(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104063747 A
(43)申请公布日 2014.09.24

(21)申请号 201410294472.2

(22)申请日 2014.06.26

(71)申请人 上海交通大学
地址 200240 上海市闵行区东川路800号

(72)发明人 曹健 杨定裕 仇沂 顾骅珊 沈琪骏 王烺

(74)专利代理机构 上海汉声知识产权代理有限公司 31236
代理人 胡晶

(51)Int.Cl.
*G06Q 10/04*(2012.01)

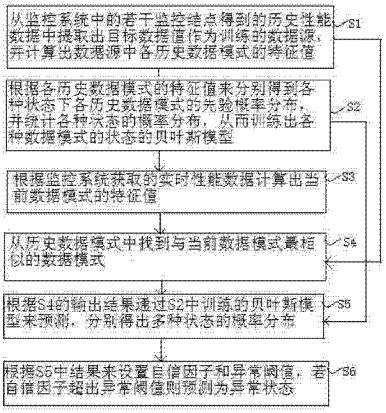
权利要求书2页 说明书10页 附图2页

(54)发明名称

一种分布式系统中的性能异常预测方法及系统

(57)摘要

本发明涉及一种分布式系统中的性能异常预测方法及系统，通过分布式环境的监控系统采集历史性能数据以及实时性能数据，采用特征值提取描述数据的特征，构建出性能变量的模式，并通过朴素贝叶斯分类训练出分类模型，由当前数据模式从历史数据模式进行比较，并在历史数据模式中找到一个与当前数据模式最相似的模式，最后根据贝叶斯预测模型推断出当前数据模式是否是异常状态。本发明本针对分布式系统中的性能异常预测，全面考虑变量的特征的问题，准确率更高，采用机器学习方法贝叶斯模型来指导预测，并实时检测出性能异常情况，并对检测出的预测通过之前得出的贝叶斯模型进行了评估分析，提供了预测的置信度，自动化程度高，提高了预测的可靠性与实用性。




CN 104063747 A

1. 一种分布式系统中的性能异常预测方法，其特征在于，包括以下步骤：

S1：从监控系统中的若干监控结点得到的历史性能数据中提取出目标数据值作为训练的数据源，并计算出数据源中各历史数据模式的特征值；

S2：根据各历史数据模式的特征值来分别得到各种状态下各历史数据模式的先验概率分布，并统计各种状态的概率分布，从而训练出各种数据模式的状态的贝叶斯模型；

S3：根据监控系统获取的实时性能数据计算出当前数据模式的特征值；

S4：从所述历史数据模式中找到与当前数据模式最相似的数据模式；

S5：根据S4的输出结果通过S2中训练的贝叶斯模型来预测，分别得出多种状态的概率分布；

S6：根据S5中结果来设置自信因子和异常阈值，若自信因子超出异常阈值则预测为异常状态。

2. 如权利要求1所述的一种分布式系统中的性能异常预测方法，其特征在于，

所述特征值包括性能值变化量、性能值变化率和性能值。

3. 如权利要求1所述的一种分布式系统中的性能异常预测方法，其特征在于，S2中将所有历史数据模式的各种特征值方差按取值大小排列，并划分为若干子空间，计算各子空间所对应的特征值方差的特定状态的先验概率。

4. 如权利要求3所述的一种分布式系统中的性能异常预测方法，其特征在于，S2中根据所述各历史数据模式的特征值训练出各历史数据模式的贝叶斯模型，分别得到各模式的多种状态的先验概率。

5. 如权利要求3所述的一种分布式系统中的性能异常预测方法，其特征在于，S4中进一步包括：

计算当前数据模式与各历史正常模式之间的特征值的标准方差；

得出与当前数据模式所有标准方差之和最小的历史数据模式为当前数据模式的最相似模式。

6. 如权利要求3所述的一种分布式系统中的性能异常预测方法，其特征在于，所述状态为异常状态，警告状态以及正常状态。

7. 如权利要求3所述的一种分布式系统中的性能异常预测方法，其特征在于，S6中还包括设置报警阈值，若自信因子在报警阈值与异常阈值之间，则预测为报警状态，若自信因子小于报警阈值则预测为正常状态。

8. 一种分布式系统中的性能异常预测系统，与分布式系统的监控系统相连，其特征在于，包括：

历史特征值计算模块，从监控系统中的若干监控结点得到的历史性能数据中提取出目标数据值作为训练的数据源，并计算出数据源中各历史数据模式的特征值；

先验概率模块，与历史特征值计算模块的输出端相连接，根据各历史数据模式的特征值来分别得到各种状态下各历史数据模式的先验概率分布，并统计各种状态的概率分布，从而训练出各种数据模式的状态的贝叶斯模型；

实时特征值计算模块，根据监控系统中的若干监控结点获取的实时性能数据计算出当前数据模式的特征值；

相似模式模块，与历史特征值计算模块的输出端以及实时特征计算模块相连接，从所

述历史数据模式中找到与当前数据模式最相似的数据模式；

概率计算模块，根据相似模式模块的输出结果通过先验概率模块中训练的贝叶斯模型来预测，分别得出所述多种状态的概率分布；以及

异常报警模块，根据概率计算模块中结果来设置自信因子和异常阈值，若自信因子超出异常阈值则预测为异常状态。

9. 如权利要求 8 所述的一种分布式系统中的性能异常预测系统，其特征在于，所述特征值包括性能值变化量、性能值变化率和性能值。

10. 如权利要求 8 所述的一种分布式系统中的性能异常预测系统，其特征在于，所述状态包括异常状态，警告状态以及正常状态。

# 一种分布式系统中的性能异常预测方法及系统

## 技术领域

[0001] 本发明涉及一种性能异常检测预测方法及系统，尤其涉及一种分布式系统中的性能异常预测方法及系统。

## 背景技术

[0002] 在分布式系统中，各个计算机是相互独立的，可以是物理上相邻的，也可以是地理上分散的，它们通过网络或者其他方式进行连接，组成一个整体。从研究上来讲，分布式计算具有以下特点：1. 资源共享；2. 可伸缩性；3. 容错性；4. 并发性。

