



长沙学院

# 高校大数据建模探索实例

报告人：许向阳

2016年12月9日

# 报告提要

一、高校大数据三个小问题

二、学生综合预警系统建模

三、系统实施计划及其思考

# 一、高校大数据三个小问题



# 高等学校智慧校园技术参考模型



老师



学生



教育管理者



社会公众

## 智能终端层

PC

智能终端

自助设备

可穿戴设备

## 智慧应用层

智慧型人才培养

智慧型科学研究

智慧型管理决策

智慧型社会服务

智慧型文化传承创新

## 大数据层

主体虚拟映像

数据服务平台

数据库  
数据仓库

元数据  
管理

数据挖掘  
引擎

数据分析  
平台

## 云计算层

HPC

Storage

虚拟化

IAAS

## 网络通信层

3G/4G

WIFI

IPV4/IPV6

## 智能感知层

ZigBee

WSN

RFID

IP CAM

信息安全保障体系

信息化运维服务体系

六层结构的参考模型

智能终端层

智慧应用层

大数据层

云计算层

网络通信层

智能感知层

# 三、学生综合预警系统建模

建模路径

1.  
业务建模

2.  
数据建模

3.  
应用建模

# 1、业务建模：高校学工痛点

---

大学生 → 跳楼、自杀、失联、裸贷

学工线 → 失明、失聪、失灵、失效

# 大学生心理素质提升示范校创建 暨特色成长辅导室建设专题会议现场

参会调研





# 业务建模：心理预警内容

分析学生在校的教学、生活、习惯等各类数据，结合已完成的学生心理健康分析，提供主动式的心理分析预警，判断有可能潜在的比较孤独、有心理潜在问题的学生。



# 业务建模：学校心理咨询数据

学院个案转介表

填表日期： 年 月 日

编号： 。

|        |                |
|--------|----------------|
| 学生姓名   | 联系方式           |
| 性别     |                |
| 年级     |                |
| 辅导老师   |                |
| 个案危机程度 | 轻度<br>中度<br>重度 |

长沙学院心理健康教育中心

系（院）学生心理状态汇总表

系（院） 心理联络员及电话 填表日期 年 月 日至 年 月 日 第 周

个案问题概述：  
1. 个案问题的具体描述  
爱问题、家庭问题、  
2. 个案背景资料

| 序号         | 姓名 | 班级   | 性别   | 联系电话 | 宿舍      | 异常心理表现及产生原因 |      |    |      | 心理联络员采取的措施及效果 |    |    | 是否建议心理中心跟进 |    | 心理中心处理情况 |    |    |
|------------|----|------|------|------|---------|-------------|------|----|------|---------------|----|----|------------|----|----------|----|----|
|            |    |      |      |      |         | 目前状况        | 入库原因 |    |      | 特殊问题          |    |    | 本月人数变动情况   |    | 退学       | 退出 |    |
| 男          | 女  | 休学状态 | 在校辅导 | 在校治疗 | 延期毕业    | 精神          | 神经   | 人格 | 其他   | 抑郁            | 自杀 | 自杀 | 伤人         | 休学 | 复学       | 退学 | 退出 |
| 2010级      |    |      |      |      |         |             |      |    |      |               |    |    |            |    |          |    |    |
| 2011级      |    |      |      |      |         |             |      |    |      |               |    |    |            |    |          |    |    |
| 2012级      |    |      |      |      |         |             |      |    |      |               |    |    |            |    |          |    |    |
| 2013级      |    |      |      |      |         |             |      |    |      |               |    |    |            |    |          |    |    |
| 2014级      |    |      |      |      |         |             |      |    |      |               |    |    |            |    |          |    |    |
| 延迟毕业       |    |      |      |      |         |             |      |    |      |               |    |    |            |    |          |    |    |
| 总数         |    |      |      |      |         |             |      |    |      |               |    |    |            |    |          |    |    |
| 本月在库出现的新情况 |    | 姓名   | 性别   | 年级班级 | 心理危机的变化 |             |      |    | 干预措施 |               |    |    |            |    |          |    |    |

