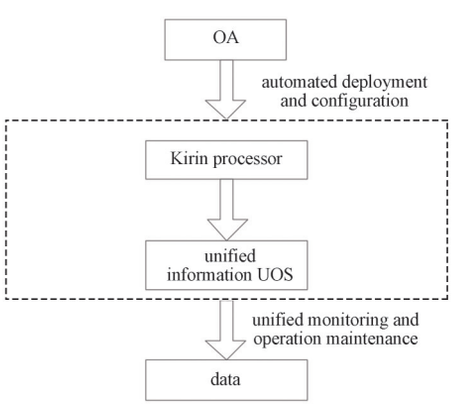

Fig.1 Cluster deployment technology
图1　集群部署技术

1.2　数据清洗

为了提高国产数据库历史数据迁移的准确率，基于国产CPU环境的国产数据库历史数据迁移技术，在孤立森林的基础上建立异常数据识别模型。该模型能够有效识别并消除国产数据库中存在的异常数据，减少数据迁移量，有效减少数据迁移耗时，降低存储空间最大占用率。

孤立森林模型的构建过程如下：

1) 随机在国产数据库训练数据矩阵$\boldsymbol{X}$中抽取一个样本矩阵$\boldsymbol{X}_z$，将其作为第b棵树根节点在孤立森林中的集合，$b=1,2,3,\cdots,a$，a代表的是根节点在森林中的数量，上述矩阵的表达式如下：

$$\begin{cases}\boldsymbol{X}=[x_1,x_2,\cdots,x_n]\\ \boldsymbol{X}_z=[x_1,x_2,\cdots,x_{n_z}]\end{cases} \tag{1}$$

式中：训练数据矩阵$\boldsymbol{X}$为$m\times n$维矩阵，其中包含n个变量，每个变量中存在m个数据样本；样本矩阵$\boldsymbol{X}_z$为$m_z\times n_z$维矩阵，其中包含n_z个变量，每个变量中存在m_z个数据样本。

2) 二叉分割处理样本矩阵$\boldsymbol{X}_z$，在样本矩阵$\boldsymbol{X}_z$中随机抽取一个列向量$\boldsymbol{x}_j$，在列向量$\boldsymbol{x}_j$中利用式(2)确定切割点Y：

$$Y=\min(\boldsymbol{x}_j)+t\left[\max(\boldsymbol{x}_j)-\min(\boldsymbol{x}_j)\right] \tag{2}$$

式中：$\min(\boldsymbol{x}_j)$、$\max(\boldsymbol{x}_j)$分别代表的是列向量$\boldsymbol{x}_j$中存在的最小值和最大值；t为随机数，在区间[0,1]内取值。

当$\boldsymbol{X}_z(i,j)$的值小于Y时，在左子树节点中存入矩阵$\boldsymbol{X}_z$第i行的全部变量；反之，将变量存储到右子树节点中。重复上述过程分割矩阵$\boldsymbol{X}_z$，最终通过矩阵变换，得到左、右子树节点集合的最终矩阵为$\boldsymbol{X}_1$和$\boldsymbol{X}_r$。

3) 用h_n表示节点中矩阵$\boldsymbol{X}_1$、$\boldsymbol{X}_r$对应的路径长度，当节点数量少于m_z或树高$h_{\max}$小于路径长度h_n时，建立单棵隔离树[5−6]。

4) 重复上述步骤，由多棵隔离树构成孤立森林模型。

在孤立森林模型中输入国产数据库中的历史数据，通过观察历史数据中不同密度的区域，对区域中存在的异常数据值进行评价，剔除掉这些评分最高的异常数据值：

$$\begin{cases} D(h_{ij},u)=2^{-R(h_{ij})/V(u)} \\ V(u)=2\left[\ln(u-1)+\psi\right]-2(u-1)/u \end{cases} \tag{3}$$

式中：$D(h_{ij},u)$为数据的异常值评分；$R(h_{ij})$为在第a棵隔离树中，国产数据库历史数据x_{ij}对应的平均路径长度；$V(u)$为历史数据在矩阵$\boldsymbol{X}$中对应的平均路径长度；u为样本数量；ψ为欧拉常数。当数据的异常值评分$D(h_{ij},u)$与1接近时属于异常数据。

