

# 学习总结

今年暑假，来自全国各地250余名可视化爱好者齐聚浙江大学，为期四天的可视化暑期学校于7月12日早上8点半正式开幕！

此次学习也是我第一次进行一次系统化地学习可视化，虽然课时不长，但是却感觉收获满满。接下来，我就每天的学习情况，进行我个人的学习总结。

Day. 1

## 陈为老师：可视化与数据可视化概论

——什么是可视化？陈老师一开场便抛出了这个问题。

陈老师先用几个鲜活的可视化实例进行引入，为大家展示并讲解了纽约上空飞行模式可视化，弦图，以及去年美国总统大选可视化，展现可视化的魅力。

然后给出数据可视化的概念：利用人眼的感知能力对数据进行交互的可视表达以增强认知的技术，称为可视化。

而可视化有三大作用：

- 表达数据
- 分析数据
- 用数据进行沟通

——为什么要可视化？

增强人的认知，做出更高效、正确的判断，帮助我们思考，看到传统统计分析所看不到的内容。

——大数据为什么要可视化？

可视化是基础自然学科的必要手段，是科学大数据发展的必需。

陈老师介绍了Nature上的两篇工作，分别是蛋白质分子的折叠，人手动做量子移动的效率比计算机要快。

可视化是进行海量行为挖掘、分析、建模、理解的重要途径，在大工程、大安全、物联网与智慧城市、互联网与社交媒体等领域都应用广泛。

## 吴向阳老师：数据以及处理

吴向阳老师着重地介绍了可视化领域在数据的获取、处理、分析阶段发挥的关键性作用。

他简要地介绍了关于数据的几个概念——

- 一、数据的类型：序数性、数值型。
- 二、数据之间的相似性比较：集合相似度、海明相似度。
- 三、对数值型数据的计量：曼哈顿距离、欧式距离、夹角余弦。
- 四、对序数型数据的计量：皮尔逊相关系数、动态时间扭曲、最大公共子序列。

而数据常见的质量问题有**数据重复**、**噪声和离群值**、**数据缺失**、**数据一致性**等。

为了解决这几个问题，我们通常需要数据清理，常见的方式包括**数据去噪**、**数据补全**、**数据更正**。

同时，为了降低数据的复杂度，通常会进行数据精简，同时保持数据的内涵特征，常用方法包括**统计**、**聚类**、**降维**、**特征抽取**等。

最后，吴向阳老师还简要地举了几个例子，如Facebook的相关工作、D-Dupe等，让我们更加清晰地了解数据处理得关键问题。

## 巫英才老师：社交媒体数据可视分析

巫老师的报告分为三个部分，分别是社交媒体大数据可视分析的概念、基本任务和案例讲解。

一、概论

社交媒体是人们撰写、分享、评价、讨论、互相沟通的互联网平台。

通过人们在社交媒体上的活动，便产生了社交媒体数据，社交媒体数据兼有**信息属性**和**社会属性**。

作为信息的载体，它具有模态多样、动态实时、规模海量、语义丰富的特点；作为网络载体，它又具有关系异构、结构复杂、规模海量、交叉演化的特点。同时它还承载海量用户行为与观点。

