

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 8 月 2 日 (02.08.2018)



(10) 国际公布号
WO 2018/137402 A1

(51) 国际专利分类号:
H04L 29/08 (2006.01)

(21) 国际申请号: PCT/CN2017/113854

(22) 国际申请日: 2017 年 11 月 30 日 (30.11.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710064154.0 2017 年 1 月 26 日 (26.01.2017) CN

(71) 申请人: 华南理工大学 (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) [CN/CN]; 中国广东省广州市南沙区环市大道南 25 号华工大广州产研院, Guangdong 511458 (CN)。

(72) 发明人: 刘发贵 (LIU, Fagui); 中国广东省广州市南沙区环市大道南 25 号华工大广州产研院,

Guangdong 511458 (CN)。王彬 (WANG, Bin); 中国广东省广州市南沙区环市大道南 25 号华工大广州产研院, Guangdong 511458 (CN)。

(74) 代理人: 广州粤高专利商标代理有限公司 (YOGO PATENT & TRADEMARK AGENCY LIMITED COMPANY); 中国广东省广州市天河区体育西路中石化大厦 B 塔 4416 室, Guangdong 510620 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

(54) Title: CLOUD DATA CENTRE ENERGY-SAVING SCHEDULING IMPLEMENTATION METHOD BASED ON ROLLING GREY PREDICTION MODEL

(54) 发明名称: 基于滚动灰色预测模型的云数据中心节能调度实现方法

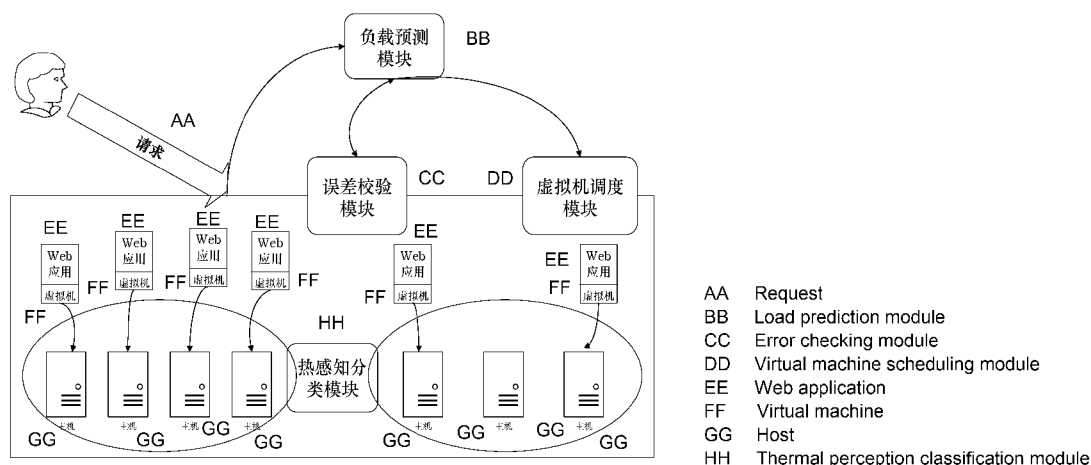


图 2

(57) Abstract: Disclosed is a cloud data centre energy-saving scheduling implementation method based on a rolling grey prediction model. In the present invention, a cloud data centre energy-saving flow is abstracted as four modules of load prediction, error checking, thermal perception classification and virtual machine scheduling. A working load of the data centre at the next moment is predicted by means of the load prediction module to obtain a load utilization rate of each host. The thermal perception classification module can divide the thermal states of all hosts according to predicted values of the load utilization rates of the hosts, wherein the utilization rate of a host in the state of being hotter is at a higher level, while the utilization rate of a host in the state of being cooler is at a lower level. In order that most of the hosts are maintained in the thermal state of being relatively mild, the virtual machine scheduling module performs operations, such as migration and integration, on a virtual machine on each of the hosts according to a classification result of the thermal states, so as to finally achieve the purposes of guaranteeing the service quality of the data centre and reducing the energy

PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

— 发明人资格 (细则4.17(iv))

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

consumption thereof. The present invention overcomes the problem that a traditional grey model has the difficulty of low precision due to the deficiency of some values.

(57) 摘要: 本发明公开了基于滚动灰色预测模型的云数据中心节能调度实现方法。本发明将云数据中心的节能流程抽象为负载预测、误差校验、热感知分类和虚拟机调度四个模块。通过负载预测模块对数据中心下一时刻的工作负载进行预测, 得到各主机的负载利用率; 热感知分类模块会根据主机负载利用率的预测值对所有主机进行热状态的划分, 处于较热状态的主机利用率处于一个较高的水平, 较冷状态的主机利用率较低; 为保持大部分主机处于较为温和的热状态, 虚拟机调度模块将依据热状态的分类结果对各主机上的虚拟机进行迁移整合等操作, 最终达到保障数据中心的的服务质量并且降低其能耗的目标。本发明克服了传统灰色模型由于部分数值匮乏而陷入精度低困境的问题。