根据数据的缺失特性[7−8]，结合上式计算结果，在国产数据库中剔除符合下述条件的历史数据：

1) 剔除符合$D(h_{ij},u)>0.75$的数据，并用0代替该历史数据；

2) 用0代替国产数据库中的缺失数据；

3) 将连续v_1个且数值相同的历史数据从国产数据库中剔除，并用0代替。

1.3 数据迁移

用M表示源数据库的总数量，用$F_{S_r}=\{(x_i^{S_r},y_i^{S_r})\}_{i=1}^{N_{S_r}}$表示其中第$i$个源数据库，其中$x_i^{S_r}$、$y_i^{S_r}$代表的是第$i$个源数据库中存在的数据量和标签量。用$N_S=\sum_{r=1}^{M}N_{S_r}$表示数据在全体国产数据库中的数量，目标数据库$F_G$由标签数据$F_{G_t}=\{x_i^G,y_i^G\}_{i=N_{G_u}+1}^{N_{G_u}+N_{G_t}}$和无标签数据$F_{G_u}=\{x_i^G\}_{i=1}^{N_{G_u}}$构成，数据总量为$N_G=N_{G_u}+N_{G_t}$。源数据库和目标数据库中共存在$N=N_S+N_G$个历史数据。通过训练各源数据库获得$M$个源分类器$\{g^{S_r}(x)\}_{r=1}^{M}$，目标分类器$g^G(x)$的表达式如下：

$$g^G(x)=w^{\mathrm{T}}\phi(x)+b \tag{4}$$

式中：$\phi(x)$为映射函数；w、b均为分类器参数。

基于国产CPU环境，国产数据库历史数据迁移技术在结构风险最小化、双加权策略、ε−不敏感损失的基础上构建了数据迁移模型：

$$\begin{aligned} \min K_m(\eta,g^G(x),w,b)=&\sum_{r=1}^{M}(\eta_r)^m\sum_{i=1}^{N_G}\sigma_i^{G_r}\left[g^G(x_i^G)-g^{S_r}(x_i^G)\right]^2+\mu_A\sum_{r=1}^{M}\sum_{i=1}^{N_{G_t}}\sigma_i^{S_r}\left[g^G(x_i^G)]^2+\mu_B\sum_{i}^{N_G}[g^G(x_i^G)-y_i^G\right]^2 \\ &+\mu_C w^{\mathrm{T}}w+\mu_D\left\{\sum_{i}^{N_G}\kappa_\varepsilon\left[w^{\mathrm{T}}\phi(x_i^G)+b-g^G(x_i^G)\right]+\sum_{r=1}^{M}\sum_{i=1}^{N_{S_r}}\left[w^{\mathrm{T}}\phi(x_i^{S_r})+b-g^G(x_i^{S_r})\right]\right\} \end{aligned} \tag{5}$$

式中：μ_A、μ_B、μ_C、μ_D均代表的是正则化参数；$\boldsymbol{\eta}=[\eta_1,\eta_2,\cdots,\eta_M]^{\mathrm{T}}$；$m$是加权指数。

迁移模型的约束条件为：

$$\left\{\begin{matrix} \sum_{\alpha=1}^{m}\delta_\alpha^{\mathrm{cpu}}\times E_{\beta,\alpha}<\vartheta_\beta^{\mathrm{cpu}} \\ \sum_{\alpha=1}^{m}\delta_\alpha^{\mathrm{mem}}\times E_{\beta,\alpha}<\vartheta_\beta^{\mathrm{mem}} \\ \sum_{\alpha=1}^{m}\delta_\alpha^{\mathrm{store}}\times E_{\beta,\alpha}<\vartheta_\beta^{\mathrm{store}} \end{matrix}\right\},\sum_{\beta=1}^{m}E_{\beta,\alpha}=1,E_{\beta,\alpha}\in\{0,1\} \tag{6}$$

式中：$\delta_\alpha^{\mathrm{cpu}}$、$\delta_\alpha^{\mathrm{mem}}$、$\delta_\alpha^{\mathrm{store}}$分别为虚拟机$\alpha$对CPU、内存和磁盘空间的请求大小；$E_{\beta,\alpha}$为第$\alpha$个数据是否被迁移到第$\beta$个数据库。