## 二、任务方法

社交媒体数据可视化的基本任务分为三个部分：

### 1. 展现和分析人们的社交行为

它包括展现信息的传播过程，影响话题的兴衰，呈现人群的移动特征，展现用户的个体行为。

### 2. 呈现和理解人们的社交网络

它包括展现社交网络结构，展现网络结构与语义信息。

### 3. 获取和探索社交媒体的信息

主要的相关工作则有：基于关键词的可视化，基于话题的可视化和基于多角度信息的可视化。

## 三、实例

在这一部分，巫老师简要地跟我们介绍了可视化技术是如何帮助分析人员迅速地从社交媒体上发现和获取有价值的信息，并向我们介绍了一个交互式可视分析系统的具体流程。

	展现和分析人们的社交行为	呈现和理解人们的社交网络	获取和探索社交媒体的信息
数据	内容（文本、图片、视频）、网络结构信息	网络结构信息	内容（文本、图片、视频）
分类	1. 展现人们的集体行为 <ul style="list-style-type: none"><li>传播信息的过程</li><li>影响话题的兴衰</li><li>呈现人群的移动</li></ul> 2. 展现用户的个体行为	1. 展现社交网络结构 <ul style="list-style-type: none"><li>静态网络</li><li>动态网络</li></ul> 2. 呈现网络与语义信息	1. 基于关键词的可视化
方法	人机交互、话题挖掘、观点挖掘、图可视化、文本可视化、时序可视化、多维数据可视化、时空信息可视化、多媒体数据可视化	人机交互、图可视化、社交计算、网络分析	人机交互、话题挖掘、观点挖掘、信息检索、命名实体识别、文本可视化、图可视化、时序数据可视化、地理信息可视化、多媒体数据可视化
应用	客户关系管理、媒体舆情监控、网络安全监控、城市公共安全、媒体危机公关、异常行为检测等	社会学研究、媒体危机公关、国土安全与反恐、商务智能决策、客户关系管理、社交市场推广等	城市公共安全、商务智能决策、自然灾害检测、国土安全与反恐、媒体资料采集等

### 张嵩老师：医学影像数据可视化

医学可视化是可视化研究领域中的一个重要的方向。

数据源主要来自：x-ray、CT、PET、MRI、超声波。

应用于：对病人的诊断，帮助医生制定手术前的计划，在医生的手术中提供支持，帮助训练病人以及教育医生。

对标量场进行可视化的方法包括：体可视化，等值面、等值线的可视化。

其中，体可视化的步骤包括：

1. 建立传递函数
  2. 射线投影
  3. 在不同方向的成像
- （没太听懂）

### 陈伟锋老师：数据可视化的基础

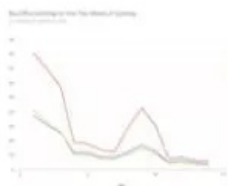
陈老师的报告分为三大部分：数据变换（统计）、数据可视化（统计图表）、数据可视化工具（统计图表制作工具）。这三部分也是可视化领域入门的必备知识，需要我们牢牢掌握的。

#### 数据变换：

一般是为了某个特定的目标去展现数据的本质。常见有线性变换和对数变换。

其中，还有一些手段可以对数据做一些简化，如采样、降维、聚类等。

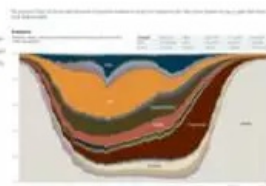
在前一部的基础上，我们使用可视化进行对数据的表达。陈老师介绍了6种基本的统计图——