基于滚动灰色预测模型的云数据中心节能调度实现方法

技术领域

本发明属于云计算节能调度领域，具体涉及基于滚动灰色预测模型的云数据中心节能调度实现方法。

背景技术

云计算作为信息通信技术行业的新兴技术已逐渐走入千家万户，它以高效益、低门槛、高扩展性的特点受到了广大企业及个体用户的青睐。随着云计算发展的逐渐成熟、用户需求的不断丰富，数据中心服务器等一系列的配套设施规模也在飞速增长。世界各地也都建立起了包含着数以万计服务节点的大规模云计算数据中心，这使得更多的计算资源及存储资源被保存在云端，但由此也引发了一系列的能耗问题。绿色和平组织在其云计算报告中称，预计到 2020 年，全球主要 IT 运营商的数据中心的能耗将接近 19630 亿千瓦时，这将超过目前德国、法国、加拿大和巴西消耗的电力总量总和。与此同时，中国政府工作报告指出在 2008 年的主要任务是“加大节能减排和环境保护力度，科技部与工信部于 2014 年制定了《2014-2015 年节能减排科技专项行动方案》，来发挥科技缓解资源环境约束的引领作用。日益不断攀升的能耗开销在加重了企业负担的同时，也增加了全球范围内影响气候变化的碳排放量，给环境问题敲响了警钟。

这些集群服务器快速攀升的能耗已经成为影响企业效益及其发展的重要因素。云计算节能领域的研究也成为了国内外新兴技术研究的重点任务。目前国内外对于云计算节能研究的出发点各不相同，也存在着一定的不足，目前对于云数据中心在流量突发情况下的工作负载监控、预测问题仍存在着较大的不足，对虚拟机的调度整合策略存在着缺陷。

发明内容

为了更加智能有效地在网络流量突发情况下减少数据中心的功耗并保障其服务质量，本发明提供基于滚动灰色预测模型的云数据中心节能调度实现方法。

本发明具体通过如下技术方案实现。

基于滚动灰色预测模型的云数据中心节能调度实现方法，其将云数据中心的节能流程抽象为负载预测、误差校验、热感知分类和虚拟机调度四个模块；通过对数据中心工作负载进行预测，从而对各主机进行状态分类，进而通过虚拟机调度算法达到节能的目的。

进一步地，所述负载预测具体是：利用滚动灰色预测模型对数据中心的工作负载进行预测，得到下一时刻数据中心各主机节点的负载利用率。

进一步地，所述误差校验具体是：对负载预测值与实际工作负载进行误差校验，确定当前预测结果的偏差值，并基于误差校验模块进行学习，校正预测结果。

进一步地，所述热感知分类具体是：根据主机当前的负载预测值对云数据中心内所有主机进行热感知分类，并引入服务水平协议作为参考指标设定主机工作负载阈值的上界和下界；当主机负载利用率在高于阈值上界、低于阈值下界、处于阈值上下界之间、负载利用率为 0 这四种情况下划分为不同的四种热状态。

进一步地，所述虚拟机调度具体是：根据当前主机的热状态对其进行虚拟机调度，通过虚拟机调度操作解决主机过载、空载的问题，将数据中心各主机的维持在一个健康的热状态下。

进一步地，通过基于滚动灰色预测模型算法，并以模块化的形式实现到云数据中心环境中，来智能地监控数据中心的负载信息、以及运行时的各性能数据。

与现有技术相比，本发明具有如下优点和技术效果：

本发明通过挖掘并分析数据中心实时工作负载数据，利用基于灰色预测模型的工作负载预测算法建立负载预测评估模型，对数据中心中各服务器下一时刻的负载状态进行预测，避免服务器的过载或空载现象，更加有效地应对目前较为普遍的网络流量突发现象。通过分析负载与能效的关联关系，建立主机的利用率热感知分类模型；并提出基于热感知机制的主机分类标准，来解决数据中心各主机直接负载不均导致的能耗性能问题。

本发明通过基于滚动灰色预测模型算法，并以模块化的形式实现到云数据中心环境中，来智能地监控数据中心的负载信息、以及运行时的各性能数据。减少数据中心工作时尤其是流量突发情况下的能耗水平，降低云服务提供商的运维成本，与此同时提升数据中心 SLA 的指标，保障用户的云服务使用体验。

附图说明

图 1 为云计算智能资源调度框架图。

图 2 为云数据中心系统节能框架图。

图 3 为云数据中心任务序列 1 的负载预测实验结果图。

图 4 为云数据中心任务序列 2 的负载预测实验结果图。

图 5 为滚动灰色预测模型和 ARIMA 模型实验数据对比图。

图 6 为滚动灰色预测模型和 ARIMA 模型平均偏差对比图。

具体实施方式

为了使本发明的技术方案及优点更加清楚明白，以下结合附图，进行进一步的详细说明，但本发明的实施和保护不限于此。

1. 策略框架

1.1 云计算资源智能调度框架

图 1 是云计算平台资源智能调度框架的架构图，从下而上分为主机层、虚拟机层、性能评估层、调度层和用户层。其中调度层和评估层是整个节能策略框架的核心所在。下面将对每一层进行阐述。

主机层是指云数据中心中的所有服务器，包含所有物理主机节点。这些硬件设备是云环境最底层的基础设施，为我们进行节能调度管理提供了硬件基础。

虚拟机层则是建立在主机层虚拟化技术基础上的，通过将多台服务器实体虚拟化，构成了虚拟机层这个资源池，能够实现云环境中的共同计算及资源共享。

性能评估层是指对云数据中心的负载利用率、能耗、SLA、PUE 等性能数据进行采集评估。评估层需要与虚拟层进行数据通信来获取虚拟机资源利用情况、各虚拟机的运行状态等信息。

调度层则在基于性能评估层所采集负载、能耗等数据的基础上，对主机进行虚拟机初次分配，以及虚拟机迁移操作，通过对虚拟机进行调度来保障主机能够运行在一个良好的负责利用率环境下。

用户层是指云计算环境中的所有用户及服务请求者，包括个人用户、企业用户以及云计算的所有使用者。用户层会时刻对数据中心发出新的服务请求。

1.2 云数据中心系统节能框架

图 2 是云计算数据中心系统节能架构的框架图，从上而下分为四个模块，分别是负载预测模块、误差校验模块、热感知分类模块以及虚拟机调度模块。下面将对每一个模块进行阐述。