本周整体学生心理健康状况

# 业务建模：需要关注的学生类别

经济困难

入党积极

考证痴迷

恋爱困难

竞赛队员

网络沉迷

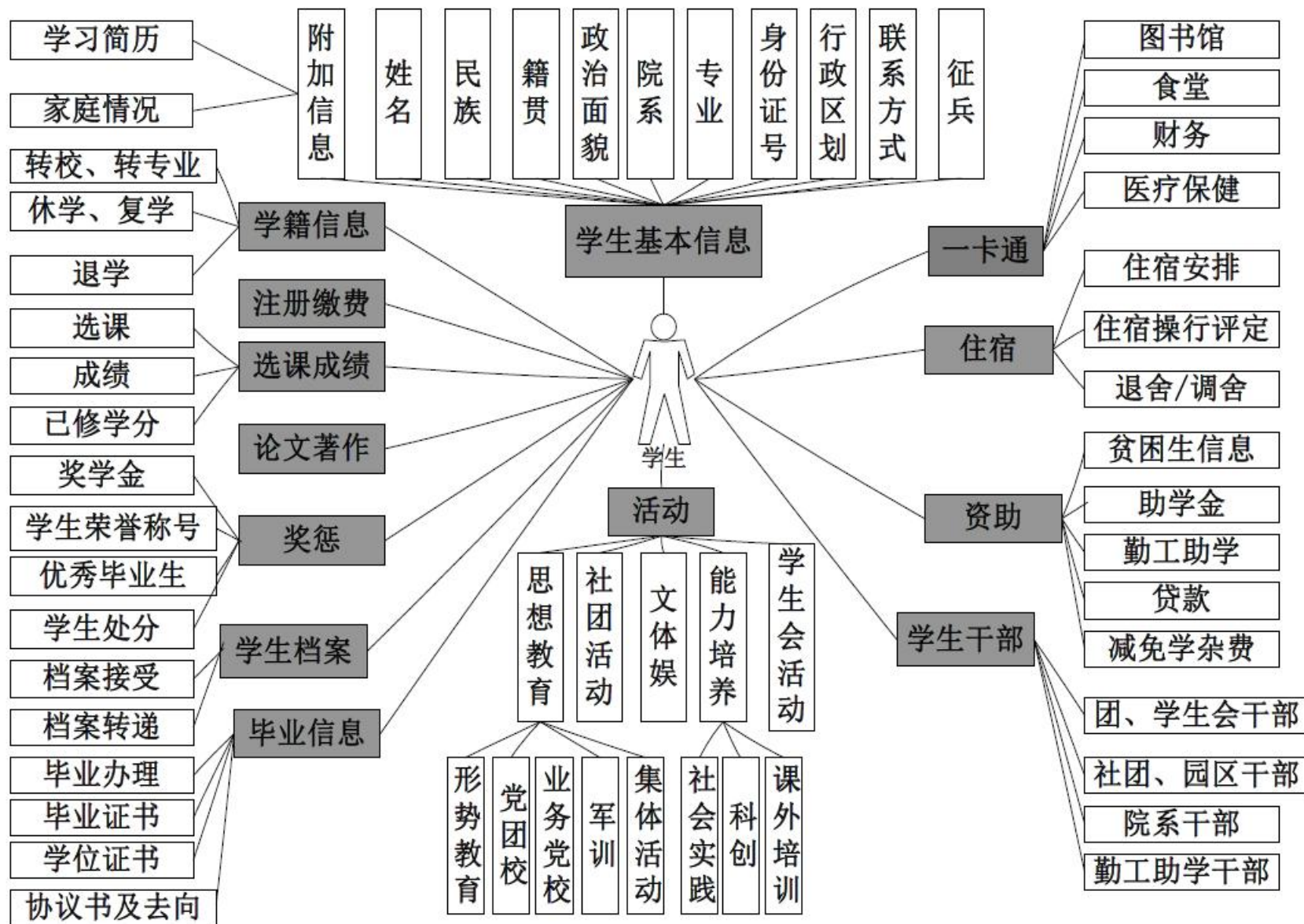
考研大军

就业迷茫

学业压力

心理问题

# 业务建模：学生个人数据档案



# 业务建模：学生个人数据来源



## 业务系统数据

✓ 教务系统

✓ 学工系统

✓ 图书管理系统

✓ 就业管理系统

迎新系统

## 硬件设备数据



✓ 上网行为认证

✓ 上网计费认证

一卡通系统

✓ 门禁系统

✓ 校园WiFi设备 手环手机APP

## 学 生 成 长 工 作 平 台



## 互联网数据

✓ 主流论坛网站：

✓ 天涯 百度贴吧 等

✓ 主流就业网站：

51job 中华英才网

✓ 主流科研网站：

✓ 知网 google学术

## 个人填报数据



✓ 上报方式：

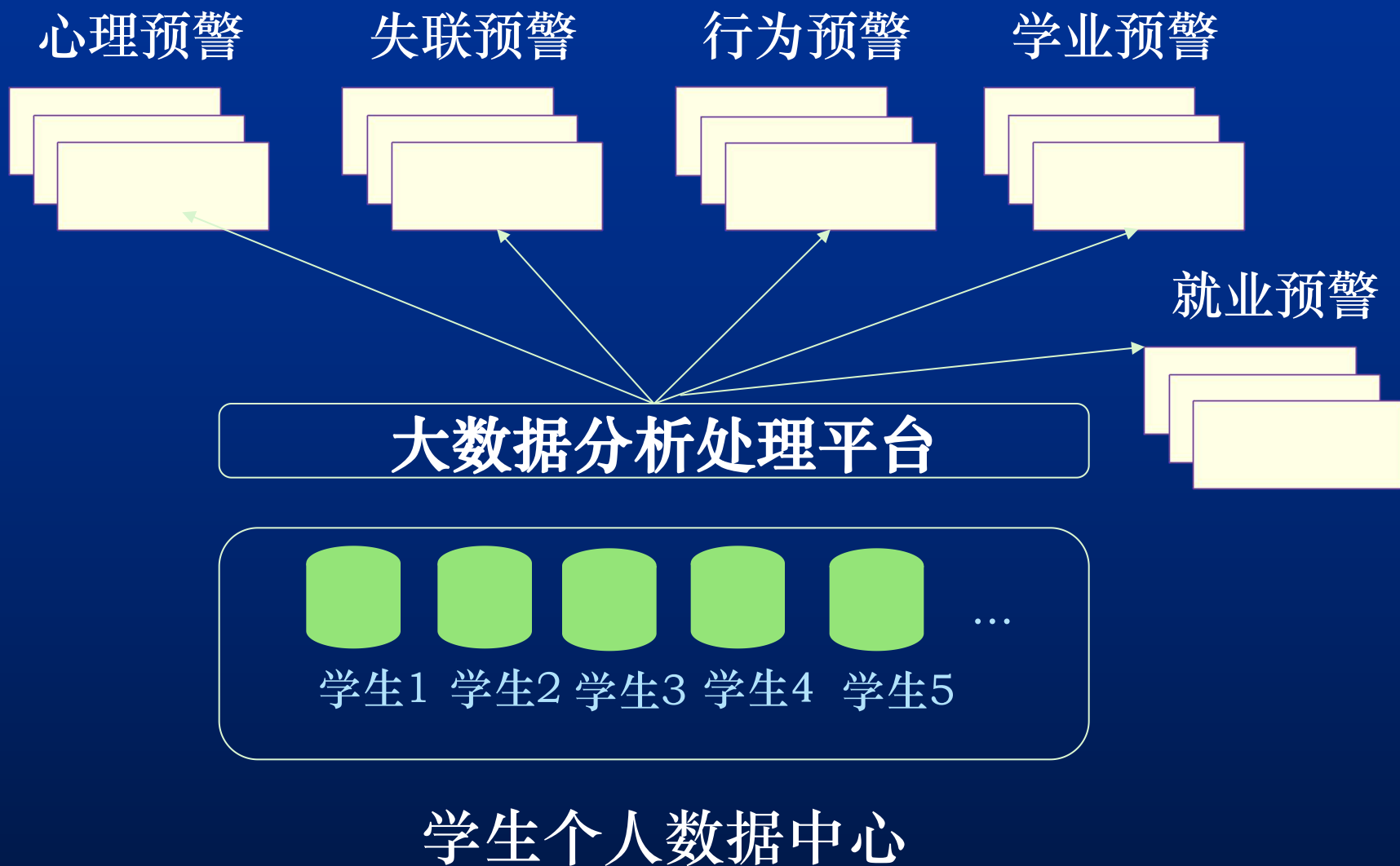
✓ PORTAL界面上报

✓ 上报人员：个人

✓ 教师 毕业生

✓ 管理员 辅导员

