应用系统负载分析与磁盘容量预测

1 项目背景

某大型企业为了信息化发展的需要,建设了办公自动化系统、人力资源管理系统、财务 管理系统、企业信息门户系统等几大企业级应用系统。因应用系统在日常运行时,会对底层 软硬件造成负荷。显著影响应用系统性能的因素包括:服务器、数据库、中间件、存储设备。 任何一种资源负载过大,都可能会引起应用系统性能下降甚至瘫痪。因此需要关注服务器、 数据库、中间件、存储设备的运行状态,及时了解当前应用系统的负载情况,以便提前预防, 确保系统安全稳定运行。

应用系统的负载率可以通过对一段时间内软硬件性能的运行状况进行综合评分而获得。 通过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,当出现应用系统的负载高或者负载趋势大的现象,代表 系统目前处于高危工作环境中。如果系统管理员不及时进行相应的处理,系统很容易出现故 障,从而导致用户无法访问系统,严重影响企业的利益。本章重点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。避免应用系统出现存储容量 耗尽的情况,从而导致应用系统负载率过高,最终引发系统故障。

2 项目目标

(1) 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。

(2) 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。

3 项目步骤

3.1 工程前期准备

3.1.1 导入数据

(1) 介绍数据

目前监控采集的性能数据主要包含 CPU 使用信息,内存使用信息,磁盘使用信息等,如表 3-1 所示。通过分析磁盘容量相关数据(见表 3-2),预测应用系统服务器磁盘空间是 否满足系统健康运行的要求。

属性名称	属性说明	属性名称	属性说明
SYS_NAME	资产所在的系统名称	ENTITY	具体的属性
NAME	资产名称	VALUE	采集到的值
	属性的标识号		
TARGET_ID	183 表示磁盘容量大小	COLLECTTIME	采集的时间
	184 表示磁盘已使用大小		
DESCRIPTION	针对属性标识的说明		

表 3-1 性能说明表

SYS_NAME	NAME	TARGET_ID	DESCRIPTION	ENTITY	VALUE	COLLECTTIME
财务管理系统	CWXT DB	184	磁盘已使用大小	C:\	34270787.33	2014/10/1
财务管理系统	CWXT DB	184	磁盘已使用大小	D:\	80262592.65	2014/10/1
财务管理系统	CWXT_DB	183	磁盘容量	C:\	52323324	2014/10/1
财务管理系统	CWXT DB	183	磁盘容量	D:\	157283328	2014/10/1
财务管理系统	CWXT DB	184	磁盘已使用大小	C:\	34328899.02	2014/10/2
财务管理系统	CWXT_DB	184	磁盘已使用大小	D:\	83200151.65	2014/10/2
财务管理系统	CWXT DB	183	磁盘容量	C:\	52323324	2014/10/2
财务管理系统	CWXT DB	183	磁盘容量	D:\	157283328	2014/10/2
财务管理系统	CWXT_DB	184	磁盘已使用大小	C:\	34327553.5	2014/10/3
财务管理系统	CWXT DB	184	磁盘已使用大小	D:\	83208320	2014/10/3
财务管理系统	CWXT DB	183	磁盘容量	C:\	52323324	2014/10/3
财务管理系统	CWXT DB	183	磁盘容量	D:\	157283328	2014/10/3
财务管理系统	CWXT_DB	184	磁盘已使用大小	C:\	34288672.21	2014/10/4
财务管理系统	CWXT DB	184	磁盘已使用大小	D:\	83099271.65	2014/10/4
财务管理系统	CWXT DB	183	磁盘容量	C:\	52323324	2014/10/4
财务管理系统	CWXT_DB	183	磁盘容量	D:\	157283328	2014/10/4
财务管理系统	CWXT DB	184	磁盘已使用大小	C:\	34190978.41	2014/10/5
财务管理系统	CWXT DB	184	磁盘已使用大小	D:\	82765171.65	2014/10/5
财务管理系统	CWXT_DB	183	磁盘容量	C:\	52323324	2014/10/5
财务管理系统	CWXT DB	183	磁盘容量	D:\	157283328	2014/10/5
财务管理系统	CWXT DB	184	磁盘已使用大小	C:\	34187614.43	2014/10/6
财务管理系统	CWXT DB	184	磁盘已使用大小	D:\	82522895	2014/10/6
财务管理系统	CWXT_DB	183	磁盘容量	C:\	52323324	2014/10/6
财务管理系统	CWXT DB	183	磁盘容量	D:\	157283328	2014/10/6