(1) 负载预测模块：数据中心每秒都会处理成千上万的服务请求，负载预测模块可以持续地监测数据中心的物理机（Power Machine）的工作负载数据，并通过分析有效历史负载数据来预测每一个 PM 在未来时刻的 CPU 利用率。负载预测模块可以帮助我们有效地区

分数据中心当前状态下过载与空载的服务器。

(2) 误差校验模块：当负载预测器完成了预测过程之后，误差校验模块会计算实际值与预测值之间的偏差，并通过分析和计算它们的相对误差来优化未来的预测结果。

(3) 热感知分类模块：根据上述模块得到的工作负载预测值，我们将物理机分为了四类类别，称为沸点 (boiling point)、温点 (warming point)、凉点 (cooling point)、冰点 (freezing point)。我们在此节能架构下引入了服务水平协议 (Service Level Agreement, SLA) 作为我们的一项重要参考指标，并利用 SLA 设定了 PM 工作负载阈值的上界和下界。当 PM 当前的负载利用率高于阈值上界时我们称之为沸点；当其利用率低于阈值下界（非 0）时我们称之为凉点；负载利用率为 0 时为冰点；当 PM 的负载利用率处于阈值上下界之间时我们称之为温点。热感知分类模块通过对 PM 进行类别区域的划分，我们可以很好地了解并调控当前数据中心内部的负载情况。

(4) 虚拟机调度模块：云计算节能的目的是使得整个数据中心能够以较高的利用率运转，并且最大程度保障用户的云服务质量。因此我们需要使得尽可能多的 PM 以温点的形式运转。因此我们需要转换更多的沸点 PM 到温点之中，并尽可能地整合一些凉点或者冰点的 PM。虚拟机调度模块会将部分运行在沸点上的虚拟机迁移至凉点之中，部分容量足够大的凉点将会被动态整合为一体。

2. 负载预测

2.1 数据中心能效指标

本文提出的云数据中心环境下的能效指标更为丰富，表 1 列出了需要获取的性能指标数据。主要有三种指标类型，分别是时间类型、环境类型和性能指标。

表 1

指标类型	属性	符号	类型
时间	数据中心的运行时间	t_{interval}	整数
	数据中心的调度周期	t_{sim}	整数
环境	云任务列表的长度大小	$N_{\text{task_size}}$	整数
	数据中心主机数量	N_{host}	整数
	数据中心虚拟机数量	N_{vms}	整数
	虚拟机的核心数	$N_{\text{vi-core}}$	整数

	主机的核心数	$N_{hi-core}$	整数
	虚拟机 i 每秒钟执行的百万指令数	N_{vi_MIPS}	整数
	主机 i 每秒钟执行的百万指令数	N_{hi_MIPS}	整数
性能	第 i 个主机在 t_j 时刻的利用率	$U_{hi,t1}$	浮点数
	第 i 个主机在 t_1-t_2 时间内的能效	$E_{hi,t1,t2}$	浮点数
	数据中心运行总能效	E_d	浮点数
	数据中心在 t_1-t_2 段时间内的能效	$E_{t1,t2}$	浮点数
	符合服务水平协议 (SLA) 的比例	R_{SLA}	浮点数
	违反服务水平协议 (SLA) 的比例	R_{SLA-V}	浮点数
	违反服务水平协议 (SLA) 的平均比例	$R_{A-SLA-V}$	浮点数
	被关闭的主机数量	$N_{hostoff}$	整数
	虚拟机发生迁移总次数	N_{mig}	整数
	主机 i 当前利用率所对应的能耗	M_{hi}	双精度型

时间类型是具有重大意义的属性，它决定了测试过程中数据中心的运行时间以及调度时间间距。环境类型定义了数据中心中各主机以及虚拟机的具体配置参数，以及当前处理的云任务列表的长度大小，这些指标共同决定了数据中心运作过程中的客观状态。性能指标主要反映当前数据中心的运转情况，包括主机在各个时刻下的利用率、能耗数据，数据中心在某时间区间下的能效，数据中心符合及违反服务水平协议的主机比例，调度过程中被关闭的主机数量，调度过程发生的虚拟机迁移总数等数据。

2.2 滚动灰色预测模型

面对流量突发现象，需要实时准确地分析处理海量的网络业务量数据，这具有很大的挑战性。而当流量突发情况发生时，数据中心中沸点及凉点的出现频率会更加频繁，服务器过载现象更为严重。本文提出一种应对流量突发情况的数据中心工作负载预测方法。

数据中心的工作负载变化具有一定的波动性与不确定性，建立工作负载的短期预测模型属于少数据建模问题，而灰色模型相比别的模型更适合解决少数据、贫信息状态预测问题，通过少量的数据便可建立工作负载的残差状态预测模型，适合用于数据中心工作负载的短期预测模型的建立中去。

该工作负载预测方法基于改进后的灰色模型提出，灰色预测方法被视为是时间序列预

测模型很好的替代策略,在云环境中,我们将数据中心中各 PM 的历史工作负载值作为历史数据序列,灰色模型使用了一次微分方程来描述整个系统的建立过程。

设 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ 为原始工作负载数列,其中 n 为数据序列的长度,其 1 次累加生成数列为 $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, 其中