# 业务建模：学生综合预警逻辑结构



# 2、数据建模：学业预警模型算法

## ● 挂科预警

对每个科目，挂科预警目前考虑三个因素缺课率 $X_1$ 、平时成绩 $X_2$ 、学生的不合格率 $X_3$ 和实际结果 $Y$ （0不合格，1合格），先验概率为科目历史不及格率。采用贝叶斯分类方法分析挂科概率，当挂科概率超过一定程度时提出预警。

实际上我们需要通过贝叶斯概率模型求一条件概率。

$$p(Y|X_1, X_2, X_3) = \frac{p(Y)p(X_1, X_2, X_3|Y)}{p(X_1, X_2, X_3)}$$

用朴素的语言可以表达为：

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

在假设独立的条件下，模型表示为， $p(Y)$ 为先验概率：

$$p(Y|X_1, X_2, X_3) = \frac{p(Y) \prod_{i=1}^3 p(X_i|Y)}{p(X_1, X_2, X_3)}$$

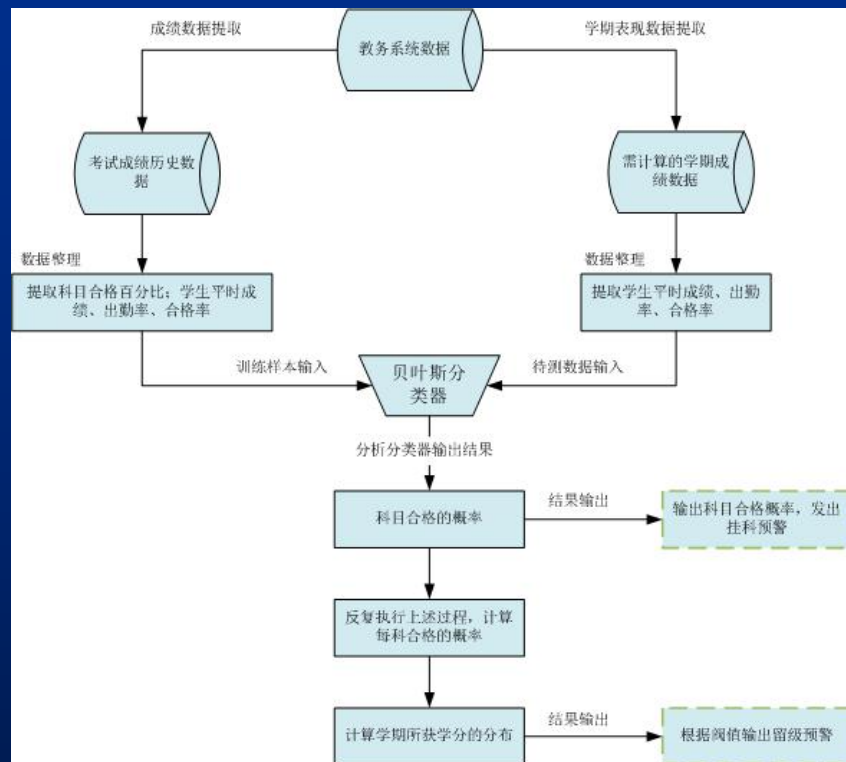
## ● 留级预警

学校对于学生都有要求每一学年必须完成的学分。根据该学期挂科概率估计该学期所获学分 $Y_1+Y_2+\dots+Y_n$ （ $Y_n$ 为科目 $n$ 的学分）分布，当学分小于所需学分 $S$ 的概率高于某阈值 $\gamma$ 时提出预警。