表 3-2 磁盘原始数据集

(2) 上传数据到 Python 数据挖掘建模平台

在新增数据源上,选择本地上传数据,如图1所示。

	Python数据挖掘建模平台						0	
▲ 首页	我的数据源共享数据源							
数据源	+ 新墳数据源 ▲		请输入表名	选择状态 🔻 请说	起译创建时间		搜索	
「工程	■ 数据来源于文件 创建人	数据来源	同步状态	创建时间	操作			
	■ 数据来源于数据库							
₩								
日子			暫无数据					

图 1 本地上传数据源

在本地路径上选择文件,填写在平台新建的目标表名,如图 2 所示。

	新建数距			×
1 文件属性 上传文件	2 预贷数 进某文件	掘		3) 字形论置
* 新建目标表名	discdata			
列分隔符	逗号 (,) 👻	文件编码	UTF-8 👻	
存储有效期 (天)	18d — +	预览设置	分页显示	
				重置下一步

图 2 本地选择文件上传

根据文件的数据,可以修改文件的字段名和类型,如图 3 所示。

	新建数据源											
1 文件属性					2 预览数							3 字段设置
注意: 【字段名】只能 原字段	<mark>提以字母开头,由小写英文</mark> : 字段名	字母、数字、下划线组成 类型		¥	度		3	精度		备注		
sys_name	sys_name	字符	-	255	-	+	0		+			
name	name	字符	-	255	-	+	0		+			
target_id	target_id	数值	-	255	-	+	0		+			
description	description	字符	-	255	-	+	0		+			
entity	entity	字符	-	255	-	+	0	-	+			
value	value	数值	-	255	-	+	2	-	+			
											E	i i i i i i i i i i i i i i i i i i i
											10	- PROALE

图 3 字段设置

上传成功,可以在平台的数据源上查看数据,单击数据源操作的查看按钮如图 4 所示, 数据预览如图 5 所示。

+ 新增数据源 ▼				请输入表名	选择	¥状态 ▼	请选择创建时间	
表名	创建人	数据来源	同步状态	创建时间		操作		
discdata	xinyou	结构化文件	同步完成	2019-05-28 08:46:49		 		
hotspotdata	xinyou	结构化文件	同步完成	2019-05-27 15:33:36		• 🕯 <		
user_dat	xinyou	结构化文件	同步完成	2019-05-27 13:59:09				