[0003] 为了更好地体现分布式计算的强大的处理数据计算的能力，对分布式计算环境进行监控将变得尤为重要和关键。系统必须协调这些任务的运行、合理分配资源使资源得到充分的利用并提升整个系统的性能。通常情况下，系统采用调度程序来管理这些任务。调度程序会采集系统中各种资源的相关信息以确定资源是否可用，然后调度算法根据资源的可用性、任务的运行时间等来确定任务的优先级并分配给它们可用的资源。然而随着任务的运行，各种资源的状态，如 CPU 负载、剩余内存、硬盘剩余空间等会随时发生改变，如果在实行调度之前，就能预测到资源在未来某个时间是否依然可用，并合理地避开异常时段对资源的使用，那么系统的调度结果将更加理想。因此，对系统中的资源进行实时监控，并在异常发生之前探测到异常的预兆具有重要的意义。

[0004] 系统性能异常是指在软件运行期间，由于资源逐渐耗尽或者运行错误逐渐累积所导致的计算机系统性能逐渐下降，最终下降到人们所不能容忍的程度的现象。系统性能异常通常是系统状态行为（如，CPU 负载，内存使用率等）不能维持现有的应用程序工作。大多数异常预测模型都只是基于回归技术的模型，而回归技术具有其特定的局限性，因此此类模型存在着各自的缺陷，或者只适用于特定的数据，或者预测误差较大等。而基于已存在的基于分类的异常预测模型，仍需要人工对历史数据分配标识，自动化程度不高，并且只是从变量值的角度去观察，不能全面考虑变量的特征，因此预测结果会存在一定的误差。

## 发明内容

[0005] 本发明的目的在于提供一种分布式系统中的性能异常预测方法及系统，解决了对分布式环境性能预测自动化程度不高、只是从变量值的角度去观察而不能全面考虑变量的特征的问题。

[0006] 为了解决上述问题，本发明涉及了一种分布式系统中的性能异常预测方法，包括以下步骤：

[0007] S1：从监控系统中的若干监控结点得到的历史性能数据中提取出目标数据值作为训练的数据源，并计算出数据源中各历史数据模式的特征值；

[0008] S2：根据各历史数据模式的特征值来分别得到各种状态下各历史数据模式的先验概率分布，并统计各种状态的概率分布，从而训练出各种数据模式的状态的贝叶斯模型；

[0009] S3：根据监控系统获取的实时性能数据计算出当前数据模式的特征值；

[0010] S4 :从所述历史数据模式中找到与当前数据模式最相似的数据模式 ;

[0011] S5 :根据 S4 的输出结果通过 S2 中训练的贝叶斯模型来预测,分别得出所述多种状态的概率分布 ;

[0012] S6 :根据 S5 中结果来设置自信因子和异常阈值,若自信因子超出异常阈值则预测为异常状态。

[0013] 较佳地,所述特征值包括性能值变化量、性能值变化率和性能值。

[0014] 较佳地, S2 中将所有历史数据模式的各种特征值方差按取值大小排列,并划分为若干子空间,计算各子空间所对应的特征值方差的特定状态的先验概率。

[0015] 较佳地,S2 中根据所述各历史数据模式的特征值训练出各历史数据模式的贝叶斯模型,分别得到各模式的多种状态的先验概率。

[0016] 较佳地, S4 中进一步包括 :

[0017] 计算当前数据模式与各历史正常模式之间的特征值的标准方差 ;

[0018] 得出与当前数据模式所有标准方差之和最小的历史数据模式为当前数据模式的最相似模式。

[0019] 较佳地,所述状态为异常状态,警告状态以及正常状态。

[0020] 较佳地,S6 中还包括设置报警阈值,若自信因子在报警阈值与异常阈值之间,则预测为报警状态,若自信因子小于报警阈值则预测为正常状态。

[0021] 为了解决上述问题,本发明还涉及了一种分布式系统中的性能异常预测系统,与分布式系统的监控系统相连,包括 :

[0022] 历史特征值计算模块,从监控系统中的若干监控结点得到的历史性能数据中提取出目标数据值作为训练的数据源,并计算出数据源中各历史数据模式的特征值 ;

[0023] 先验概率模块,与历史特征值计算模块的输出端相连接,根据各历史数据模式的特征值来分别得到各种状态下各历史数据模式的先验概率分布,并统计各种状态的概率分布,从而训练出各种数据模式的状态的贝叶斯模型 ;

[0024] 实时特征值计算模块,根据监控系统中的若干监控结点获取的实时性能数据计算出当前数据模式的特征值 ;

[0025] 相似模式模块,与历史特征值计算模块的输出端以及实时特征计算模块的输出端相连接,从所述历史数据模式中找到与当前数据模式最相似的数据模式 ;