迁移的本质是将数据从源数据库移动到目标数据库，假设目标映射关系为$\pi(\cdot)$，那么迁移过程可理解为$\pi_0(\cdot)\to\pi(\cdot)$，因此设立优化目标：

$$\min\left|\bigcup_{wb_\alpha\in VM}\left\{\pi\left(F_{G_t},F_{G_\alpha}\right)\right\}\right|_{\sum_{\pi(wb_\alpha)=\theta m_\beta}\phi(x)\times l_i\leqslant Cap_\beta;\forall\theta m_\beta\in\theta\times M}\tag{7}$$

式中：θm_β为迁移到第β个数据库的流量代价；θ为全局迁移代价。通过式(5)所示的数据迁移模型中包含的5项内容，共同完成迁移模型的优化。

数据迁移模型中的各项含义如下：

数据迁移模型中存在的第1项代表的是双加权损失，其中，$(\eta_r)^m$为第r个源数据库在迁移学习过程中的重要性，符合下式条件：

$$\sum_{r=1}^{M}\eta_r=1,\eta_r>0\tag{8}$$

在数据迁移模型中，加权指数m的主要作用是控制各源数据库迁移的重要性程度以及目标函数的凹凸性，以此来降低迁移学习过程中的错误率。当加权指数m的值大于1时，可通过交替优化策略[9-10]求解数据迁移模型。

$g^G(x_i^G)$、$g^{S_r}(x_i^G)$分别代表的是针对第i个目标数据，目标分类器和第r个源分类器得到的预测结果。$\sigma_i^{G_r}$可以衡量$g^G(x_i^G)$对$g^{S_r}(x_i^G)$逼近准确性的置信度，其计算公式如下：

$$\sigma_i^{G_r}=\frac{\upsilon_{G_r}^{(c)}\mathrm{e}^{-\varsigma_{G_r}z_{G_r,i}^{(c)}}}{o_G}\tag{9}$$

式中：因子$\upsilon_{G_r}^{(c)}=N_{C_u}/|F_{G_r}^{(c)}|$，$F_{G_r}^{(c)}$是目标数据库无标数据集合；$\varsigma_{G_r}$是尺度因子；$z_{G_r,i}^{(c)}$是超平面与集合$F_{G_r}^{(c)}$中数据$x_i^G$之间存在的符号距离；$o_G$是标准化因子。

数据迁移模型中，第2项被设计为源数据库数据的Universum正则项，基于国产CPU环境的国产数据库历史数据迁移技术，将各源数据库中存在的数据作为Universum数据。在国产CPU环境下，当这些Universum数据与目标数据库中的数据越接近，与目标分类面之间的距离越小。所提方法利用权重$\sigma_i^{S_r}$衡量Universum数据$x_i^{S_r}$与目标国产数据库之间的接近程度，其计算公式如下：

$$\sigma_i^{S_r}=\frac{\upsilon_{S_r}^{(c)}e^{-\varsigma_{S_r}z_{S_r,i}^{(c)}}}{o_{S_r}}\tag{10}$$

式中：$\upsilon_{S_r}^{(c)}=N_{S_r}/|F_{S_r}^{(c)}|$，$F_{S_r}^{(c)}$为第$r$个源数据库类标为$c$的数据集合；$\varsigma_{S_r}$为尺度因子；$z_{S_r,i}^{(c)}$为超平面与集合$F_{S_r}^{(c)}$中数据$x_i^{S_r}$之间存在的符号距离；$o_{S_r}$为标准化因子。

数据迁移模型中的第3项涉及目标数据库标签数据的正则项，使用目标分类器对目标数据库中的标签数据F_{G_t}进行预测，即$g^G(x_i^G)$，并力求这一预测结果逼近标签数据的真实类别y_i^G[11-12]。

数据迁移模型中的第4项为L2正则项[13-14]，其主要目的是降低目标分类器$g^G(x)$的复杂度，避免$g^G(x)$出现过拟合现象。

数据迁移模型中的第5项代表的是ε-不敏感损失正则项[15]，其主要目的是保持数据的完整性，以提高访问效率。κ_ε为稀疏表示中的参数，目标数据为该项的首要组成部分，而Universum数据的ε-不敏感损失为该项中的末项。