图 4 单击预览数据按钮

sys_name name target_id description entity value collectime 财务管理系统 CWXT_DB 184 磁盘已使用大小 C 34270787.33 2014-10-01 财务管理系统 CWXT_DB 184 磁盘已使用大小 D 80262592.65 2014-10-01 财务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-01 财务管理系统 CWXT_DB 183 磁盘容量 D 157283328 2014-10-01 财务管理系统 CWXT_DB 183 磁盘容量 D 157283328 2014-10-01 财务管理系统 CWXT_DB 184 磁盘容量 D 3432889.02 2014-10-01 财务管理系统 CWXT_DB 184 磁盘它使用大小 C 3432889.02 2014-10-02 财务管理系统 CWXT_DB 184 磁盘已使用大小 D 83200151.65 2014-10-02 财务管理系统 CWXT_DB 183 磁盘容量 C 5232324 2014-10-02 财务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-				预	急数据 (分页加载)		
財务管理系统 CWXT_DB 184 磁盘已使用大小 C 34270787.33 2014-10-01 財务管理系统 CWXT_DB 184 磁盘已使用大小 D 800262592.65 2014-10-01 財务管理系统 CWXT_DB 183 磁盘容量 C 5232324 2014-10-01 財务管理系统 CWXT_DB 183 磁盘容量 D 5523324 2014-10-01 財务管理系统 CWXT_DB 183 磁盘容量 D 157283328 2014-10-01 財务管理系统 CWXT_DB 184 磁盘容量 D 157283328 2014-10-01 財务管理系统 CWXT_DB 184 磁盘已使用大小 C 3432899.02 2014-10-02 財务管理系统 CWXT_DB 184 磁盘已使用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 183 磁盘已使用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-02 <td>sys_name</td> <td>name</td> <td>target_id</td> <td>description</td> <td>entity</td> <td>value</td> <td>collecttime</td>	sys_name	name	target_id	description	entity	value	collecttime
財务管理系统 CWXT_DB 184 磁盘它使用大小 D 80262592.65 2014-10-01 財务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-01 財务管理系统 CWXT_DB 183 磁盘容量 D 157283328 2014-10-01 財务管理系统 CWXT_DB 184 磁盘合使用大小 C 3432889.02 2014-10-01 財务管理系统 CWXT_DB 184 磁盘合使用大小 C 3432889.02 2014-10-02 財务管理系统 CWXT_DB 184 磁盘已使用大小 C 83200151.65 2014-10-02 財务管理系统 CWXT_DB 183 磁盘合量 C 5232324 2014-10-02 財务管理系统 CWXT_DB 183 磁盘合量 C 5232324 2014-10-02 財务管理系统 CWXT_DB 183 磁盘合量 C 5232324 2014-10-02	财务管理系统	CWXT_DB	184	磁盘已使用大小	С	34270787.33	2014-10-01
財务管理系统 CWXT_DB 183 磁盘容量 C 5232324 2014-10-01 財务管理系统 CWXT_DB 183 磁盘容量 D 157283328 2014-10-01 財务管理系统 CWXT_DB 184 磁盘已使用大小 C 3432889.02 2014-10-02 財务管理系统 CWXT_DB 184 磁盘已使用大小 C 83200151.65 2014-10-02 財务管理系统 CWXT_DB 184 磁盘已使用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 5232324 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 5232324 2014-10-02	财务管理系统	CWXT_DB	184	磁盘已使用大小	D	80262592.65	2014-10-01
財务管理系统 CWXT_DB 183 磁盘容量 D 157283328 2014-10-01 財务管理系统 CWXT_DB 184 磁盘已使用大小 CC 3432899.02 2014-10-02 財务管理系统 CWXT_DB 184 磁盘已使用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 184 磁盘白度用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 CC 5232324 2014-10-02 財务管理系统 CWAT_DB 183 磁盘容量 C 5232324 2014-10-02	财务管理系统	CWXT_DB	183	磁盘容量	С	52323324	2014-10-01
財务管理系统 CWXT_DB 184 磁盘已使用大小 C 34328899.02 2014-10-02 財务管理系统 CWXT_DB 184 磁盘已使用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 184 磁盘已使用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-02 財务管理系统 CWAT_DB 193 磁盘容量 D 1457089328 2014-10-02	财务管理系统	CWXT_DB	183	磁盘容量	D	157283328	2014-10-01
財务管理系统 CWXT_DB 184 磁盘已使用大小 D 83200151.65 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-02 財务管理系统 CWXT_DB 183 磁盘容量 C 52323324 2014-10-02	财务管理系统	CWXT_DB	184	磁盘已使用大小	С	34328899.02	2014-10-02
財务管理系统 CWXT_DB 183 磁盘容量 C 52333324 2014-10-02	财务管理系统	CWXT_DB	184	磁盘已使用大小	D	83200151.65	2014-10-02
HSS TATE 192 正中宗皇 D 157292329 2014 10 02	财务管理系统	CWXT_DB	183	磁盘容量	С	52323324	2014-10-02
则为自注意的L 0WX1_DD 103 00m台里 D 107203220 2014-10-02	财务管理系统	CWXT_DB	183	磁盘容量	D	157283328	2014-10-02