[0026] 概率计算模块,根据相似模式模块的输出结果通过先验概率模块中训练的贝叶斯模型来预测,分别得出所述多种状态的概率分布 ;以及

[0027] 异常报警模块,根据概率计算模块中结果来设置自信因子和异常阈值,若自信因子超出异常阈值则预测为异常状态。

[0028] 较佳地,所述特征值包括性能值变化量、性能值变化率和性能值。

[0029] 较佳地,所述状态包括异常状态,警告状态以及正常状态。

[0030] 本发明由于采用以上技术方案,与现有技术相比,具有以下的优点和积极效果 :

[0031] 1) 本发明本针对分布式系统中的性能异常预测,通过对分布式节点的性能通过特制值和划分数据模式进行分析,全面考虑变量的特征的问题,准确率更高 ;

[0032] 2) 本发明采用机器学习方法贝叶斯模型来指导预测,并实时检测出性能异常情况,并对检测出的预测通过之前得出的贝叶斯模型进行了评估分析,提供了预测的置信度,

自动化程度高，提高了预测的可靠性与实用性；

[0033] 3）本发明将各历史数据模式的特征值标准方差转化成多个子空间，将这些子空间作为贝叶斯模型的参数进行训练，计算个子空间所对应的特定状态的先验概率，进一步提升了异常预测的准确率。

## 附图说明

[0034] 图 1 为本发明一种分布式系统中的性能异常预测方法的流程图；

[0035] 图 2 为本发明一种分布式系统中的性能异常预测系统的系统框图。

## 具体实施方式

[0036] 以下将结合本发明的附图，对本发明实施例中的技术方案进行清楚、完整的描述，显然，这里所描述的仅仅是本发明的一部分实例，并不是全部的实例，基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都属于本发明的保护范围。

[0037] 为了便于对本发明实施例的理解，下面将结合附图以具体实施例为例作进一步的解释说明，且各个实施例不构成对本发明实施例的限定。

[0038] 实施例一

[0039] 请参考图 1，本发明提供了一种分布式系统中的性能异常预测方法，主要包括以下步骤：

[0040] S1：从监控系统中的若干监控结点得到的历史性能数据中提取出目标数据值作为训练的数据源，并计算出数据源中各历史数据模式的特征值；

[0041] 本实施例中，用三个方面的特征值来描述一个数据点，包括性能值变化量（Change Value，CV）、性能值变化率（Change Rate，CR）和性能值（Value，V）。性能值是一个时刻 $t_1$ 的性能度量的值。

[0042] 性能值变化量是一个时刻 $t_1$ 与另一个时刻 $t_2$ 的性能度量的差值：

[0043] $$CV(t_i)=V_{t_i}-V_{t_{i-1}}$$

[0044] 其中，$V_{t_1}$ ——时刻 $t_i$ 的性能度量的值，i ＝ 0，1，…，n；

[0045] $V_{t_{i-1}}$ ——时刻 $t_{i-1}$ 的性能度量的值，i ＝ 1，…，n。

[0046] 性能值变化率是性能度量的变化比例，等于性能值变化量除以当前时刻 $t_1$ 的性能值：

$$CR(t_i)=\frac{V_{t_i}-V_{t_{i-1}}}{V_{t_i}}*100\%$$

[0047] 其中，$V_{t_1}$ ——时刻 $t_i$ 的性能度量的值，i ＝ 0，1，…，n；

[0048] $V_{t_{i-1}}$ ——时刻 $t_{i-1}$ 的性能度量的值，i ＝ 1，…，n。

[0049] S2：根据各历史数据模式的特征值来分别得到各种状态下各历史数据模式的先验概率分布，并统计各种状态的概率分布，从而训练出各种数据模式的状态的贝叶斯模型；

[0050] 根据 S1 的数据特征结果，把历史数据分成多个模式，并对这些模式进行三个状态

的标识，即异常状态，警告状态和正常状态，然后通过三个状态训练出先验概率分布，统计出各个模式的在各个状态的概率分布，训练出各种模式的贝叶斯模型，为了进一步提升模型正确性，将模式的特征转化成多个子空间，将这些子空间作为贝叶斯模型的参数进行训练。

[0051] S2 中，可以根据各历史数据模式的特征值训练出各历史数据模式的贝叶斯模型，分别得到各模式的多种状态的先验概率。多种状态可以为异常状态，警告状态以及正常状态。

[0052] 使用朴素贝叶斯分类器来建立一个分类模型。朴素贝叶斯分类的使用限制是各参数之间是互相独立的，所得到的模式中的正式是互相独立的三个参数，因此符合朴素贝叶斯分类的要求。

[0053] 假设当前时刻为 $t_i$，那么从 $t_{i-L}$ 到 $t_i$ 的时间段内的所有数据相关的特征值构成当前数据模式，其中 L 为当前数据模式的长度。

[0054] 在训练的时候，为训练数据中的每个模式进行添加标签，来指明该模式的状态，即一个模式可以表示为 (Vt1, Vt2, …, Vtn, Status)。使用含标签的训练数据集，可以得到三个状态的所有模式的先验概率分布 (prior distribution)：

[0055] $P((SD_{CV}, SD_{CR}, SD_V) | status)$

[0056] 其中，status——正常状态 normal，报警状态 alert 或异常状态 abnormal。

[0057] 最相似模式的三个标准方差分别为 $SD_{CV}$，$SD_{CR}$，$SD_V$，该模式所对应的状态为 status 的概率大小。根据训练数据，还可以得到各个状态的分布情况 P(status)。