$$P(Y_1+Y_2+\dots+Y_n < S) \geq \gamma$$

## ● 毕业预警

当所获取学分 $F$ 和剩余学分 $R$ 之和接近或小于毕业所需学分 $T$ 时预警。 $F + R < T + \epsilon$



图：学业预警流程

# 数据建模：行为预警模型算法

## ● 作息时间异常

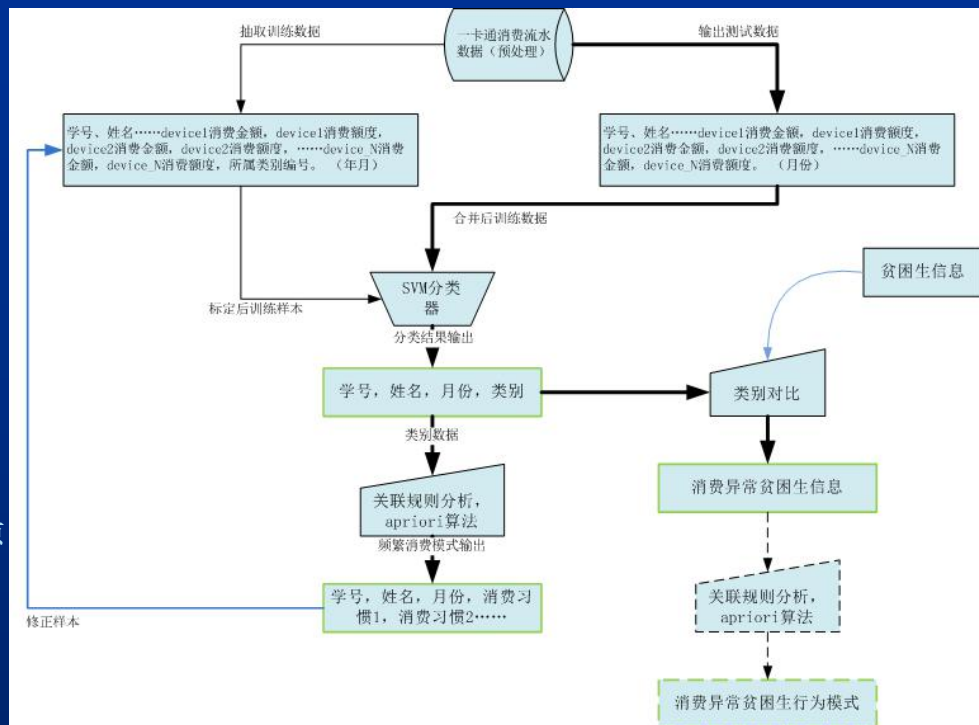
从学生的**深夜上网行为**、**上课考勤**、**早午晚餐时间**、**自修时间**情况分析学生的**作息时间是否正常**，主要采用与普遍正常作息时间对比的方法。对一段时间内作息时间异常的学生作预警。

## ● 长期缺席异常

对有**长期请假**、**缺课**的学生做出预警，结合请假记录分析其内在原因，对有疾病、家庭变故、感情变故等缺席内在因素的学生。

## ● 消费行为异常

通过横向和纵向的方法分析学生的消费情况，横向对比某学生与总体消费的情况对比，**分析学生中的奢侈和贫困情况**，并作出适当的预警；纵向可以使用时间序列分析的方法（具体可以用回归、移动平均的方法）分析某学生消费的变化情况，作出适当预警。



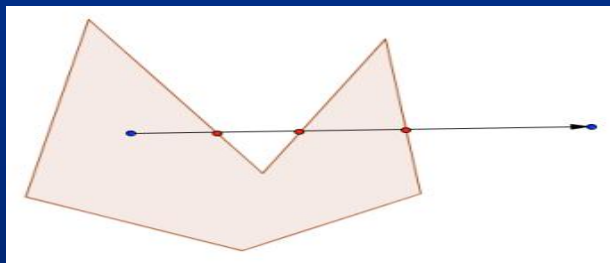
图：消费行为分析



# 数据建模：行为预警模型算法

## ● 校园轨迹异常

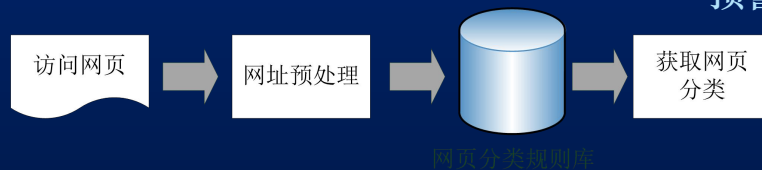
当某学生在某时间出现在不该出现的地方，或在该出现的地方不出现的时候作出预警。**考虑使用门禁记录、WIFI定位、考勤记录数据**，对诸如学生深夜出现在教学楼、学生长期不在宿舍等情况作预警。