在此目标模型中，η为数据库中的历史数据对应的迁移重要性程度；参数$\sigma_i^{G_r}$和$\sigma_i^{S_r}$分别用于衡量目标数据和Universum数据在国产数据库中迁移的重要性程度。

基于国产CPU环境的国产数据库历史数据迁移技术，采用交替优化策略求解数据迁移模型，完成数据库与数据迁移信息的适配预测输出平衡。

2　实验与分析

为了验证本文提出的基于国产CPU环境的国产数据库历史数据迁移技术的整体有效性，需要对其展开测试。选择某企业机房内自身服务器部署和运行搭建的互联网数据中心(Internet Data Center，IDC)，即IDC自建数据库为实验对象，将该数据库中历史积累的2 000 GB数据作为测试数据，在Win10系统Matlab 2016a软件中进行测

试。将文献[3]方法和文献[4]方法作为测试的对比方法，与本文提出的基于国产CPU环境的国产数据库历史数据迁移技术共同进行数据迁移测试。分别采用3种不同方法对IDC自建数据库中2 000 GB历史数据进行迁移处理，对历史数据迁移时间、迁移完成后储存空间最大占用率、主动学习评价指标和被访问数据在线概率进行横向对比，以此验证本文提出方法的实际应用性能。

2.1 迁移时间

海量数据在进行迁移时需要耗费一定的时间，耗时越短说明迁移技术稳定性越好，迁移工作效率越高。对比不同方法的数据迁移时间，测试结果如图2所示。

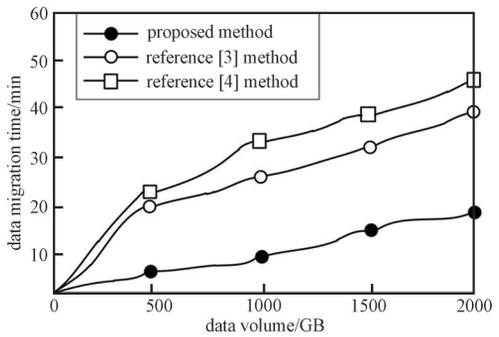

Fig.2 Data migration time

图2 数据迁移时间

通过图2可知，随着数据量的增加，不同方法数据迁移的时间都有所增加。在数据量最大为2 000 GB情况下，采用文献[3]方法进行数据迁移的时间为38 min，采用文献[4]方法进行数据迁移的时间为45 min；而采用所提方法进行数据迁移的时间为18 min，耗时最短，说明所提方法的迁移技术具有更优的稳定性，且迁移工作效率显著提升。这是由于本文采用随机森林算法有效缩减数据迁移量，从而大幅减少数据迁移过程中的耗时，提高了数据迁移的工作效率。

2.2 储存空间最大占用率

数据迁移的成本可依据系统存储空间的最大占用率来衡量，存储空间的最大占用率越低，说明消耗的系统存储空间越小，表明数据迁移的成本越低。对比不同方法的存储空间最大占用率，测试结果如图3所示。

分析图3可知，在数据迁移过程中，不同的数据迁移方法都会在某个特定阶段内，经历系统存储空间最大占用率的波动。采用文献[3]方法进行数据迁移时系统储存空间的最大占用率在35%~60%之间波动，采用文献[4]方法进行数据迁移时系统储存空间的最大占用率在65%~82%之间波动，储存空间最大占用率较高；而采用所提方法进行数据迁移时，系统储存空间最大占用率在10%~25%之间波动，波动较小，并且储存空间最大占用率较低，说明消耗的系统存储空间较小，数据迁移的成本较低。这是因为所提方法利用孤立森林模型剔除了国产数据库中存在的异常数据，减少了需要迁移的数据量，进而降低了系统存储空间的消耗，减小了数据迁移成本。