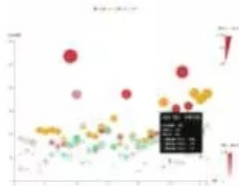
折线图



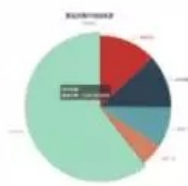
柱状图



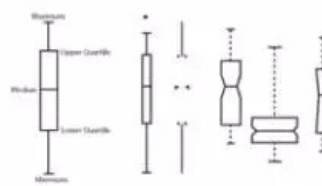
堆积/堆叠图



散点图



饼状图



盒须图

随后，陈老师又列举了大量的工具和配色选取工具，帮助没有经验的同学们进行进一步的可视化学习。

## Day. 2

鲁爱东老师：时变数据可视化

鲁老师介绍了随着时间变化、带有时间属性的时变数据，并且进一步地带我们了解了时变数据的可视化方法要点。她也举了几个例子，让我们对其印象更加深刻。

目的是为了更好地对数据进行预测，并方便人们进行计划。

其中，设计方案和可视化的交互实现都是时变数据处理的难点。

设计方法：

鲁老师提出了一个问题：我们如何在设计可视化方案中，可以一目了然地看出时间属性对数据的影响呢？

这里，她给出了几个设计的思路：

1. 小图：不同时间不同属性变化比较。固定参数选择区域染色一样，选取数据的角度也要一致

举例：可视化利用率高 一到二 二到多

ThemeRiver：整体性、增加减少 线性的表达

螺旋状图：非线性 调整周期率寻找周期

有多个属性（举例：地理位置 时间轴）

Visual Sedimentation：上面 下面 四周 散点的数据加进来

不适合分析严肃数据

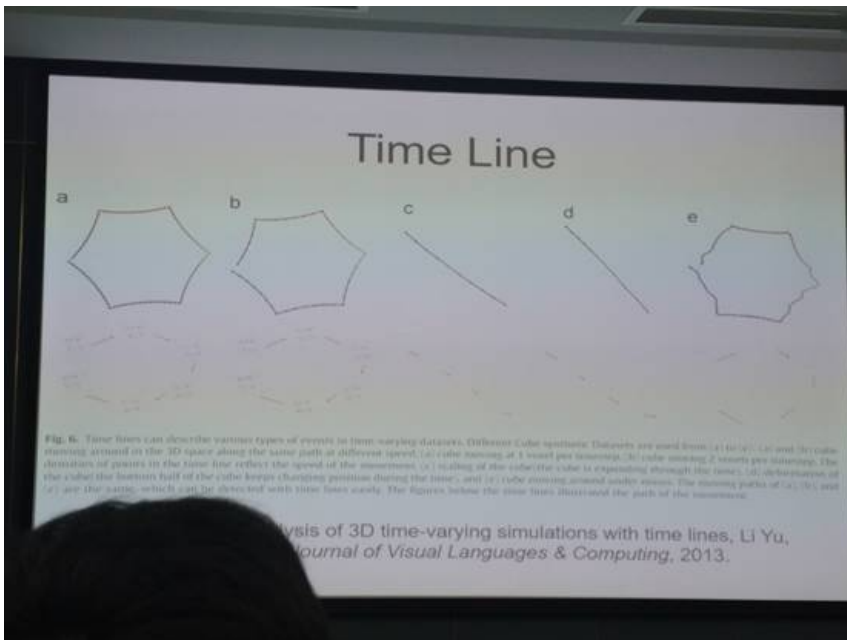
Whisper：向日葵 中间长出 两边成熟

严格的数据

Twitter Monito

2. 简单线条描述时序图（Time Line）：不同线性变化计算出来的

线条形状用来描绘不同事件的变化、时序数据



高维数据 -> 线条数据:

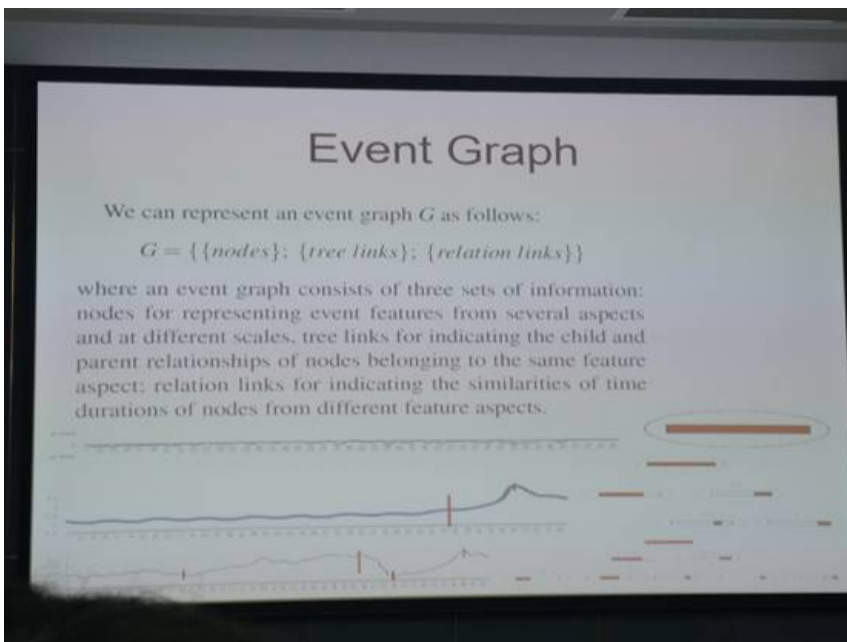
每一个点不是同样重要: 数据的边缘位置、最大值、拐角特殊位置

(选择相对较少的一个点, 点分别代表边缘, 描述数据的不同特征)

例子: Virtual Words

不同时间相似度是多少?