图 5 数据预览

3.1.2 新建空白工程

右击我的工程,新建一个空白的工程,如图 6 所示。

	Python数据挖掘建模	要平台	4 🛔 🗭 🛛
▼ 倉页	工程 〇	🛓 Q Q 120% % 🖺 O	工程信息
))) 数据源	 ▶ ○□AA-〒<□ ● 新建工程 ▲ 导入工程 		未选择工程
工程	 ▲ 导出工程 ■ 添加文件夹 		
⑦ ↑人组件	⇒		
◆ 模型	输入内容进行过滤 ▶ 系统组件		
TTT 任务	▶ 个人组件		
	模型		

图 6 新建工程

填写工程的信息,包括工程名称和工程描述,如图 7 所示。

	创建工程	×
* 工程名称	应用系统负载分析与磁盘容量预测	
工程描述	通过分析磁盘容量相关数据,预测应用系统服务器磁盘空间是否满足系统健康运行的要求。案例针对历史磁盘数据,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。并根据用户	*
工程位置	▼我的工程	
	重置确定	

图 7 填写工程信息

3.2 数据预处理

读取 discdata 数据,步骤如图 8 所示。

(1)选择工程。

- (2) 选择输入源组件。
- (3) 拖入输入源组件。
- (4) 填写数据表名。
- (5) 单击更新按钮,更新出数据。

工程 〇	📩 Q Q 🛛 120%) % 🖺 D	◇ 字段属性			
▼ 我的工程		数据表			0
▲ 竞赛网站用户	≓ 输入源	discdata			
▲ 气象与输电线					
▲ 应用系统负载		字段信息			
<u>_</u>		2			
组件		字段	类型	取值范围	
输入内容进行过滤		sys_name	字符	财务管	
▼ 系统组件		name	字符	CWXT	
➡ 输入源	•	target_id	数值	183-184	
≓ 輸出源		descriptio	字符	磁盘容	
▶ 预处理		n	515		
▶ 统计分析					
▶分类					
▶ 聚类					
▶ 时序模型					
▶ 关联规则					
▲ 档开店亚社		N 1010100-0			
模型		/ 出行+细述			

图 8 输入源组件

3.2.1 数据筛选

选择数据筛选,步骤如图 9、图 10 所示。

(1) 找到预处理→数据筛选。

(2) 拖入数据筛选组件,将输入源和数据筛选组件连接。

(3) 单击更新按钮,勾选全部字段作为输出字段。

(4) 选择参数设置,点击添加项,然后点击刷新,字段选择 target_id,函数选择=,值填入 184。

(5) 对数据筛选组件右键,选择运行该节点。

工程 〇	🛓 I Q I Q 🛛 120% 🛛 🗞 I 🖹 I 💿	~ 穷	0属性		
▼ 我的工程		特征			0
▲ 竞赛网站用户		2			
▲ 气象与输电线 ▲ 应用系统负载		添加亨	假过滤字符串		
	数 数据演选		字段	类型	取值范围
组件			sys_name	字符	财务管理系统
100000			name	字符	CWXT_DB
网 土曜日井 図 修改类型			target_id	数值	183-184
■ 数据集划分			description	字符	磁盘容量,磁
		-	antite -	-shy-firty	0.0
▲ 赤山山/丁シリ ■ 表堆叠					
X SQL脚本		L			
■ 排序 ■ 数据筛选					
2011年日 2011年日 2011年日					
A TSHASULA X Python脚本		> 参数	女设置		
模型		> 组	件描述		

图 9 数据筛选组件_字段属性

工程 O	📩 Q Q <mark>120%</mark> % 🖹 D	> 字段屬性
▼我的工程	MAIN NET	◆ 参数设置
▲ 克赉网站用户		条件
▲ 气象与输电线		
▲ 应用系统负载		+ 0
	数据筛选	and 🐨 target 🐨 = 🐨 184 🗃
组件		
翼 主键合并		
又修改类型		
X 数据集划分		
🗙 记录去重		
🗙 新增序列		
🗙 表堆叠		
X SQL脚本		
■ 排序		
■ 数据筛选		
■ 分组聚合		
■ 修改列名		
X Python脚本		
模型		> 组件描述