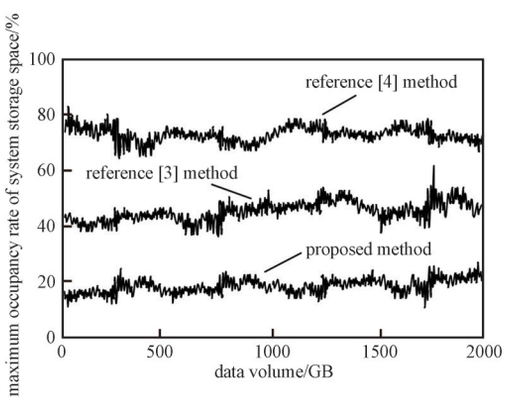

Fig.3 Maximum occupancy of the storage space

图3 存储空间的最大占用率

2.3 ALC指标

ALC指标是指基础模型达到最佳性能时，迁移学习的有效性百分比，用来评价数据迁移学习过程中方法的优劣，ALC指标值越高，表明迁移学习的错误率越低。用$p_i=[p_0,p_1,\cdots,p_N]$表示迁移学习曲线，该迁移学习曲线中存在$N+1$个元素，ALC指标的计算公式如下：

$$ALC=\frac{\sum_{i=1}^{N}(p_{i-1}+p_i)}{2N} \tag{11}$$

采用所提方法、文献[3]方法和文献[4]方法对ALC指标值进行测试，测试结果如图4所示。

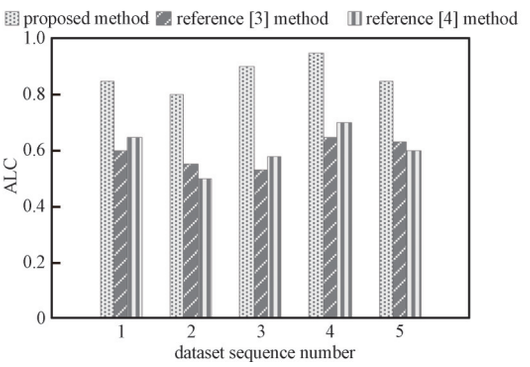

Fig.4 ALC for different methods

图4 不同方法的ALC

分析图4可知，文献[3]方法的ALC指标值为0.55~0.65；文献[4]方法的ALC指标值为0.51~0.70；而所提方法的ALC指标值为0.78~0.95；所提方法在不同数据集中的ALC指标值均高于文献[3]方法和文献[4]方法，表明所提方法在数据迁移学习过程中的错误率较低，可精准地完成国产数据库中历史数据的迁移学习。这是由于本文采用了双

加权策略，能够控制各源数据库的迁移重要性程度以及目标函数的凹凸性，有效降低了迁移学习过程中的错误率。

2.4 被访问数据在线概率

在完成IDC自建数据库中历史数据迁移后，对已迁移数据的被访问情况进行对比。数据在线概率越高表明数据库的访问效率高，即数据迁移效果好。对比不同方法被访问数据在线概率，测试结果如表1所示。

表1 被访问数据在线概率

Table1 Online probability of the accessed data

t / day	online probability of the accessed data/%		
	proposed method	reference [3] method	reference [4] method
10	100	100	100
20	100	90.2	88.9
30	100	84.7	80.2
40	99.7	73.5	71.4
50	98.6	68.4	62.7
60	97.3	59.3	54.8

由表1中的数据可知，随着访问天数的增长，3种方法的被访问数据在线概率呈现出不同程度的降低。文献[3]方法与文献[4]方法的被访问数据在线概率下降速度较快，直至访问60天时，被访问数据在线概率已分别下降至59.3%与54.8%，被访问数据在线概率较低。

而所提方法在访问第40天时，被访问数据在线概率才开始下降，直至访问60天时，被访问数据在线概率下降至97.3%，下降趋势非常小，且在线概率始终保持在97%以上，远高于文献[3]方法和文献[4]方法，验证了所提方法能够确保较高的被访问数据在线概率，表明数据库的访问效率高，数据迁移效果较好。这是由于本文在ε-不敏感损失的基础上，构建历史数据迁移模型，能够保持数据的完整性，提高被访问数据在线概率，具有较高访问效率，数据迁移效果较好。

综合实验测试可得出结论，本文所提的基于国产CPU环境的国产数据库历史数据迁移技术，数据迁移耗时较短，存储空间最大占用率较小，迁移学习过程中的错误率较低，被访问数据在线概率较高，表明数据库的访问效率高，数据迁移成本低，迁移工作效率高，数据迁移效果较好。

