

# 应用系统负载分析与磁盘容量预测

通过对系统的当前负载率与历史平均负载率进行比较，获得负载率的当前趋势。通过负载率以及负载趋势可对系统进行负载分析。分析存储设备中磁盘容量预测，通过对磁盘容量进行预测，可预测磁盘未来的负载情况，避免应用系统因出现存储容量耗尽的情况而导致应用系统负载率过高，最终引发系统故障。

- 针对历史磁盘数据，采用时间序列分析方法，预测应用系统服务器磁盘已使用空间大小。
- 根据用户需求设置不同的预警等级，将预测值与容量值进行比较，对其结果进行预警判断，为系统管理员提供定制化的预警提示。

## 分析方法与过程

- 1) 从数据源中选择性抽取历史数据与每天定时抽取数据。
- 2) 对抽取的数据进行周期性分析以及数据清洗、数据变换等操作后，形成建模数据。
- 3) 采用时间序列分析法对建模数据进行模型的构建，利用模型预测服务器磁盘已使用情况。
- 4) 应用模型预测服务器磁盘将要使用情况，通过预测到的磁盘使用大小与磁盘容量大小按照定制化标准进行判断，将结果反馈给系统管理员，提示管理员需要注意磁盘的使用情况。

## 数据抽取

## 数据探索分析

对数据平稳性进行检验。

## 数据预处理

数据清洗  
属性构造

## 模型构建

### 容量预测模型

首先需要对观测值序列进行平稳性检验，如果不平稳，则对其进行差分处理直到差分后的数据平稳。在数据平稳后，对其进行白噪声检验。如果没有通过白噪声检验，就进行模型识别，识别其模型属于 AR、MA 和 ARMA 中的哪一种模型。并且通过 BIC 信息准则对模型进行定阶，确定 ARIMA 模型的  $p$ 、 $q$  参数。在模型识别后需进行模型检验，检测模型残差序列是否为白噪声序列。如果模型没有通过检测，需要对其进行重新识别。对已通过检验的模型采用极大似然估计方法进行模型参数估计。最后，应用模型进行预测，将实际值与预测值进行误差分析。如果误差比较小（误差阈值需通过业务分析进行设定），表明模型拟合效果较

好，则模型可以结束。反之需要重新估计参数。

- **平稳性检验**：为了确定原始数据序列中没有随机趋势或确定趋势，需要对数据进行平稳性检验，否则将会产生“伪回归”的现象。采用单位根检验（ADF）的方法或时序图的方法进行平稳性检验。
- **白噪声检验**：为了检验序列中有用的信息是否已被全部提取完毕，需要对序列进行白噪声检验。如果序列检验为白噪声，说明序列中有用信息已经被提取完毕，剩下的全是随机扰动，无法进行预测和使用。
- **模型识别**：采用极大似然比方法进行模型的参数估计，估计各个参数的值。然后针对各个不同的模型，采用 BIC 信息准则对模型进行定阶，确定  $p, q$  参数，从而选择最优模型。
- **模型检验**：模型确定后，检验其残差序列是否为白噪声。如果不是白噪声，说明残存中还存在有用的信息，需要修改模型或者进一步提取。
- **模型预测**：利用测试机与实际值进行比较

### 模型评价

评价时序预测模型效果的好坏，采用平均绝对误差、均方根误差和平均均对百分误差，从不同方面反应算法的预测精度。

### 模型应用

1) 从系统中每日定时抽取服务器磁盘数据。

2) 对定时抽取的数据进行数据清洗、数据变换预处理操作。

3) 将预处理后的定时数据存放到模型的初始数据中，获得模型的输入数据，调用模型对服务器磁盘已使用空间进行预测，预测后 5 天的磁盘已使用空间大小。

4) 将预测值与磁盘的总容量进行比较，获得预测的磁盘使用率。如果某一天预测的使用率达到业务设置的预警级别，就会以预警的方式提醒系统管理员。

模型应用的预警流程图如图 11-7 所示。

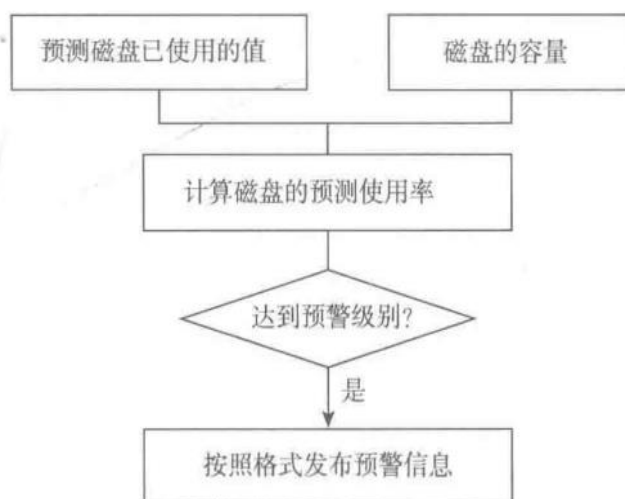


图 11-7 预警流程图

因为模型采用历史数据进行建模，随着时间的变化，每天会定时地将新增数据加入初始建模数据中。在正常的情况下，模型需要重新调整。但考虑到建模的复杂性高，且磁盘的已使用大小每天的变化量相对很小，对于整个模型的预测影响较小。因此，结合实际业务情况，每半个月对模型进行一次调整。