图 10 数据筛选组件_参数设置

(6)运行完成后,对数据筛选组件右键,选择查看数据,数据筛选的输出表结果如图 11

所示。

			预览数据			>
sys_name	name	target_id	description	entity	value	collecttime
财务管理系统	CWXT_DB	184	磁盘已使用大小	с	34270787.33	2014-10-01
财务管理系统	CWXT_DB	184	磁盘已使用大小	D	80262592.65	2014-10-01
财务管理系统	CWXT_DB	184	磁盘已使用大小	С	34328899.02	2014-10-02
财务管理系统	CWXT_DB	184	磁盘已使用大小	D	83200151.65	2014-10-02
财务管理系统	CWXT_DB	184	磁盘已使用大小	С	34327553.5	2014-10-03
财务管理系统	CWXT_DB	184	磁盘已使用大小	D	83208320	2014-10-03
财务管理系统	CWXT_DB	184	磁盘已使用大小	С	34288672.21	2014-10-04
财务管理系统	CWXT_DB	184	磁盘已使用大小	D	83099271.65	2014-10-04
		共 94 条 2	25条/页 🔻 🤇 1 2 3 4	〉 前往 1 页		

图 11 数据筛选结果

3.2.2 属性变换

接下来进行属性变换,步骤如图 12 所示。

- (1) 找到预处理→Python 脚本组件。
- (2) 拖入 Python 脚本组件,并将数据筛选和 Python 脚本组件连接。
- (3)选择字段属性,在脚本处填入数据变换代码,如表 3-3 所示。
- (4) 对 Python 脚本组件右键,选择运行该节点。



图 12 属性变换组件

```
表 3-3 属性变换代码
```

data_in = db_utils.query(conn, 'select * from ' + inputs['input1'])
import warnings
warnings.filterwarnings("ignore")
import pandas as pd

data_group = data_in.groupby('collecttime') #以时间分组

def attr_trans(x): #定义属性变换函数 result = pd.Series(index = ['sys_name', 'cwxt_db_184_c_', 'cwxt_db_184_d_', 'collecttime']) result['sys_name'] = x['sys_name'].iloc[0] result['collecttime'] = x['collecttime'].iloc[0] result['cwxt_db_184_c_'] = x['value'].iloc[1] result['cwxt_db_184_d_'] = x['value'].iloc[0] return result

data_processed = data_group.apply(attr_trans) #逐组处理

data_out = pd.DataFrame(data_processed)

return(data_out)

(5)运行完成后,对 Python 脚本组件右键,选择查看数据,如图 13 所示。

预览数据				
sys_name	cwxt_db_184_c_	cwxt_db_184_d_	collecttime	
财务管理系统	80262592.65	34270787.33	2014-10-01	
财务管理系统	83200151.65	34328899.02	2014-10-02	
财务管理系统	83208320	34327553.5	2014-10-03	
财务管理系统	83099271.65	34288672.21	2014-10-04	
财务管理系统	82765171.65	34190978.41	2014-10-05	
财务管理系统	82522895	34187614.43	2014-10-06	
财务管理系统	82590885	34285280.22	2014-10-07	
财务管理系统	82368173.3	34290578.41	2014-10-08	
	共 47 条 25 条/页 ▼	〈 1 2 〉 前往 1 页		