[0058] 根据以上先验概率，可以求得在该方差值得情况下，计算出一个特定的状态的概率大小，使用贝叶斯分类得到：

[0059] $$P(\text{status}|(SD_{CV}, SD_{CR}, SD_V)) = \frac{P((SD_{CV}, SD_{CR}, SD_V)|status)P(status)}{P((SD_{cv}, SD_{CR}, SD_V))}$$

[0060] 正如前面所述，三个参数之间是彼此独立的，因此可以表示成：

[0061] $$P(status|(SD_{CV}, SD_{CR}, SD_V)) = \frac{P(SD_{CV}|status)P(SD_{CR}|status)P(SD_V|status)P(status)}{P(SD_{CV})P(SD_{CR})P(SD_V)}$$

[0062] 为了进一步提升模型正确性，也可以为将所有历史数据模式的各种特征值方差按取值大小排列，并划分为若干子空间，计算各子空间所对应的特征值方差的特定状态的先验概率，特定状态可以为异常状态，警告状态或正常状态。

[0063] 将模式空间划分为若干个子空间，每一个子空间包含了一个连续的取值范围内存在的所有特定特征值，因此得到了若干个离散的子空间，将这些子空间作为朴素贝叶斯分类的参数。例如，性能值变化率方差 $SD_{CR}$ 的所有取值范围是 r = [a, b]，其中 a 是性能值变化率方差所取到的最小值，b 是性能值变化率方差所取到的最大值。将该空间划分为 m 个子空间，则每个子空间的长度是：

[0064] $$\Delta r = \frac{b-a}{m}$$

[0065] 所以，各子空间可表示为：

[0066] $S_{SDCR1}=[a,a+\Delta r],S_{SDCR2}=[a+\Delta r,a+2*\Delta r],\ldots,S_{SDCR1}=[b-\Delta r,b]$

[0067] 对于每个性能值变化率方差，只要将它投入到合适的子空间中即可。因此，不需要计算每个方差所对应特定状态的先验概率，只需要计算个子空间所对应的特定状态的先验概率即可：

[0068] $$P(status|(SD_{CV},SD_{CR},SD_V))=\frac{P(S_{SDCVi}|status)P(S_{SDCVj}|status)P(S_{SDCVk}|status)}{P(S_{SDCVi})P(S_{SDCRj})P(S_{SDVk})}$$

[0069] 其中，$S_{SDCVi}$——性能值变化量方差 $SD_{CV}$ 所对应的某个子空间；

[0070] $S_{SDCRj}$——性能值变化率方差 $SD_{CR}$ 所对应的某个子空间；

[0071] $S_{SDVk}$——性能值方差 $SD_V$ 所对应的某个子空间；

[0072] Status——某个特定的状态，normal，alert 或者 abnormal。

[0073] S3：根据监控系统获取的实时性能数据计算出当前数据模式的特征值。假设当前时刻为 $t_i$，那么从 $t_i$-L 到 $t_i$ 的时间段内的所有数据相关的特征构成当前数据模式，其中 L 为当前数据模式的长度。

[0074] S4：从历史数据模式中找到与当前数据模式最相似的数据模式；

[0075] 具体为：

[0076] S41：计算当前数据模式与各历史正常模式之间的特征值的标准方差；

[0077] 每个时刻 $t_i$ 的数据都有三个特征，即 $(CV(t_i),CR(t_i),V(ti))$。假设当前时刻为 $t_i$，那么从 $t_{i-L}$ 到 $t_i$ 的时间段内的所有数据相关的特征构成当前的性能度量的模式，其中 L 为当前数据模式的长度。

[0078] 如图 2，将当前的模式与历史正常模式进行比较，并在历史正常模式中找到一个与当前数据模式最相似的模式。计算当前数据模式与各历史正常模式之间的各个特征的标准方差 (Standard Deviation)。如果一个历史数据模式从时刻 $t_j$-L 开始，到 $t_j$ 结束，当前数据模式与该历史数据模式之间的性能值变化量标准方差记为 $SD_{CV}(t_j)$，当前数据模式与该历史数据模式之间的性能值变化率标准方差记为 $SD_{CR}(t_j)$，当前数据模式与该历史数据模式之间的值得标准方差 $SD_V(t_j)$。将当前数据模式与之前的历史数据模式一一对比，

[0079] S42：满足当前数据模式与一历史数据模式的所有标准方差之和最小，则设该历史数据模式为当前数据模式的最相似模式。

[0080] 当历史数据中的一个模式满足以下公式时：

[0081] $$SD_{CV}(t_k)+SD_{CR}(t_k)+SD_V(t_k)=\min_j\left\{SD_{CV}(t_j)+SD_{CR}(t_j)+SD_V(t_j)\right\}$$

[0082] 其中，$\{SD_{CV}(t_j)+SD_{CR}(t_j)+SD_V(t_j)\}$——当前数据模式与所有历史数据模式之间特征的标准方差构成的集合；

[0083] min——集合中的最小值。

[0084] 即，满足当前数据模式与这个历史数据模式的所有标准方差之和最小，那么就称该历史数据模式是当前数据模式的最相似模式。因此，对于每一个当前数据模式，都能找到一个历史中最相似的模式：