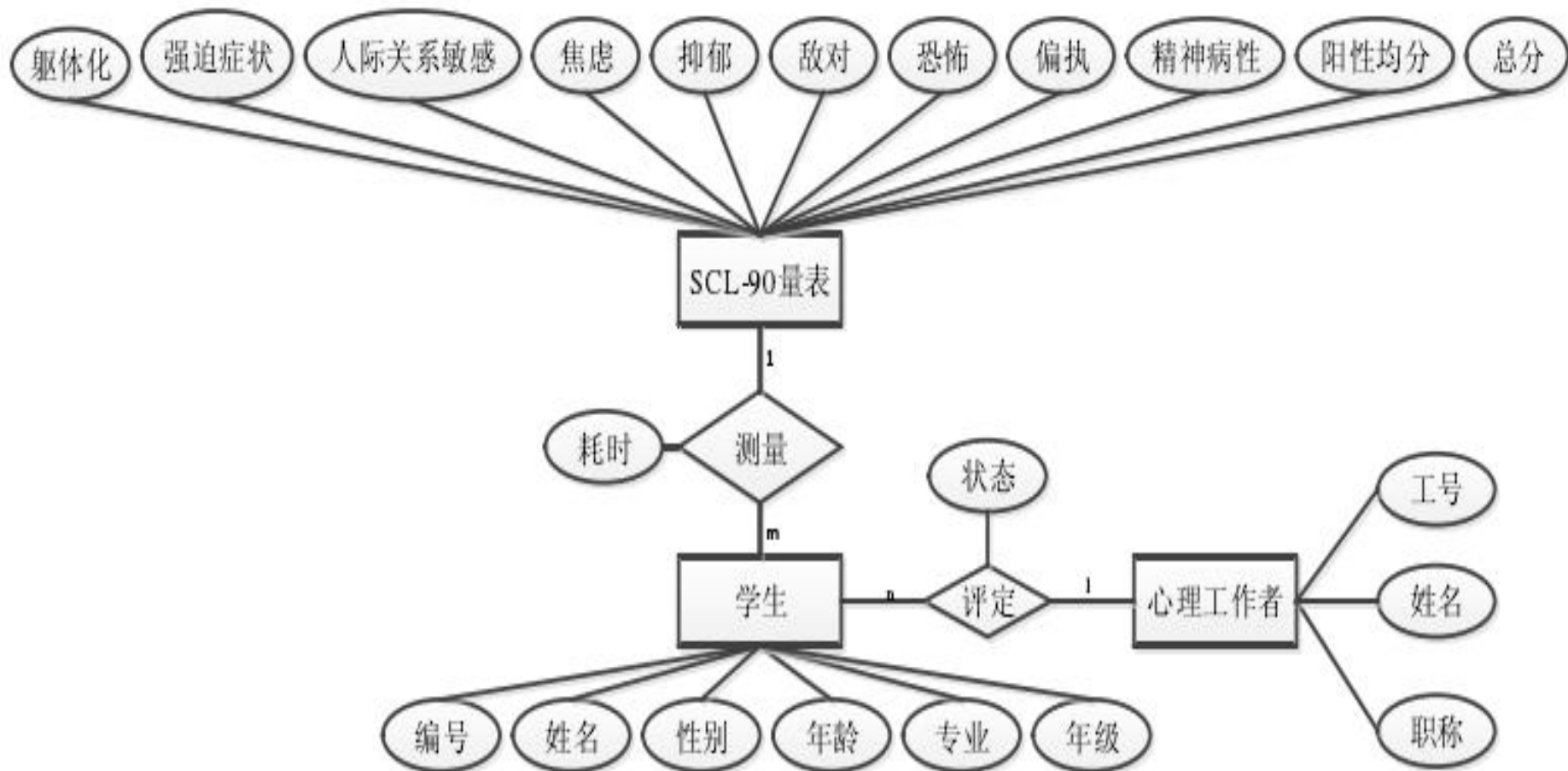
图：射线法判断  
学生所在区域

## ● 上网行为异常

根据上网审计系统记录，分析学生访问的网站信息，**分析学生极其过度沉迷游戏网站、购物网站、视频网站、黄色网站.....**的情况，作出适当预警。

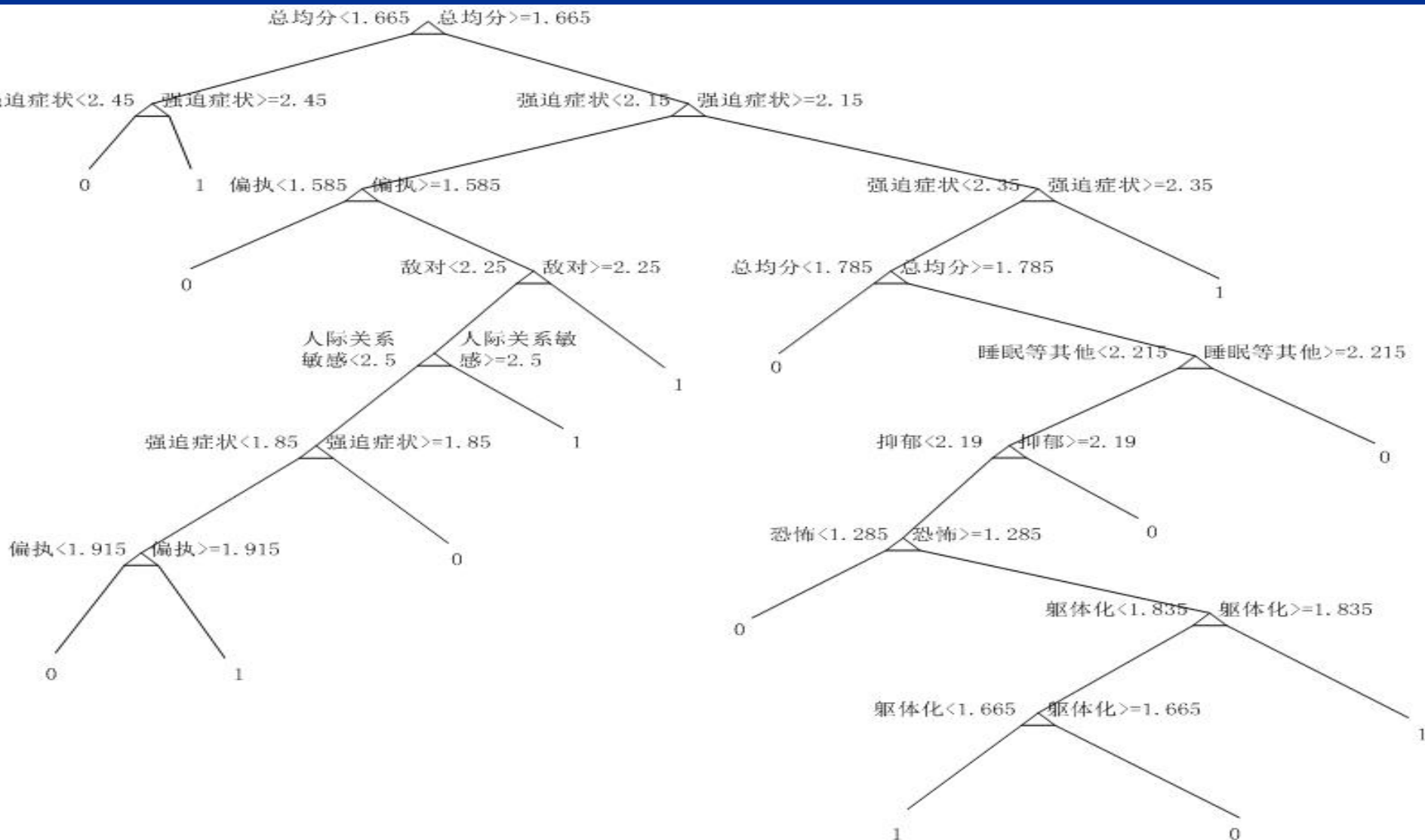


# 数据建模：心理预警异常分析表



世界著名的SCL-90心理分析图

# 数据建模：心理异常预警算法



CART 决策树分析结果 (0正常, 1异常)

# 3、应用建模：个人数据中心查询展示



个人基础信息展示，包括姓名、性别、出生日期、身份证号、家庭住址、学籍信息等。

|         |                   |       |             |       |  |
|---------|-------------------|-------|-------------|-------|--|
| 姓名:     | 陈华妮               | 性别:   | 女           | 出生日期: |  |
| 身份证号:   | 46000619951122021 | 民族:   | 汉族          | 政治面貌: |  |
| 学籍编号:   | 2014-09-01        | 家庭住址: | 海南省 海口市 万宁市 | 籍贯:   |  |
| 毕业学校:   | 2014-09-01        | 专业名称: | 地理与城乡规划学院   | 专业名称: |  |
| 毕业中学:   | 万宁市第二中学           | 入学日期: |             | 毕业日期: |  |