图 13 属性变换结果

(6) 运行完成后,对 Python 脚本组件右键,重命名为属性变换。

3.2.3 生成训练数据

选择数据筛选,步骤如图 14、图 15 所示。

(1) 找到预处理→数据筛选。

(2) 拖入数据筛选组件,将属性变换和数据筛选组件连接。

(3) 单击更新按钮,勾选全部字段作为输出字段。

(4) 选择参数设置,点击添加项,然后点击刷新,字段选择 collecttime,函数选择<,值 填入 2014-11-12。

(5) 对数据筛选组件右键,选择运行该节点。

	🛓 Q Q (120%) % 🖪 O	~判	2属性		
▼我的工程		特征			0
▲ 竞赛网站用户	11 11 11 11 11 11 11 11 11 11 11 11 11				
▲ 气象与输电线 ▲ 应用系统合教		添加亨	假过滤字符串		
· 121110000232	区 教廷告诉法		字段	类型	取值范围
184t			sys_name	字符	
▲记录去重	区 属性变换		cwxt_db_184 _c_	数值	-
■ 新增序列 ■ 表堆叠	全 生成训练数据		cwxt_db_184 _d_	数值	
X SQL脚本			collecttime	日期	
置排序 函数脂肪透 如分用聚合 如何改列名 如Python脚本。 就是法信负进理 如数学类函数					
■ 特征构造		> 参发	文设置		
個型		> 组	牛描述		

图 14 生成训练数据组件_字段属性

工程 〇	🛓 Q Q 🚺 🗞 🖺 🖸	> 字段屬性
▼我的工程		✓ 参数设置
▲ 竞赛网站用户	1111 1111 1111 1111 1111 1111 1111 1111 1111	条件
▲ 气象与输电线		JAIT 0
▲ 应用系统负载	Refresterate	+ 0
	SUB-may	and 〒 collec 〒 < 〒 2014-11-1 窗
49.01		
Willster	区 属性变换	
M ICRZ出 M 新聞の		
X 30140/1775		
X SOL 期末	这 生成3 临场费双语	
24年7月1日		
■ 数据筛选		
其 分组聚合		
鼠 修改列名		
X Python脚本		
■ 缺失值处理		
■ 数学类函数		
■ 特征构造		
■ 数据意散化		
模型		> 组件描述

图 15 生成训练数据组件_参数设置

(6)运行完成后,对数据筛选组件右键,选择查看数据,数据筛选的输出表结果如图 16 所示。

	预览数	牧据	
sys_name	cwxt_db_184_c_	cwxt_db_184_d_	collecttime
财务管理系统	80262592.65	34270787.33	2014-10-01
财务管理系统	83200151.65	34328899.02	2014-10-02
财务管理系统	83208320	34327553.5	2014-10-03
财务管理系统	83099271.65	34288672.21	2014-10-04
财务管理系统	82765171.65	34190978.41	2014-10-05
财务管理系统	82522895	34187614.43	2014-10-06
财务管理系统	82590885	34285280.22	2014-10-07
财务管理系统	82368173.3	34290578.41	2014-10-08
	共 42 条 25 条/页 🔍	〈 1 2 〉 前往 1 页	

图 16 生成训练数据结果

(7)运行完成后,对数据筛选组件右键,重命名为生成训练数据。

3.2.4 平稳性检验

选择平稳性检验,步骤如图 17 所示。

- (1) 找到统计分析→平稳性检验。
- (2) 拖入平稳性检验组件,将生成训练数据和平稳性检验组件连接。

(3) 单击更新按钮,时序特征勾选 cwxt_db_184_d_字段作为检验字段。

(4) 对平稳性检验组件右键,选择运行该节点。

工程 O	📩 Q Q 120% % 🖺 O	◇ 字段属性		
▼ 我的工程		数据		
▲ 寬賽网站用户	(二) 输入源			
▲ 飞驶与轴电线		字段 类型 取值范围		
	数据 渐选	sys_name 字符		
组件	赵 属性变换	cwxt_db		
G Mann-Whitney G K-S检验		cwxt_db_ 184 d 数值 -		
C 全表统计		collecttim e 日期		
© 方差齐性检验		时序特征	6	
C Granger因果	C TRILLING			
 主成分分析 頻数统计 		cwxt_db_184_d_	v	
© 因子分析 ● 平稳性检验				
▶ 分类				
▶ 聚类				
模型		>> 担件描述		