[0085] $(SD_{CV}(t_k),SD_{CR}(t_k),SD_V(t_k))$。

[0086] S5：根据S4的输出结果通过S2中训练的贝叶斯模型来预测，分别得出多种状态的概率分布；

[0087] 本实施例中根据S4的最相似模式（$SD_{CV}(t_k)$，$SD_{CR}(t_k)$，$SD_V(t_k)$），从S2中训练的贝叶斯模型来指导预测，得到模式的状态的概率情况：

[0088] $$P(\text{status}|(SD_{CV},SD_{CR},SD_V))=\frac{P((SD_{CV},SD_{CR},SD_V)|status)P(status)}{P((SD_{cv},SD_{CR},SD_V))}$$

[0089] 得到模式的概率来确定模式的状态，准确地判断模式的状态，就可以捕捉到异常发生的前兆，从而实现异常预测。

[0090] S6：根据S5中结果来设置自信因子和异常阈值，若自信因子超出异常阈值则预测为异常状态。

[0091] 还包括设置报警阈值，若自信因子在报警阈值与异常阈值之间，则预测为报警状态，若自信因子小于报警阈值则预测为正常状态。还需要设立报警机制，并通过预设的报警机制采取报警后的防御处理措施。

[0092] 本实施例中，对于当前的模式（$SD_{CV}$，$SD_{CR}$，$SD_V$），根据以上方法得到了对应三种状态的概率：

[0093] $P(\text{normal}|(SD_{cv},SD_{CR},SD_V))$

[0094] $P(\text{alert}|(SD_{cv},SD_{CR},SD_V))$

[0095] $P(\text{abnormal}|(SD_{cv},SD_{CR},SD_V))$

[0096] 为了确定该模式是处于哪种状态，将以上三个状态的概率做相应的比较：

[0097] $\delta_1 = \log P(\text{alert}|(SD_{CV},SD_{CR},SD_V))-\log P(\text{normal}|(SD_{CV},SD_{CR},SD_V))$

[0098] $\delta_2 = \log P(\text{alert}|(SD_{CV},SD_{CR},SD_V))-\log P(\text{abnormal}|(SD_{CV},SD_{CR},SD_V))$

[0099] 如果满足如下条件，则判断当前数据模式处于警报状态，在接下来可能会发生异常：

[0100] $\delta_1 \geqslant 0$ 并且 $\delta_2 \geqslant 0$

[0101] $\delta_1$表示当前数据模式处于报警状态的可能性和处于正常状态的可能性哪个更大，$\delta_2$表示当前数据模式处于报警状态的可能性和处于异常状态的可能性那个更大。如果满足式（3-10），则说明当前数据模式处于报警状态的可能性比处于正常或者异常状态的可能性都更大，可以判断接下来有可能发生异常。

[0102] 当发出预测到异常的警报时，如果 $\delta_1 \geqslant 0$，并且 $\delta_1$值越大，则表明该模式是报警状态的可能性显著大于是正常状态的可能性。同样地，如果 $\delta_2 \geqslant 0$，并且 $\delta_2$值越大，则表明该模式是报警状态的可能性显著大于是异常状态的可能性。可以说 $|\delta_1|$ 和 $|\delta_2|$ 的值越大，预测结果的可信度越高，因此可以将 $|\delta_1|$ 和 $|\delta_2|$ 作为异常预测的可信程度的参考指标。对做出的每一个异常预测分配一个自信因子（Confidence Factor，CF），自信因子的计算如下：

[0103] $CF = \delta_1 + \delta_2$

[0104] 显然地，如果该模式是alert状态的可能性越大，那么CF值越大，因此这是个有效地衡量异常预测可信度的方式。根据CF值，可以知道预测的可信度是多大，并根据可信度大小来确定报警阈值，若自信因子在报警阈值与异常阈值之间，则预测为报警状态，若自信因子小于报警阈值则预测为正常状态，还需要设立报警机制，并通过预设的报警机制在报

警状态和异常状态时采取防御处理措施来防止异常发生或者减少异常发生带来的损失。

[0105] 实施例二

[0106] 请参考图 2，本发明提供了一种分布式系统中的性能异常预测系统，与分布式系统的监控系统相连，主要包括：历史特征值计算模块、先验概率模块、实时特征值计算模块、相似模式模块、概率计算模块以及异常报警模块。

[0107] 历史特征值计算模块，从监控系统中的若干监控结点得到的历史性能数据中提取出目标数据值作为训练的数据源，并计算出数据源中各历史数据模式的特征值；

[0108] 本实施例中，用三个方面的特征值来描述一个数据点，包括性能值变化量 (Change Value，CV)、性能值变化率 (Change Rate，CR) 和性能值 (Value，V)。性能值是一个时刻 $t_1$ 的性能度量的值。

[0109] 性能值变化量是一个时刻 $t_1$ 与另一个时刻 $t_2$ 的性能度量的差值：

[0110] $$CV(t_i)=V_{t_i}-V_{t_{i-1}}$$

[0111] 其中，$V_{t_i}$ ——时刻 $t_i$ 的性能度量的值，i ＝ 0，1，…，n；

[0112] $V_{t_{i-1}}$ ——时刻 $t_{i-1}$ 的性能度量的值，i ＝ 1，…，n。

[0113] 性能值变化率是性能度量的变化比例，等于性能值变化量除以当前时刻 $t_1$ 的性能值：

$$CR(t_i)=\frac{V_{t_i}-V_{t_{i-1}}}{V_{t_i}}*100\%$$

[0114] 其中，$V_{t_i}$ ——时刻 $t_i$ 的性能度量的值，i ＝ 0，1，…，n；

[0115] $V_{t_{i-1}}$ ——时刻 $t_{i-1}$ 的性能度量的值，i ＝ 1，…，n。

[0116] 先验概率模块，与历史特征值计算模块的输出端相连接，根据各历史数据模式的特征值来分别得到各种状态下各历史数据模式的先验概率分布，并统计各种状态的概率分布，从而训练出各种数据模式的状态的贝叶斯模型；