| E-Mail: |                   | 电话:   |             | 微信号:  |  |
| 备注:     |                   | 手机号码: |             | QQ号:  |  |

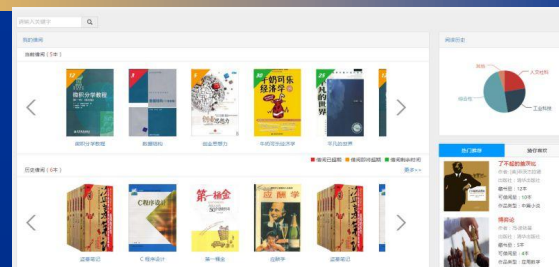
## 个人基础信息



个人课程信息展示，包含课程列表、课程名称、学分、学时、成绩等。

| 序号 | 课程名称                  | 学分 | 学时 | 成绩 |
|----|-----------------------|----|----|----|
| 1  | 2014-2015-1学期 高等数学    | 4  | 4  | 80 |
| 2  | 2014-2015-1学期 大学英语    | 2  | 2  | 85 |
| 3  | 2014-2015-1学期 计算机组成原理 | 2  | 2  | 85 |
| 4  | 2014-2015-1学期 英语(二)   | 1  | 2  | 85 |
| 5  | 2014-2015-1学期 大学英语(二) | 2  | 2  | 85 |
| 6  | 2014-2015-1学期 大学英语(二) | 5  | 6  | 85 |
| 7  | 2014-2015-1学期 大学英语(二) | 3  | 0  | 85 |
| 8  | 2014-2015-1学期 大学英语(二) | 4  | 3  | 85 |
| 9  | 2014-2015-1学期 大学英语(二) | 1  | 2  | 85 |
| 10 | 2014-2015-1学期 大学英语(二) | 3  | 2  | 85 |
| 11 | 2014-2015-1学期 大学英语(二) | 3  | 3  | 85 |