例子: Storytelling——动态转换方法



在课程的后期, 鲁教授讲解了Visual Sedimentation、Whisper等多个实例, 向我们生动地展现了时变数据可视化的广泛应用、有效性以及趣味性。

最后鲁教授通过对她与浙江大学合作完成的两项工作的详细介绍, 让同学们对时变数据的可视化设计和实现过程有了一定的了解。

张小龙: 文本数据可视化 (可视化第一步)

我们在可视化的进行过程中, 第一部就是要进行划分: 根据形状/颜色/移动来对数据进行划分。

特点: 快、无干扰

文本数据可视化的设计有三个重要因素:

1. Preattentive cues for text of interest;

2. Focus + Context
3. Coordinated Multiple View

举例：Coordinated Multiple Views多视图：不同角度联系在一起  
 鱼眼：牺牲空间换取信息

文本数据的特点：

1. 一组文章可能有多个主题
2. 每一个主题有多个词组成

为了处理海量的文档，从中分析文档之间的联系，挖掘有效信息，我们需要引入主题模型，利用自然语言处理技术，使用由若干个高频词组成的主题来表达每个文档。主题模型使得我们可以通过计算获得文档、主题和词语这三者之间的关系，分析这些关系来快速和直观地获取和分析信息。

主题词：多个文档，文档由多个主题，由主题词分类

隐含的话题搜索LDA：机器学习/输出：一组词

LDA：是一种统计方法 很难捕捉零星但重要的信息

他也给我们提议，要自己做一个词典（标记）

王叙萌：可视化中的交互

在可视化中，人机的交互占了重要的地位。基于大量实例，王博士依次对选择、浏览、重构、编码、抽象/具象、过滤和关联七种常见交互方式进行了详细的讲解。

## 七种常见可视化类别

类别	内容
选择	当可视化中的元素较多时，选择操作有助于用户在可视化中追踪感兴趣的数据元素。
浏览	浏览操作让用户主动寻找并调动可视化程序去寻找感兴趣的数据。
重配	通过对可视化元素的重新布局，有助于解释蕴含于数据中的信息。
编码	改变可视化编码，有针对性地表达数据中的特征。
抽象/具象	面对大规模的数据，可视化通常需要先简化数据再进行显示用户关注的细节数据。
过滤	通过过滤操作，相关数据被更好展示，便于用户观察可视化结果中的图案。
关联	高亮显示数据对象间的联系，用户可以同时观察数据的不同属性。

用户的探索基于表达

过程：

解决的问题 -> 需要的数据 -> 编码设计 -> 怎么样操作？操作抽象化 -> 操作交互实现

### 一、选择

操作：鼠标单击/悬停

间接选择/直接选择

### 二、浏览

通过改变用户的视角

### 三、重配

显示不同的排列提供不同的角度

例：重组视图/重新排序



#### 四、编码

显示一个不用的表示方式

是可视化的河西要素

改变视觉外观

例：颜色 大小 方向 字体 形状

偏向于应用系统，用起来舒服

#### 五、抽象/具象

显示更多或更少的数据细节

调整的抽象级别

例：展开子类别/缩放

（提供一个小窗口

概览+细节

尺度问题：数据太大，不能再屏幕上显示完整

可能包含太多的事例/变量

解决：数据的表达、用户交互、两者兼备

#### 六、过滤

显示符合条件的某些数据子集

基于某些特定的条件改变现实的数据集

例：动态查询

#### 七、关联

不同的纬度看数据

可视化设计：

自下而上的探究方法

翁荻：可视化系统搭建

1. 多平台支持
2. 良好的交互响应性
3. 对开发人员要求高

mean.js框架介绍——

一、存储数据：数据库

**MongoDB**是NoSQL数据库， 可以进行复杂的查询

2dsphere索引

利用索引检索地理位置。。

二、搭建可视化系统后端

①方便快捷地处理数据库的数据

Node.js与MongoDB对接

②大量并发连接

单线程异步模型

同一时间仅有一次javascript代码被执行

### ③复杂的可视化数据计算

javascript性能高

Node.js的C++ Addon: 提供可被javascript调用的C++程序

例: 数据挖掘模块

头文件 -> 模块 -> 使用node-gyp编写原生模块(使算法消耗时间减少到系统可执行时间)。。

### 基于html模板的渲染

### ④开发服务器难度降低

#### 三、前端框架

angularJS非常流行的框架

MVW框架

动态模板加载

客户加载

双向数据绑定:

①javascript -> 网页

②页面 -> 变量代码

只读的数据绑定

与可视化结合:

一套可视化的代码

①用angular的directive封装

②用angular.js可视化组件

例: 散点图可视化

建立model、编写view、添加交互(绑定事件)

例子:

处理数据:

①python+pymongo

超过24小时

python字符串处理效率低

②C++、libbson和mongoDB的原生驱动

几小时

③独立的MongoDB提交线程、简单线程池和双缓冲机制

6分钟

前端可视化:

Boxplot扇形图

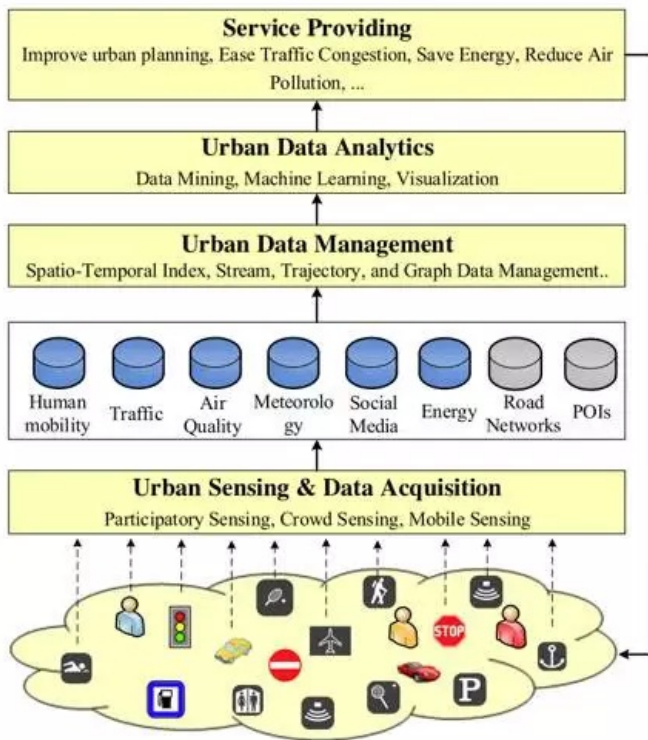
最小值、中位数、最大值: 金融数据、重建数据的趋势

Day. 3

孙道国老师: 城市数据可视化(空间、时空上的……)