[0117] 根据历史特征值计算模块输出的数据特征结果，把历史数据分成多个模式，并对这些模式进行三个状态的标识，即异常状态，警告状态和正常状态，然后通过三个状态训练出先验概率分布，统计出各个模式的在各个状态的概率分布，训练出各种模式的贝叶斯模型，为了进一步提升模型正确性，将模式的特征转化成多个子空间，将这些子空间作为贝叶斯模型的参数进行训练。

[0118] 先验概率模块中，可以根据各历史数据模式的特征值训练出各历史数据模式的贝叶斯模型，分别得到各模式的多种状态的先验概率。多种状态可以为异常状态，警告状态以及正常状态。

[0119] 使用朴素贝叶斯分类器来建立一个分类模型。朴素贝叶斯分类的使用限制是各参数之间是互相独立的，所得到的模式中的正式是互相独立的三个参数，因此符合朴素贝叶斯分类的要求。

[0120] 假设当前时刻为 $t_i$，那么从 $t_{i-L}$ 到 $t_i$ 的时间段内的所有数据相关的特征值构成当前数据模式，其中 L 为当前数据模式的长度。

[0121] 在训练的时候，为训练数据中的每个模式进行添加标签，来指明该模式的状态，即一个模式可以表示为 (Vt1, Vt2, …, Vtn, Status)。使用含标签的训练数据集，可以得到三个状态的所有模式的先验概率分布 (prior distribution)：

[0122] $P((SD_{CV}, SD_{CR}, SD_V)|status)$

[0123] 其中，status——正常状态 normal，报警状态 alert 或异常状态 abnormal。

[0124] 最相似模式的三个标准方差分别为 $SD_{CV}$，$SD_{CR}$，$SD_V$，该模式所对应的状态为 status 的概率大小。根据训练数据，还可以得到各个状态的分布情况 P(status)。

[0125] 根据以上先验概率，可以求得在该方差值得情况下，计算出一个特定的状态的概率大小，使用贝叶斯分类得到：

[0126] $$P(\text{status}|(SD_{CV}, SD_{CR}, SD_V))=\frac{P((SD_{CV}, SD_{CR}, SD_V)|status)P(status)}{P((SD_{cv}, SD_{CR}, SD_V))}$$

[0127] 正如前面所述，三个参数之间是彼此独立的，因此可以表示成：

[0128] $$\mathrm{P}(status|(SD_{CV}, SD_{CR}, SD_V)) = \frac{P(SD_{CV}|status)P(SD_{CR}|status)P(SD_V|status)P(status)}{P(SD_{CV})P(SD_{CR})P(SD_V)}$$

[0129] 为了进一步提升模型正确性，也可以为将所有历史数据模式的各种特征值方差按取值大小排列，并划分为若干子空间，计算各子空间所对应的特征值方差的特定状态的先验概率，特定状态可以为异常状态，警告状态或正常状态。

[0130] 将模式空间划分为若干个子空间，每一个子空间包含了一个连续的取值范围内存在的所有特定特征值，因此得到了若干个离散的子空间，将这些子空间作为朴素贝叶斯分类的参数。例如，性能值变化率方差 $SD_{CR}$ 的所有取值范围是 r = [a, b]，其中 a 是性能值变化率方差所取到的最小值，b 是性能值变化率方差所取到的最大值。将该空间划分为 m 个子空间，则每个子空间的长度是：

[0131] $$\Delta r=\frac{b-a}{m}$$

[0132] 所以，各子空间可表示为：

[0133] $S_{SDCR1} = [a, a+\Delta r], S_{SDCR2} = [a+\Delta r, a+2*\Delta r], \ldots, S_{SDCR1} = [b-\Delta r, b]$

[0134] 对于每个性能值变化率方差，只要将它投入到合适的子空间中即可。因此，不需要计算每个方差所对应特定状态的先验概率，只需要计算个子空间所对应的特定状态的先验概率即可：

[0135] $$\mathrm{P}(status|(SD_{CV}, SD_{CR}, SD_V)) = \frac{P(S_{SDCVi}|status)P(S_{SDCVj}|status)P(S_{SDCVk}|status)}{P(S_{SDCVi})P(S_{SDCRj})P(S_{SDVk})}$$

[0136] 其中，$S_{SDCVi}$——性能值变化量方差 $SD_{CV}$ 所对应的某个子空间；

[0137] $S_{SDCRj}$——性能值变化率方差 $SD_{CR}$ 所对应的某个子空间；

[0138] $S_{SDVk}$——性能值方差 $SD_V$ 所对应的某个子空间；

[0139] Status——某个特定的状态，正常状态 normal，报警状态 alert 或异常状态

abnormal。

[0140] 实时特征值计算模块，根据监控系统中的若干监控结点获取的实时性能数据计算出当前数据模式的特征值。假设当前时刻为 $t_i$，那么从 $t_i$-L 到 $t_i$ 的时间段内的所有数据相关的特征构成当前数据模式，其中 L 为当前数据模式的长度。