## 个人课程信息



图书借阅信息展示，包含借阅记录、图书封面、借阅日期、归还日期等。

| 序号 | 书名    | 借阅日期       | 归还日期       |
|----|-------|------------|------------|
| 1  | 《...》 | 2014-10-01 | 2014-10-15 |
| 2  | 《...》 | 2014-10-01 | 2014-10-15 |
| 3  | 《...》 | 2014-10-01 | 2014-10-15 |
| 4  | 《...》 | 2014-10-01 | 2014-10-15 |
| 5  | 《...》 | 2014-10-01 | 2014-10-15 |

## 图书借阅信息



网络访问信息展示，包含访问时间、IP地址、流量、访问次数等。

| 序号 | 访问时间                | IP地址      | 流量(M)          | 访问次数 |
|----|---------------------|-----------|----------------|------|
| 1  | 2016-03-07 17:35:54 | 169.94.43 | 172.20.17.85   | 0    |
| 2  | 2016-03-07 00:57:30 | 56        | 172.20.17.85   | 0    |
| 3  | 2016-03-08 18:01:02 | 169.94.33 | 172.20.115     | 0    |
| 4  | 2016-03-09 18:30:57 | 218.94.50 | 172.20.108.43  | 0    |
| 5  | 2016-03-10 22:46:35 | 57        | 172.20.134     | 0    |
| 6  | 2016-03-11 13:18:59 | 106       | 172.20.224     | 0    |
| 7  | 2016-03-12 18:09:12 | 153.99.50 | 172.20.196     | 0    |
| 8  | 2016-03-15 18:17:31 | 439       | 172.20.17.85   | 0    |
| 9  | 2016-03-15 18:16:23 | 18.94.8   | 172.20.19.38   | 0    |
| 10 | 2016-03-15 18:19:13 | 439.94.33 | 172.20.108.43  | 0    |
| 11 | 2016-03-16 18:17:31 | 439       | 172.20.17.85   | 0    |
| 12 | 2016-03-15 18:16:23 | 1.94.8    | 172.20.19.38   | 0    |
| 13 | 2016-03-15 18:19:13 | 439.94.33 | 172.20.108.43  | 0    |
| 14 | 2016-03-14 18:27:49 | 18.94.8   | 172.20.131.253 | 0    |

## 网络访问信息



校园消费信息展示，包含消费记录、消费金额、消费地点、消费时间等。

| 序号 | 消费时间                | 消费地点       | 消费金额 |
|----|---------------------|------------|------|
| 1  | 2016-03-01 10:00:00 | 南区二食堂二楼消费点 | 8    |
| 2  | 2016-03-01 11:58:59 | 南区二食堂二楼消费点 | 9    |
| 3  | 2016-03-01 13:19:10 | 南区二食堂二楼消费点 | 7    |
| 4  | 2016-03-04 8:22:22  | 南区二食堂二楼消费点 | 21   |
| 5  | 2016-03-04 15:49:40 | 南区二食堂二楼消费点 | 5    |
| 6  | 2016-03-04 17:53:14 | 南区二食堂二楼消费点 | 5    |
| 7  | 2016-03-04 17:54:15 | 南区二食堂二楼消费点 | 5    |

## 校园消费信息



校园轨迹信息展示，包含学生姓名、日期、轨迹图、轨迹描述等。

| 学生姓名 | 日期         | 轨迹描述       |
|------|------------|------------|
| 陈华妮  | 2016-10-21 | 图书馆-教学楼-宿舍 |
| 陈华妮  | 2016-10-21 | 图书馆-教学楼-宿舍 |
| 陈华妮  | 2016-10-21 | 图书馆-教学楼-宿舍 |

## 校园轨迹信息



个人成绩信息展示，包含课程名称、学分、成绩、排名、平均分、标准差等。

| 学期          | 科目           | 开课学院     | 性质   | 学分 | 成绩    | 排名   | 平均分 | 标准差 |         |       |      |
|-------------|--------------|----------|------|----|-------|------|-----|-----|---------|-------|------|
| 2014-2015-1 | 计算机科学与技术系列课程 | 德县科学技术学院 | 专业必修 | 2  | 90.00 | 4.00 | 优秀  | 20  | 100.00% | 90.41 | 1.18 |
| 2014-2015-1 | 计算机科学与技术     | 德县科学技术学院 | 专业必修 | 2  | 80.00 | 3.00 | 良好  | 9   | 100.00% | 74.59 | 6.62 |
| 2014-2015-1 | 计算机组成原理      | 海南政法职业学院 | 公共必修 | 2  | 74.00 | 2.40 | 中等  | 58  | 100.00% | 74.59 | 5.40 |
| 2014-2015-1 | 大学英语(一)      | 海南政法职业学院 | 公共必修 | 3  | 62.00 | 0.00 | 及格  | 29  | 89.13%  | 64.76 | 8.39 |
| 2014-2015-1 | 大学生职业生涯规划    | 海南政法职业学院 | 公共必修 | 1  | 88.00 | 3.80 | 良好  | 31  | 100.00% | 87.02 | 6.61 |
| 2014-2015-1 | 英语(二)        | 海南政法职业学院 | 公共必修 | 1  | 67.00 | 1.70 | 及格  | 19  | 100.00% | 78.47 | 8.17 |