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n,$$

定义 $x^{(1)}$ 的灰导数为

$$d(k) = x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1).$$

令 $z^{(1)}$ 为数列 $x^{(1)}$ 的邻值生成数列, 即

$$z^{(1)}(k) = \alpha x^{(1)}(k) + (1 - \alpha)x^{(1)}(k-1),$$

其中, α 为差异系数, $\alpha \in (0, 1)$ 。

于是定义 GM (1, 1) 的灰微分方程模型为

$$d(k) + az^{(1)}(k) = b, \text{ 或 } x^{(0)}(k) + az^{(1)}(k) = b,$$

$$x^{(0)}(k) + az^{(1)}(k) = b,$$

在式 (1) 中, $x^{(0)}(k)$ 称为灰导数, a 称为发展系数, $z^{(1)}(k)$ 称为白化背景值, b 称为灰作用量。

将时刻表 $k=2, 3, \dots, n$ 代入 (1) 式并引入矩阵向量记号:

$$\begin{cases} x^{(0)}(2) + az^{(1)}(2) = b, \\ x^{(0)}(3) + az^{(1)}(3) = b, \\ \dots \dots \dots \\ x^{(0)}(n) + az^{(1)}(n) = b, \end{cases}$$

$$u = \begin{bmatrix} a \\ b \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

则 GM(1, 1) 模型可以表示为矩阵 $Y = Bu$. 其中 u 为 a 和 b 的拟合系数矩阵, B 为原始序列提供的信息矩阵。

用回归分析求得 a, b 的估计值, 于是数列相应的白化模型为

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b,$$

$$\text{解得 } x^{(1)}(t) = (x^{(0)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a}.$$

于是得到预测值

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n-1,$$

从而相应地得到下一时刻工作负载预测值:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k = 1, 2, \dots, n-1,$$

传统的 GM(1, 1) 模型很有可能会由于数据及部分数值的匮乏而陷入预测精度低的困境。采用滚动 GM(1, 1) 模型后, 每进行下一次预测都将重新建立预测模型, 模型会不断采用较新的数据进行预测, 过旧的数据将会被摒弃。利用前 100 个调度周期历史数据预测第 101 个周期结果, 并不断更新, 可以大大提升预测的精度。我们将根据数据中心历史负载动态更新模型中平移变换常量 C 的值, 确保预测过程中灰色模型始终处于可用状态。通过迭代

选出序列 $\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 2, 3, \dots, n$ 中 $\lambda(k)$ 的最大值 $Max\lambda(k)$ 以及最小值 $Min\lambda(k)$, 确保 $Min\lambda(k)$ 和 $Max\lambda(k)$ 均在可容空间 $X = (e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}})$ 内。

3. 基于热感知的虚拟机分类策略

传统的虚拟机调度策略的建立大多基于启发式算法 (遗传算法、模拟退火、粒子群算法等), 而启发式算法会无法避免陷入局部最优解的问题, 启发式算法本质上是一种贪心策略, 导致不符合贪心规则的最优解会被错过。基于热感知的虚拟机分类调度策略根据主机的“冷热”状态对物理节点之间的虚拟机进行重分配, 可以有效减少数据中心中的过载与空载主机数, 达到能耗与 SLA 的平衡。

热感知模型持续监控数据中心物理节点的热状态信息, 将当前物理节点根据“冷热”状态划分为四种类型, 分别为沸点 (boiling point)、温点 (warming point)、凉点 (cooling point)、冰点 (freezing point), 并利用 SLA 设定了 PM 工作负载阈值的上界和下界。当 PM 当前的负载利用率高于阈值上界时我们称之为沸点; 当其利用率低于阈值下界 (非 0) 时我们称之为凉点; 负载利用率为 0 时为冰点; 当 PM 的负载利用率处于阈值上下界之间时我们称之为温点。我们需要使得尽可能多的 PM 以温点的形式运转。因此我们需要转换更多

的沸点 PM 到温点之中, 并尽可能地整合一些凉点或者冰点的 PM。虚拟机调度模块会将部分运行在沸点上的虚拟机迁移至凉点之中, 部分容量足够大的凉点将会被动态整合为一体, 处于冰点的物理机将会被关闭以节能。

4. 实验说明

对于滚动灰色预测模型来说, 我们首先需要先让数据中心运行一段时间 t , 获取一定数量的历史负载数据作为输入样本提供给滚动灰色预测模型学习。时间 t 越大, 滚动灰色预测模型预测的结果就越准, 此处实验 t 取 100 个采样周期长度。

滚动灰色预测模型实验测试的平均绝对偏差值与平均绝对百分误差的计算方法如定义如下:

$$\text{平均绝对偏差} = \frac{1}{n} \sum_{k=1}^n |x^{(0)}(k) - \hat{x}^{(0)}(k)|, \quad (k = 1, 2, \dots, n)$$

$$\text{平均绝对百分比误差} = \frac{1}{n} \sum_{k=1}^n \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right|, \quad (k = 1, 2, \dots, n)$$

其中 $x^{(0)}(k)$ 为负载利用率的实际值, $\hat{x}^{(0)}(k)$ 为滚动灰色预测模型的预测结果。

我们采集了数据中心 100 个采样周期内的负载数据作为滚动灰色预测模型的初始输入序列, 紧接着运行滚动灰色预测模型算法, 算法通过对前 100 个采样周期的数据序列进行计算预测得到第 101 个周期时刻的负载数据, 再对比计算第 101 个周期时刻的工作负载实际值与预测值的平均绝对偏差和平均绝对百分误差来调整原有的滚动模型, 并更新输入数据序列。之后依此预测第 102、103 及以后周期的工作负载。