图 17 平稳性检验组件

(5)运行完成后,对平稳性检验组件右键,选择查看报告,平稳性检验的报告结果如图 18 所示。

	算法运行服告	×
		. 1
	平稳性检验结果	
(检验结果	
	Test statistic: -0.5635118147156265 p-value: 0.8790155702809297 Number of lags used: 3 Number of observations used for the ADF regression and calculation of the critical values: 38 Critical values for the test statistic at the 5 %: -2.941262357486514 Critical values for the test statistic at the 1 %: -3.6155091011809297 Critical values for the test statistic at the 10 %: -2.6091995013850418	
(ACF	
		下载

图 18 平稳性检验报告

3.2.5 纯随机性检验

选择平稳性检验,步骤如图 19 所示。

(1) 找到统计分析→纯随机性检验。

(2) 拖入平稳性检验组件,将生成训练数据和纯随机性检验组件连接。

(3) 单击更新按钮,特征勾选 cwxt_db_184_d_字段作为检验字段。

(4) 对纯随机性检验组件右键,选择运行该节点。

工程 〇	📩 Q Q <mark>120%</mark> % 🖺 D	→ 字段属性	
 ▼我的工程 ▲ 竞赛网站用户 ▲ 气象与输电线 	(C)	数据 3	0
▲应用系统负载		字段 类型 取值范围	
	数 据筛选	sys_name 字符	
组件	図 属性变换	cwxt_db	
© 二项分布检验		cwxt_db_ 数值 -	
✿ 双样本T检验		collecttim	-
C 正态性检验	至 生DG/I的後的店	e 日期	
Mann-Whitney.		特征	0
G K-S 恒短 の	平稳性检验		
▲ 独随机性检验		cwat db 184 d	
© 方差齐性检验			
C Granger因果			
C 主成分分析			
C 频数统计			
© 因子分析			
模型		> 组件描述	

图 19 纯随机性检验组件

(5)运行完成后,对纯随机性检验组件右键,选择查看报告,纯随机性检验的报告结果 如图 20 所示。

		算法运行报告	
		如果p值小于0.05时,可以证明通过白噪声检验!	
lags	pvalue		
1	1.0609907508070775e-08		
2	3.186627383120427e-13		
3	9.234410613191885e-17		
4	2.742740357005688e-20		
5	6.029061100701443e-23		
6	1.2370671988737928e-24		
7	1.58594492078062e-25		
-			

图 20 纯随机性检验报告

3.3 模型构建

3.3.1 ARIMA 算法

选择 ARIMA 算法模型,步骤如图 21、图 22 所示。

(1) 找到时序模型→ARIMA 组件。

(2) 拖入 ARIMA 组件,将生成训练数据和 ARIMA 组件连接。

(3)选择字段属性,单击更新数据,时序列勾选 cwxt_db_184_d_字段,时间列勾选 collecttime 字段。

(4) 选择参数设置,设置预测周期数的值为5,设置自回归项数p的值为0,设置差分次数d的值为1,设置移动平均项数q的值为3。



图 21 ARIMA 组件_字段属性

I程 O	📩 ତ୍ ବ୍ <mark>120%</mark> % 🖹 오	> 字段屬性	
▼ 我的工程	(=) té \ 30	◇ 基础参数	
▲ 見要网站用户		预测周期数	0
▲ 应用系统负载		10	
	▲ 数据师选	自回归项数p	0
组件	國性变换	0	
▶ 緊美		差分次數d	0
3 ARIMA	AL-P-11/2#24HB	1	¥.]
3 ARCH		移动平均项数q	0
つ指数平滑法		3	
り GM(1,1)	● 平稳性检验		
D GARCH			
▶ 关联规则	(4) 种酶机性检验		
▶ 模型评估			
▶ 模型预测		1	
▶ 案例			
▶ 协同过渡			
横型		> 组件描述	

图 22 ARIMA 组件_参数设置

(5)运行完成后,对 ARIMA 组件右键,选择查看报告, ARIMA 的报告如图 23 所示。

		算法运行报告
		模型参数
		需要配置的参数及其取值如下。
参数名称	参数值	
自回归项数p	0	
差分次数d	1	
移动平均项数q	3	
(模型具体信息
		模型具体信息如下.
		Statespace Model Results

图 23 ARIMA 的报告