孙老师在对城市数据进行可视化上, 主要分为三大部分进行讲解, 分别为: 城市数据的概论、可视化方法以及具体案例讲解。

一、概述



## 1、城市数据——

交通数据（GPS、卡口）

用户行为（微博签到、手机信号数据）

环境数据（空气质量、城市噪音）

经济数据（地区收入）

POI（兴趣点）、地铁数据……（静态数据：几年内不会变化）

**特点：高度异构、高纬度、高动态、冗余与稀疏并存、高动态、不确定性、非结构化**

框架：

数据 -传感器-> 不同类型数据 -数据库-> 存储、检索 -> 可视化（机器学习……） --> 城市化计算（用数据来治理城市）

挑战：

数据质量欠佳、不同城市数据的可是编码不一、时空可视探索问题（基础问题，寻找方案）  
解读得到的可视化、可视化换转化为知识

## 二、方法

时间数据可视化：

1. 线性时间（xy轴）折线图

2. 周期性时间（辐射性）

<http://browser.timeviz.net/>

空间数据可视化：

1. 点

热力图

2. 线条

路况 拿破仑东征图

地铁图 传播图

3. 区域（面）

人口迁移图

树图



基于节点（每个区域作为一点）

时空数据可视化：

1. 链接视图

跟踪

2. 集成视图

遮挡

三、案例（经济、交通……）

1. 交通流、公共交通时间表、基于轨迹透镜（Trajectory Lenses）的交互

2. 人群移动数据（签到）、时空社交媒体数据可视化、时空信息传播可视化

3. 空气质量可视化、南极臭氧层空洞可视化、多维度地理动态数据可视化

四、总结、挑战

1. 方法、系统适用性：数据敏感、不确定性

2. 多源异构数据问题：数据源之间层型/隐形关系、静态数据和流数据的结合

3. 系统可伸缩性：适合存储城市数据的数据结构、数据需求

4. 专家参与、评估

夏佳志老师：高维数据可视化

高维数据是指维度太高，以至于难以从维度集合中提取有意义的关联信息的数据。对于海量的高维数据，由于人的认知成本、视觉成本、计算机成本都是有限的，传统的统计和基本分析的方法显得捉襟见肘。我们需要新的方法弥补用户在领域知识和数据洞察上的不足。

一、概论

基于坐标轴的方法	基于图标的方法	基于像素图的方法
<ul style="list-style-type: none"><li>散点图矩阵</li><li>平行坐标图</li><li>平行集</li><li>星形图</li><li>雷达图</li></ul>	<ul style="list-style-type: none"><li>星形图</li><li>切尔诺夫脸谱图</li><li>Small Multiples</li></ul>	<ul style="list-style-type: none"><li>像素图</li><li>像素柱状图</li><li>马赛克图</li></ul>

二维数据：平面坐标表达

三维：三维空间坐标系（二维表达：变化为表格）

（慎用——遮挡、不直观）

高维多元数据：High-Dimensional

多维：数据具有多个独立属性

多元：多个相关属性

例子：

Iris：片/花瓣长、宽

问题：在视觉空间（二维平面、三维平面）上表达？

答：数据驱动+任务导向（取舍）。

展示：像素、图标方法

二、可视映射（必须）

基于坐标轴的方法——

散点图矩阵：使用一个二维散点图表达每对纬度、直观

例子：Iris

平行坐标轴：2、3维坐标系中坐标轴相互垂直（纬度相关性）

交互：“选取”、“刷取”

半透明平行坐标

重排坐标轴

针对类别型数据的拓展：平行集

1. 适合于类别型属性
2. 坐标轴排序对结果影响较大
3. 可以与平行坐标联合使用

星形图

雷达图：弹簧模型

切尔诺夫脸谱图：用人脸特征编码不同变量的数值

Small Multiples：将高维数据的多个视图“以邮票大小”排列

基于像素图的方法——

像素图：1. 以单个像素（或近似像素大小的矩形）作为数据表达的基本单位

2. 按一定顺序排列

像素柱状图、马赛克图（类别型数据、按照指定属性顺序进行平面分割）

三、数据转化（可选）

以分析为中心（任务导向）

降维——线性、非线性变化，把高维投射在低维空间、保留重要的关系（保持信息量、保持数据区分……）

降维与投影方法——

主成分分析（PCA）：将数据映射到新的空间、新空间的基按重要程度递减排列

矩阵相乘  $\rightarrow$  新的数据（原来的数据有新的表达）

解释：1. 最大方差理论

2. 最小平方误差理论

关键步骤：

1. 对数据样本的协方差矩阵进行特征分解，得到一系列相互正交的特征向量
2. 选取其中最大的k个特征值对应的特征向量

多维尺度分析（MDS）：

基于样本之间的距离（不相似度）

关注点和点的距离，纬度没有很明显被定义

例：美国地图（MDS）

衍化：ISOMAP

测地距离  $\langle - \rangle$  欧式距离

联系：KPCA内积\PCA欧式距离\MDS\测地距离ISOMAP

t-SNE：

具有全局聚类结构的能力

从保持距离（MDS）转向保持距离的条件概率分布（t-SNE）

（小范围点的远近

复杂数据

投影中的交互：星型坐标

在投影和描点（纬度）之间建立联系

（每一个点画成描点，远近代表投影的贡献

投影寻踪：寻找一个有趣的投影

低维表达：高维数据本征结构往往是低维的

单个结构：降维PCA、ICA 流形聚类：ISOMAP、LE、LLE

多个结构：

交互：数据+纬度 双视图

四、案例

五、总结

（信息可视化转换流程图）

分析任务更清晰

指导信息更充分

强调数据预审查

数据规范化

陶煜波老师：科学计算可视化

科学计算可视化的目标是通过计算机图形学的方法，以图形、图像的形式直观地展示科学与工程中的数据。应用领域十分广泛，如医学、生物中的三维数据、气象中的多变量数据、模拟仿真等。