## 个人成绩信息



在线学习信息展示，包含课程名称、课程简介、课程视频、课程评价等。

| 课程名称           | 课程简介                        | 课程视频 | 课程评价 |
|----------------|-----------------------------|------|------|
| 北京大学公开课：世界文化地理 | 本课程旨在帮助学生了解世界文化地理知识，提高地理素养。 | 视频播放 | 好评如潮 |
| 宇宙时空之旅         | 本课程带你探索宇宙的奥秘，感受时空的奇妙。       | 视频播放 | 引人入胜 |
| 斯坦福：总统选举中的地理学  | 本课程探讨总统选举中的地理因素，分析地理对政治的影响。 | 视频播放 | 见解独到 |

## 在线学习信息

# 应用建模：学生综合预警界面

各类预警可视化界面

### 今日学业预警详情

546 条

**心理预警分析**

- 189条 逃课预警
- 150条 低分预警
- 91条 学习敷衍预警
- 43条 期末学业预警

刷新 一级预警 二级预警 三级预警
0% → 周

| #  | 持续时间 | 开始日期      | 学号         | 姓名  | 性别 | 年级   | 辅导员 | 学院      | 专业       | 预警说明          | 预警级别 |
|----|------|-----------|------------|-----|----|------|-----|---------|----------|---------------|------|
| 1  | 5天   | 2016.7.29 | 1143021234 | 杭武  | 男  | 2015 | 林强  | 计算机学院   | 计算机科学与技术 | 该学生当前处于心理异常状态 | 一级   |
| 2  | 3天   | 2016.7.29 | 1143021234 | 梁泽  | 女  | 2015 | 徐强  | 软件学院    | 软件工程     | 该学生当前处于心理边缘状态 | 一级   |
| 3  | 1天   | 2016.7.30 | 1143021234 | 齐露  | 女  | 2015 | 刘强  | 文学与新闻学院 | 中文学      | 该学生当前处于心理状态不好 | 三级   |
| 4  | 3天   | 2016.7.31 | 1143021234 | 李守善 | 女  | 2015 | 李强  | 历史文化学院  | 历史学      | 该学生当前处于心理边缘状态 | 二级   |
| 5  | 5天   | 2016.8.1  | 1143021234 | 武青肃 | 男  | 2015 | 张强  | 建筑学院    | 建筑学      | 该学生当前处于心理异常状态 | 三级   |
| 6  | 5天   | 2016.8.1  | 1143021234 | 徐春风 | 女  | 2015 | 钟强  | 制造学院    | 机械制造学    | 该学生当前处于心理异常状态 | 二级   |
| 7  | 1天   | 2016.8.2  | 1143021234 | 向以辰 | 男  | 2015 | 丁强  | 临床医学院   | 临床医学     | 该学生当前处于心理状态不好 | 三级   |
| 8  | 1天   | 2016.8.2  | 1143021234 | 向以辰 | 男  | 2015 | 丁强  | 临床医学院   | 临床医学     | 该学生当前处于心理状态不好 | 三级   |
| 9  | 5天   | 2016.8.1  | 1143021234 | 武青肃 | 男  | 2015 | 张强  | 建筑学院    | 建筑学      | 该学生当前处于心理异常状态 | 一级   |
| 10 | 3天   | 2016.7.31 | 1143021234 | 李守善 | 女  | 2015 | 李强  | 历史文化学院  | 历史学      | 该学生当前处于心理边缘状态 | 三级   |
| 11 | 3天   | 2016.7.29 | 1143021234 | 梁泽  | 女  | 2015 | 徐强  | 软件学院    | 软件工程     | 该学生当前处于心理边缘状态 | 二级   |

邮箱 短信

### /\*计算机学院\*/近期预警情况

Analysis Result

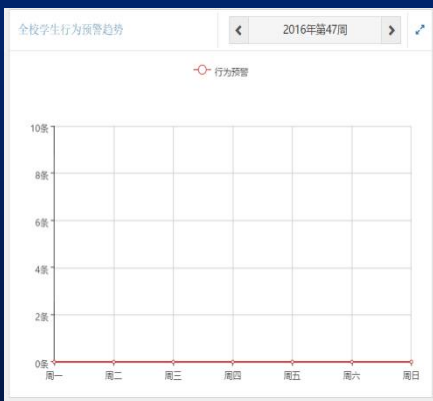
**数据分析**

- 逃课日均消费过低
- 高分日均消费过低
- 资金存可能存在心理问题
- 高分日均消费过低

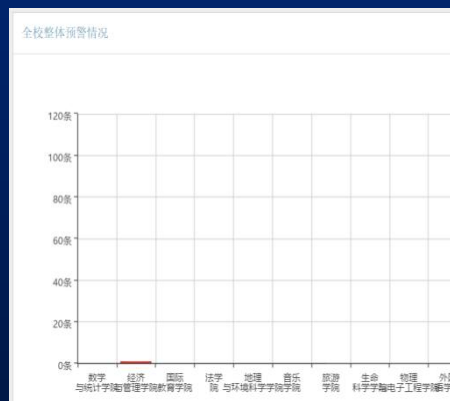
**数据总结**

- 今日全校平均上网时长为：2h
- 上网时长最高学院：计算机学院

原因分析报告



预警后期趋势



危机干预曲线

## 三、系统实施计划及其思考

- 1、学工牵头，技术推动
- 2、依托学科，共建共享
- 3、校企合作，定制开发
- 4、不断优化，迭代收敛
- 5、适度预算，多种来源

---

**欢迎交流，共同切磋！**