3　结论

针对目前数据迁移方法存在的被访问数据在线概率低、存储空间最大占用率较高、数据迁移耗时长和迁移学习错误率高的问题，提出基于国产CPU环境的国产数据库历史数据迁移技术，该方法在国产CPU环境中对数据展开了清洗处理，剔除数据库中存在的异常值，其次建立数据迁移模型，结合交替优化策略完成国产数据库历史数据的迁移，解决了目前方法中存在的问题，提高了国产数据库的访问率。

参考文献：

[1] 张承圣,邵振国,陈飞雄,等. 基于条件深度卷积生成对抗网络的新能源发电场景数据迁移方法[J]. 电网技术, 2022,46(6):2182-2189. (ZHANG Chengsheng,SHAO Zhenguo,CHEN Feixiong,et al. Renewable power generation data transferring based on conditional deep convolutions generative adversarial network[J]. Power System Technology, 2022,46(6):2182-2189.) doi:10.13335/j.1000-3673.pst.2021.1008.

[2] 陈金浩,蒋大鹏,张怡卓,等. 实木板材抗弯强度的SWCSS-GFK-SVM数据迁移建模方法[J]. 光谱学与光谱分析, 2022,42(5):1471-1477. (CHEN Jinhao, JIANG Dapeng, ZHANG Yizhuo, et al. Research on data migration modeling method for bending strength of solid wood based on SWCSS-GFK-SVM[J]. Spectroscopy and Spectral Analysis, 2022,42(5):1471-1477.) doi:10.3964/j.issn.1000-0593(2022)05-1471-07.

[3] 蒲勇霖,于炯,鲁亮,等. Storm平台下的线程重分配与数据迁移节能策略[J]. 软件学报, 2021,32(8):2557-2579. (PU Yonglin, YU Jiong,LU Liang,et al. Energy-efficient strategy based on executor reallocation and data migration in Storm[J]. Journal of Software, 2021,32(8):2557-2579.) doi:10.13328/j.cnki.jos.006074.

[4] 郭辉,芮兰兰,高志鹏. 车辆边缘网络中基于多参数MDP模型的动态服务迁移策略[J]. 通信学报, 2020,41(1):1-14. (GUO Hui,RUI Lanlan,GAO Zhipeng. Dynamic service migration strategy based on MDP model with multiple parameter in vehicular edge network[J]. Journal on Communications, 2020,41(1):1-14.) doi:10.11959/j.issn.1000-436x.2020012.

[5] 杨建,王力,宋冬然,等. 基于孤立森林与稀疏高斯过程回归的风电机组偏航角零点漂移诊断方法[J]. 中国电机工程学报, 2021,41(18):6198-6211. (YANG Jian,WANG Li,SONG Dongran,et al. Diagnostic method of zero-point shifting of wind turbine

yaw angle based on isolated forest and sparse Gaussian process regression[J]. Proceedings of the CSEE, 2021,41(18): 6198–6211.) doi:10.13334/j.0258–8013.pcsee.202224.

[6] 李国成,陆俊,王赟,等. 基于Bagging二次加权集成的孤立森林窃电检测算法[J]. 电力系统自动化, 2022,46(2):92–100. (LI Guocheng,LU Jun,WANG Yun,et al. Isolated–forest electricity theft detection algorithm based on Bagging secondary weighted ensemble[J]. Automation of Electric Power Systems, 2022,46(2):92–100.) doi:10.7500/AEPS20210323005.

[7] 王布宏,罗鹏,李腾耀,等. 基于粒子群优化多核支持向量数据描述的广播式自动相关监视异常数据检测模型[J]. 电子与信息学报, 2020,42(11):2727–2734. (WANG Buhong,LUO Peng,LI Tengyao,et al. ADS–B anomalous data detection model based on PSO–MKSVDD[J]. Journal of Electronics & Information Technology, 2020,42(11):2727–2734.) doi:10.11999/JEIT190767.