[0141] 相似模式模块，与实时特征值计算模块和历史特征值计算模块相连接，从历史数据模式中找到与当前数据模式最相似的数据模式；

[0142] 具体为：

[0143] 历史数据比较模块，与实时特征值计算模块和历史特征值计算模块相连接，计算当前数据模式与各历史正常模式之间的特征值的标准方差；

[0144] 每个时刻 $t_i$ 的数据都有三个特征，即 $(CV(t_i),CR(t_i),V(ti))$。假设当前时刻为 $t_i$，那么从 $t_{i-L}$ 到 $t_i$ 的时间段内的所有数据相关的特征构成当前的性能度量的模式，其中 L 为当前数据模式的长度。

[0145] 如图 2，将当前的模式与历史正常模式进行比较，并在历史正常模式中找到一个与当前数据模式最相似的模式。计算当前数据模式与各历史正常模式之间的各个特征的标准方差 (Standard Deviation)。如果一个历史数据模式从时刻 $t_j$-L 开始，到 $t_j$ 结束，当前数据模式与该历史数据模式之间的性能值变化量标准方差记为 $SD_{CV}(t_j)$，当前数据模式与该历史数据模式之间的性能值变化率标准方差记为 $SD_{CR}(t_j)$，当前数据模式与该历史数据模式之间的值得标准方差 $SD_V(t_j)$。将当前数据模式与之前的历史数据模式一一对比，

[0146] 最小方差获取模块，与历史数据比较模块的输出端相连接，满足当前数据模式与一历史数据模式的所有标准方差之和最小，则设该历史数据模式为当前数据模式的最相似模式。

[0147] 当历史数据中的一个模式满足以下公式时：

[0148] $$SD_{CV}(t_k)+SD_{CR}(t_k)+SD_V(t_k)=\min_j\left\{SD_{CV}(t_j)+SD_{CR}(t_j)+SD_V(t_j)\right\}$$

[0149] 其中，$\{SD_{CV}(t_j)+SD_{CR}(t_j)+SD_V(t_j)\}$——当前数据模式与所有历史数据模式之间特征的标准方差构成的集合；min——集合中的最小值。

[0150] 即，满足当前数据模式与这个历史数据模式的所有标准方差之和最小，那么就称该历史数据模式是当前数据模式的最相似模式。因此，对于每一个当前数据模式，都能找到一个历史中最相似的模式：

[0151] $(SD_{CV}(t_k),SD_{CR}(t_k),SD_V(t_k))$。

[0152] 概率计算模块，根据相似模式模块的输出结果通过先验概率模块中训练的贝叶斯模型来预测，分别得出多种状态的概率分布；

[0153] 本实施例中根据最小方差获取模块的最相似模式：

[0154] $(SD_{CV}(t_k),SD_{CR}(t_k),SD_V(t_k))$，从先验概率模块中训练的贝叶斯模型来指导预测，得到模式的各状态的概率情况：

[0155] $$P(\text{status}\,|\,(SD_{CV},SD_{CR},SD_V))=\frac{P((SD_{CV},SD_{CR},SD_V)\,|\,status)P(status)}{P((SD_{cv},SD_{CR},SD_V))}$$

[0156] 得到模式的概率来确定模式的状态，准确地判断模式的状态，就可以捕捉到异常发生的前兆，从而实现异常预测。

[0157] 异常报警模块，根据概率计算模块中输出结果来设置自信因子和异常阈值，若自信因子超出异常阈值则预测为异常状态。

[0158] 一般还包括设置报警阈值，若自信因子在报警阈值与异常阈值之间，则预测为报警状态，若自信因子小于报警阈值则预测为正常状态。还需要设立报警机制，并通过预设的报警机制采取报警后的防御处理措施。

[0159] 本实施例中，对于当前的模式 $(SD_{CV}, SD_{CR}, SD_V)$，根据以上方法得到了对应三种状态的概率：

[0160] $P(normal|(SD_{cv}, SD_{CR}, SD_V))$

[0161] $P(alert|(SD_{cv}, SD_{CR}, SD_V))$

[0162] $P(abnormal|(SD_{cv}, SD_{CR}, SD_V))$

[0163] 为了确定该模式是处于哪种状态，将以上三个状态的概率做相应的比较：

[0164] $\delta_1 = \log P(alert|(SD_{CV}, SD_{CR}, SD_V)) - \log P(normal|(SD_{CV}, SD_{CR}, SD_V))$

[0165] $\delta_2 = \log P(alert|(SD_{CV}, SD_{CR}, SD_V)) - \log P(abnormal|(SD_{CV}, SD_{CR}, SD_V))$

[0166] 如果满足如下条件，则判断当前数据模式处于警报状态，在接下来可能会发生异常：

[0167] $\delta_1 \geq 0$ 并且 $\delta_2 \geq 0$

[0168] $\delta_1$ 表示当前数据模式处于报警状态的可能性和处于正常状态的可能性哪个更大，$\delta_2$ 表示当前数据模式处于报警状态的可能性和处于异常状态的可能性那个更大。如果满足式 (3-10)，则说明当前数据模式处于报警状态的可能性比处于正常或者异常状态的可能性都更大，可以判断接下来有可能发生异常。

[0169] 当发出预测到异常的警报时，如果 $\delta_1 \geq 0$，并且 $\delta_1$ 值越大，则表明该模式是 alert 状态的可能性显著大于是正常状态的可能性。同样地，如果 $\delta_2 \geq 0$，并且 $\delta_2$ 值越大，则表明该模式是报警状态的可能性显著大于是异常状态的可能性。可以说 $|\delta_1|$ 和 $|\delta_2|$ 的值越大，预测结果的可信度越高，因此可以将 $|\delta_1|$ 和 $|\delta_2|$ 作为异常预测的可信程度的参考指标。对做出的每一个异常预测分配一个自信因子 (Confidence Factor，CF)，自信因子的计算如下：