图 3 和图 4 是对两组云任务序列进行工作负载预测的实验结果, 其中图 3 的云任务序列在第 1-5 小时、第 22-24 小时发生了流量突发事件, 图 4 的云任务序列在第 10 小时和第 20 小时发生了流量突发事件。实验用滚动灰色预测模型与自回归移动平均 (ARIMA) 模型以及实际负载值进行了对比, 两组实验结果都显示滚动灰色预测模型对于数据中心的工作负载的预测精度更为接近实际值 (Actual), 相比 ARIMA 模型, 滚动灰色预测模型对于流量突发情况下的反馈更为及时。

图 5 是在运行图 4 所示云任务序列之后的实验结果, 结果显示滚动灰色预测模型的在云数据中心的负载预测过程中的平均偏差比为 6.93%, 而 ARIMA 模型的平均偏差比为 10.35%, 滚动灰色预测模型的整体预测效果更优。图 6 为滚动灰色预测模型与 ARIMA 模型实验后的偏差比面积图, 若折线与 X 轴围成的面积越小, 则代表其预测偏差越小, 精度越高。如图

所示 ARIMA 模型的偏差比面积要远大于滚动灰色预测模型，更凸显出滚动灰色预测模型在云数据中心负载预测中的优势。

权利要求书

1. 基于滚动灰色预测模型的云数据中心节能调度实现方法，其特征在于将云数据中心的节能流程抽象为负载预测、误差校验、热感知分类和虚拟机调度四个模块；通过对数据中心工作负载进行预测，从而对各主机进行状态分类，进而通过虚拟机调度算法达到节能的目的。

2. 根据权利要求1所述的基于滚动灰色预测模型的云数据中心节能调度实现方法，其特征在于所述负载预测具体是：利用滚动灰色预测模型对数据中心的工作负载进行预测，得到下一时刻数据中心各主机节点的负载利用率。

3. 根据权利要求1所述的基于滚动灰色预测模型的云数据中心节能调度实现方法，其特征在于所述误差校验具体是：对负载预测值与实际工作负载进行误差校验，确定当前预测结果的偏差值，并基于误差校验模块进行学习，校正预测结果。

4. 根据权利要求1所述的基于滚动灰色预测模型的云数据中心节能调度实现方法，其特征在于所述热感知分类具体是：根据主机当前的负载预测值对云数据中心内所有主机进行热感知分类，并引入服务水平协议作为参考指标设定主机工作负载阈值的上界和下界；当主机负载利用率在高于阈值上界、低于阈值下界、处于阈值上下界之间、负载利用率为0这四种情况下划分为不同的四种热状态。

5. 根据权利要求1所述的基于滚动灰色预测模型的云数据中心节能调度实现方法，其特征在于所述虚拟机调度具体是：根据当前主机的热状态对其进行虚拟机调度，通过虚拟机调度操作解决主机过载、空载的问题，将数据中心各主机的维持在一个健康的热状态下。

6. 根据权利要求1~5任一项所述的基于滚动灰色预测模型的云数据中心节能调度实现方法，其特征在于通过基于滚动灰色预测模型算法，并以模块化的形式实现到云数据中心环境中，来智能地监控数据中心的负载信息、以及运行时的各性能数据。