[8] 陈利军,王畅. 基于DBSCAN的地震电离层扰动异常数据检测方法[J]. 地震工程学报, 2020,42(2):410–415. (CHEN Lijun, WANG Chang. Detection method for seismic ionospheric disturbance anomaly data based on DBSCAN[J]. China Earthquake Engineering Journal, 2020,42(2):410–415.) doi:10.3969/j.issn.1000–0844.2020.02.410.

[9] 沙林秀,聂凡,高倩,等. 基于布朗运动与梯度信息的交替优化算法[J]. 计算机应用, 2022,42(7):2139–2145. (SHA Linxiu,NIE Fan,GAO Qian,et al. Alternately optimizing algorithm based on Brownian movement and gradient information[J]. Journal of Computer Applications, 2022,42(7):2139–2145.) doi:10.11772/j.issn.1001–9081.2021050839.

[10] 王继豪,王安东,孙福春,等. 基于交替迭代优化的同步调相机电气参数分步辨识方法[J]. 电测与仪表, 2022,59(1):99–105. (WANG Jihao,WANG Andong,SUN Fuchun,et al. A step–by–step identification method of synchronous condenser electrical parameters based on alternative iterative optimization[J]. Electrical Measurement & Instrumentation, 2022,59(1):99–105.) doi: 10.19753/j.issn1001–1390.2022.01.013.

[11] 丁尹,桑楠,李晓瑜,等. 基于循环神经网络的电信行业容量数据预测方法[J]. 计算机应用, 2021,41(8):2373–2378. (DING Yin,SANG Nan,LI Xiaoyu,et al. Prediction method of capacity data in telecom industry based on recurrent neural network[J]. Journal of Computer Applications, 2021,41(8):2373–2378.) doi:10.11772/j.issn.1001–9081.2020101677.

[12] 戴品远,余小金,谢纬华,等. 基于条件高斯贝叶斯网络的代谢组学数据分类预测研究[J]. 中国卫生统计, 2021,38(5):656–660. (DAI Pinyuan,YU Xiaojin,XIE Weihua,et al. Study on metabolomics data classification and prediction based on conditional Gaussian Bayesian network[J]. Chinese Journal of Health Statistics, 2021,38(5):656–660.) doi:10.3969/j.issn.1002–3674.2021.05.004.

[13] 田德艳,张小川,邹司宸,等. L2正则化粒子滤波在水下无人平台纯方位角跟踪的应用[J]. 舰船科学技术, 2020,42(23):111–116. (TIAN Deyan,ZHANG Xiaochuan,ZOU Sichen,et al. Application of L2 regularized particle filter in pure azimuth tracking of underwater unmanned platform[J]. Ship Science and Technology, 2020,42(23):111–116.) doi:10.3404/j.issn.1672–7649.2020.12.022.

[14] 杨婷,朱恒东,马盈仓,等. 基于L_(2,1)范数和流形正则项的半监督谱聚类算法[J]. 山东大学学报(理学版), 2021,56(3):67–76. (YANG Ting,ZHU Hengdong,MA Yingcang,et al. Semi–supervised spectral clustering algorithm based on L_(2,1) norm and regular term of manifold[J]. Journal of Shandong University(Natural Science), 2021,56(3):67–76.)

[15] 雍皓,韩铎,张俊杰,等. 地震数据自适应多层字典学习稀疏表示方法[J]. 石油地球物理勘探, 2022,57(3):525–531. (YONG Hao,HAN Duo,ZHANG Junjie,et al. Sparse representation method for adaptive multilayer dictionary learning of seismic data[J]. Oil Geophysical Prospecting, 2022, 57(3): 525–531.) doi:10.13810/j.cnki.issn.1000–7210.2022.03.003.

作者简介：

毛 冬(1991–)，男，硕士，工程师，主要研究方向为电力信息与通信技术.email:bbbdcf_kk123@163.com.

张 辰(1990–)，男，硕士，主要研究方向为电力信息与通信技术.

陈又咏(1978–)，男，本科，工程师，主要研究方向为信息技术应用创新、电力信息化等.

刘永清(1980–)，男，硕士，教授级高级工程师，主要研究方向为信息技术应用创新、电力信息标准化、科技管理.

焦艳斌(1982–)，男，硕士，工程师，主要研究方向为信息技术应用创新、电力信息标准化等.