[0170] $CF = \delta_1 + \delta_2$

[0171] 显然地，如果该模式是 alert 状态的可能性越大，那么 CF 值越大，因此这是个有效地衡量异常预测可信度的方式。根据 CF 值，可以知道预测的可信度是多大，并根据可信度大小来确定报警阈值，若自信因子在报警阈值与异常阈值之间，则预测为报警状态，若自信因子小于报警阈值则预测为正常状态，还需要设立报警机制，并通过预设的报警机制在报警状态和异常状态时采取防御处理措施来防止异常发生或者减少异常发生带来的损失。

[0172] 以上所述，仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应该以权利要求的保护范围为准。


从监控系统中的若干监控结点得到的历史性能数据中提取出目标数据值作为训练的数据源，并计算出数据源中各历史数据模式的特征值
S1
根据各历史数据模式的特征值来分别得到各种状态下各历史数据模式的先验概率分布，并统计各种状态的概率分布，从而训练出各种数据模式的状态的贝叶斯模型
S2
根据监控系统获取的实时性能数据计算出当前数据模式的特征值
S3
从历史数据模式中找到与当前数据模式最相似的数据模式
S4
根据S4的输出结果通过S2中训练的贝叶斯模型来预测，分别得出多种状态的概率分布
S5
根据S5中结果来设置自信因子和异常阈值，若自信因子超出异常阈值则预测为异常状态
S6

图 1

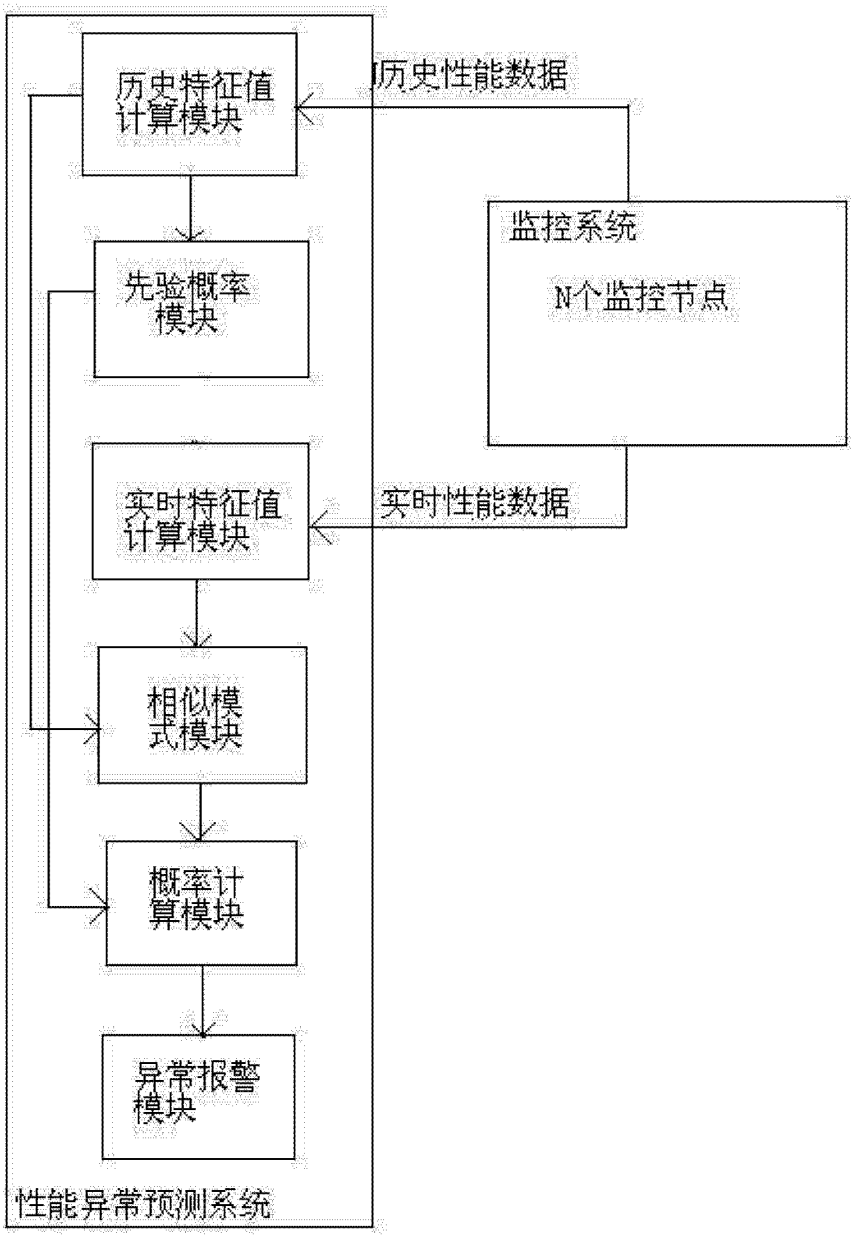

历史特征值
计算模块
历史性能数据
监控系统
N个监控节点
先验概率
模块
实时特征值
计算模块
实时性能数据
相似模
式模块
概率计
算模块
异常报警
模块
性能异常预测系统

图 2