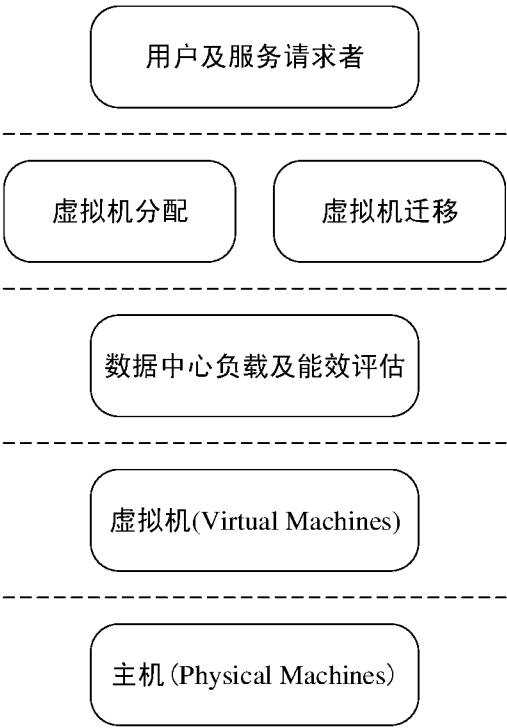


图 1

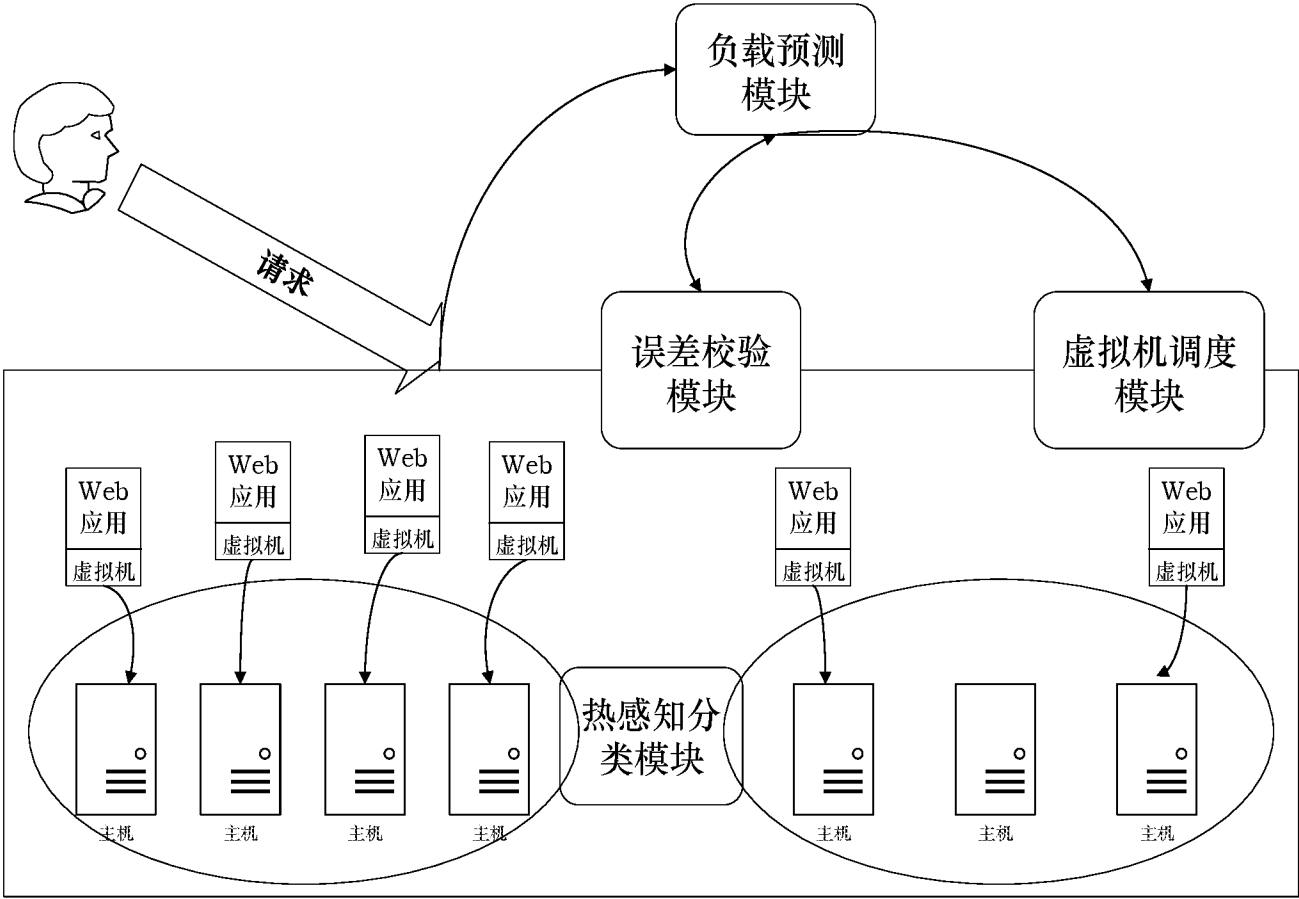


图 2

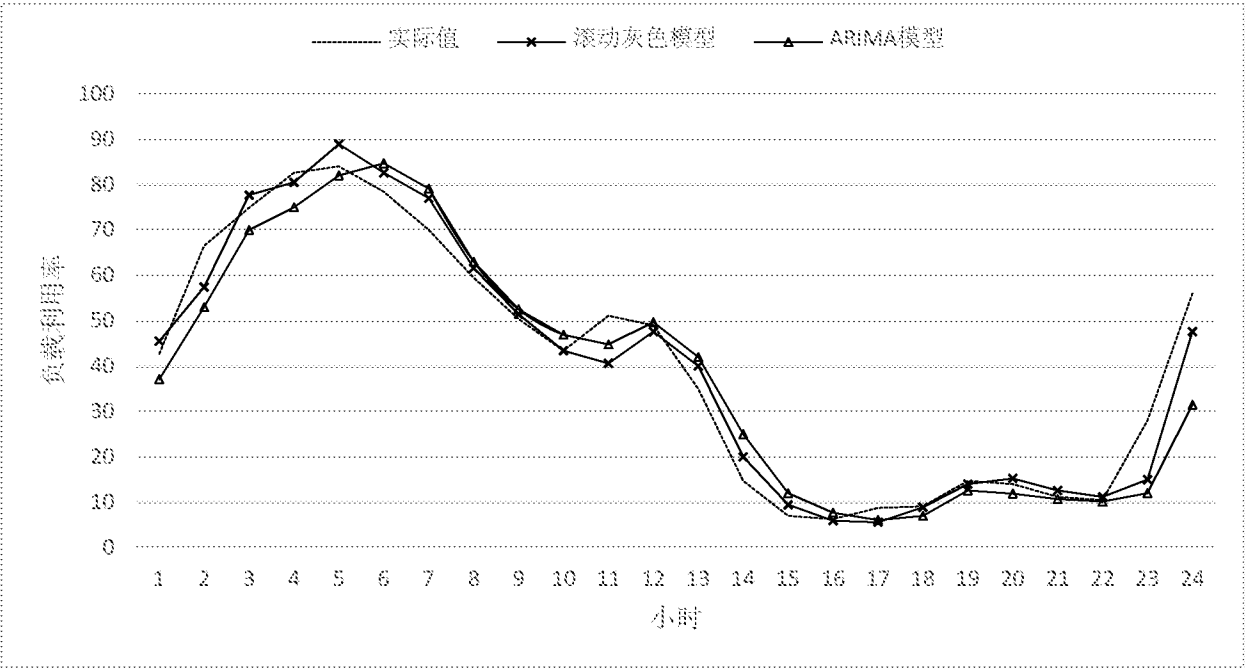


图 3

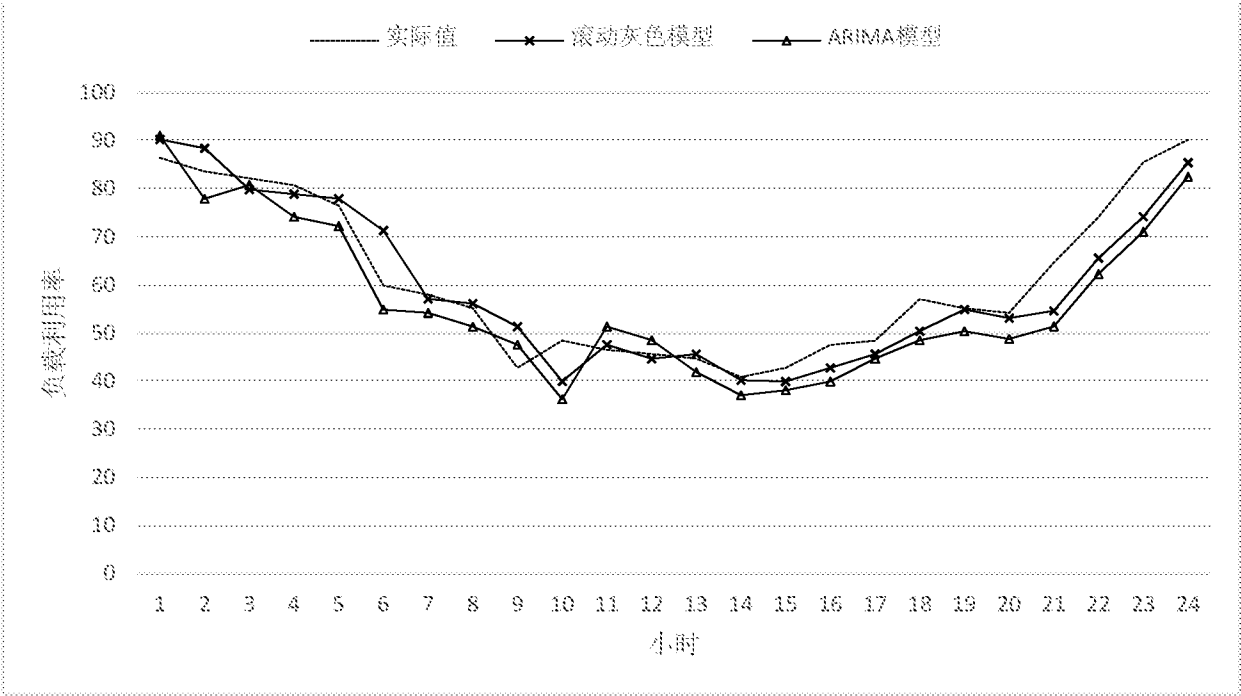


图 4

小时	实际值	ARIMA模型			滚动灰色模型		
		预测值	偏差值	偏差比	预测值	偏差值	偏差比
1	86.42	91.02	4.6	5.3228419	90.25	3.83	4.4318445
2	83.6	77.9	-5.7	6.8181818	88.41	4.81	4.5813397
3	82.17	80.75	-1.42	1.7281246	79.84	-2.33	2.8355848
4	80.74	74.15	-6.59	8.1620015	78.86	-1.88	2.3284617
5	76.46	72.22	-4.24	5.5453832	77.94	1.48	1.9356526
6	59.87	54.86	-5.01	8.368131	71.28	11.41	19.057959
7	57.98	54.19	-3.79	6.5367368	57.07	-0.91	1.5695067
8	55.14	51.32	-3.82	6.9278201	56.12	0.98	1.7772942
9	42.73	47.57	4.84	11.326937	51.33	8.6	20.126375
10	48.42	36.22	-12.2	25.1962	39.94	-8.48	17.513424
11	46.51	51.34	4.83	10.384863	47.58	1.07	2.3005805
12	45.63	48.52	2.89	6.3335525	44.65	-0.98	2.1477098
13	44.68	41.88	-2.8	6.266786	45.6	0.92	2.0590868
14	40.84	37.05	-3.79	9.2801175	40.19	-0.65	1.5915769
15	42.76	38.08	-4.68	10.944808	39.92	-2.84	6.6417212
16	47.52	39.91	-7.61	16.01431	42.75	-4.77	10.037879
17	48.4	44.66	-3.74	7.7272727	45.61	-2.79	5.7644628
18	57.05	48.5	-8.55	14.986854	50.38	-6.67	11.691499
19	55.11	50.36	-4.75	8.6191254	54.89	-0.22	0.3992016
20	54.18	48.78	-5.4	9.9667774	53.1	-1.08	1.9933555
21	64.62	51.37	-13.25	20.504488	54.63	-9.99	15.45961
22	74.1	62.29	-11.81	15.937922	65.56	-8.54	11.524966
23	85.49	71.06	-14.43	16.879167	74.17	-11.32	13.241315
24	90.18	82.51	-7.67	8.5052118	85.41	-4.77	5.2894212
平均偏差比(%)		10.345151			6.9291595		

图 5

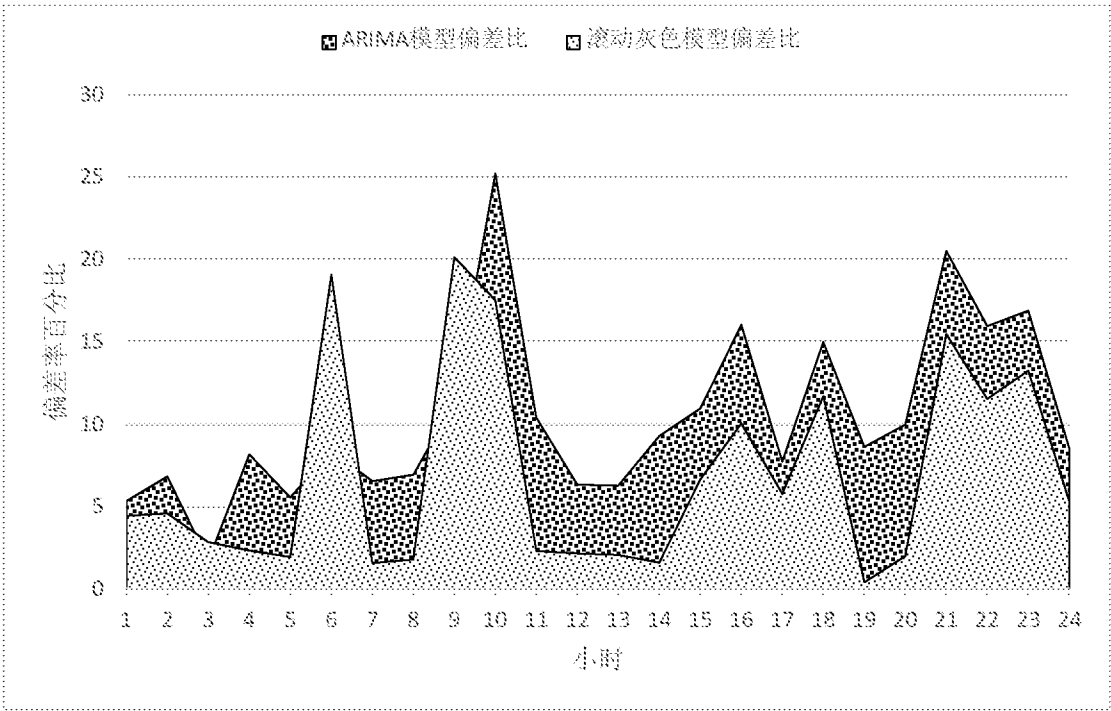


图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/113854

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/08 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI: 灰色, 调度, 预测, 预估, 平衡, 储存, 存储, 云, 估计, 均衡, 负载, 数据, 误差, 利用率

VEN; WOTXT: grey, dispatch, forecast, calculate, balance, store, cloud, estimate, load, data, error, usage ratio, use efficiency

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106899660 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY), 27 June 2017 (27.06.2017), claims 1-6	1-6
Y	CN 106020934 A (INSPUR ELECTRONIC INFORMATION INDUSTRY CO., LTD.), 12 October 2016 (12.10.2016), description, paragraphs 0007-0017	1-6
Y	CN 102833326 A (GUANGDONG UNIVERSITY OF TECHNOLOGY), 19 December 2012 (19.12.2012), description, paragraphs 0003-0011	1-6
Y	CN 103916438 A (SHANGHAI DEVELOPMENT CENTRE OF COMPUTER SOFTWARE TECHNOLOGY et al.), 09 July 2014 (09.07.2014), description, paragraphs 0006-0015	1-6
Y	张磊, 徐达宇, “基于多步优化 GM(1,1)模型的云计算资源负荷短期预测”, 计算机工程与应用, 2014(10), 31 May 2014 (31.05.2014), ISSN: 1002-8331, chapter 2.1, (ZHANG, Lei; XU, Dayu; “Multi-step Optimized GM(1,1) Model-based Short Term Resource Load Prediction in Cloud Computing”, Computer Engineering and Applications)	1-6
A	EP 2570922 A1 (ALCATEL LUCENT), 20 March 2013 (20.03.2013), entire document	1-6

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	
“E” earlier application or patent but published on or after the international filing date	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	“&” document member of the same patent family

Date of the actual completion of the international search 12 January 2018	Date of mailing of the international search report 22 February 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer LUO, Fangjie Telephone No. (86-10) 62089981

International application No.
PCT/CN2017/113854

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/113854

A. 主题的分类

H04L 29/08(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;CNKI:灰色, 调度, 预测, 预估, 平衡, 储存, 存储, 云, 估计, 均衡, 负载, 数据, 误差, 利用率
 VEN;WOTXT:gray, dispatch, forecast, calculate, balance, store, cloud, estimate, load, data, error,
 useage ratio, use efficiency

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 106899660 A (华南理工大学) 2017年 6月 27日 (2017 - 06 - 27) 权利要求1-6	1-6
Y	CN 106020934 A (浪潮电子信息产业股份有限公司) 2016年 10月 12日 (2016 - 10 - 12) 说明书第0007-0017段	1-6
Y	CN 102833326 A (广东工业大学) 2012年 12月 19日 (2012 - 12 - 19) 说明书第0003-0011段	1-6
Y	CN 103916438 A (上海计算机软件技术开发中心 等) 2014年 7月 9日 (2014 - 07 - 09) 说明书第0006-0015段	1-6
Y	张磊, 徐达宇. “基于多步优化GM(1, 1)模型的云计算资源负荷短期预测” 计算机工程与应用, 第2014卷, 第10期, 2014年 5月 31日 (2014 - 05 - 31), ISSN: 1002-8331, 第2.1节	1-6
A	EP 2570922 A1 (ALCATEL LUCENT) 2013年 3月 20日 (2013 - 03 - 20) 全文	1-6

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2018年 1月 12日

国际检索报告邮寄日期

2018年 2月 22日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
 中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

罗芳洁

电话号码 (86-10)62089981

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/113854

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	106899660	A	2017年 6月 27日	无			
CN	106020934	A	2016年 10月 12日	无			
CN	102833326	A	2012年 12月 19日	无			
CN	103916438	A	2014年 7月 9日	CN	103916438	B	2017年 4月 12日
EP	2570922	A1	2013年 3月 20日	WO	2013037619	A1	2013年 3月 21日

表 PCT/ISA/210 (同族专利附件) (2009年7月)