一、空间数据分类

网格：

规则网格

非规则网格

散乱点

应用：医学（人脑构造可视化）、生物（共聚焦显微镜 原子结构）、地质勘探、气象、模拟仿真、材料

二、标量科学计算可视化方法

1. 标量场数据

连续三维数据（设备采集、模拟仿真） -采样-> 离散标量场 -重采样-> 二维图像、三维数据（线性插值）

2. 二维标量场可视化

颜色映射：用颜色编码数据（连续或分段映射）

高度图（Hhightmap）：用高度编码数据

等值线（isocontour）：用曲线表示相同的数值，等高线、等温线

3. 三维标量场可视化

采样获得的数据（设备采集、模拟仿真） -> 体素仿真 -> 体素空间 ->像素空间

3.1 间接体绘制-将标量场转换为几何表面进行绘制

3.2 直接体绘制-标量场的可视表达（线面直接画出）

例：发射-吸收光学模型

体绘制积分-数值实现

```
for each pixel{
  //从前往后技术
  for i from 1 to n{
    ....
  }
}
```

### 3.3 数据分类-传递函数/传输函数

定义了数据值及其相关属性、颜色、不透明度等视觉元素之间的映射关系函数

不透明度：特征

### 3.4 采样点的属性（定义域）-特征空间

局部属性-标量值、梯度值、曲率、纹理、沿梯度的采样值、局部直方图

全局属性-频率、形状、遮挡

### 3.5

## 4. 传递函数设计

映射规则

分类-> 分层 -> 特征度

光学属性设计

方法：

1. 反复尝试
2. 以图像为中心
3. 以数据为中心
4. 混合图像和数据的智能传递函数设计

目的：

省时省力

以图像为中心的传递函数设计-**遗传算法**

基于随机搜索技术的传递函数交互式和自动优化流程

(.....)

以数据为中心

拓展传递函数的定义域，增强数据的区分能力

特征空间自动分析

基于反馈的传递函数设计-可见性

混合图像和数据的智能传递函数设计

具有实时反馈的所见即所得（WYSIWYG）的数据分类

可视化类库与工具：VIK、Voreen

图像处理、图形绘制、信息/科学可视化

向量法可视化：

1. 图标法：线条、箭头、方向标识符
2. 曲线/曲面可视化：流线、时间线、纹线.....
3. 基于纹理可视化

实例：数据获取 -> 数据可视化-流线

## 四、多变量数据分析方法

多变量：气压、温度.....）描述某个物理现象的多个方面，常见于模拟仿真应用

多趟数据/组集数据：

改变输入参数，生成多组运行结果

分析模拟仿真模型的不确定性

多变量数据分析：

1. 基于整体的变量间关系分析：  
相关关系：皮尔森相关系数  
-线性相关关系



## 2. 网络攻击

分析：将网络攻击时的流量与连接特征/网络活动/主机行为特征可视化出来

## 3. 攻击特征分析

## 4. 关联分析

多步攻击的关联分析与路径分析

多种网络安全事件的3W(when, what, where)关联分析

雷达图

Day. 4

## 时磊老师：网络数据可视化

### 一、网络可视化基础

(.bat文件)

新型信息网络——

社会网络、人口迁移网络（节点是人）

文本网络（节点是字）

基因网络、脑部网络

社会网络数据构成：

1. 节点（信息实体）

2. 边（关联关系：好友/评论/转发……）

图布局：

首要质量判定规则——最小化（边交叉数目）

更多的边交叉导致用户理解能力下降

图布局算法：

Tutte基于重心的布局算法

基于Planarity的图布局算法

基于力导向的图布局算法（一般的图）

力求满足图布局的首要规则：最小化交叉数目

其它定制化图布局方法：

层次图布局

正交图布局

辐射状图布局（更容易看树的结构，但边看不清楚）

### 动态网络可视化（难点）

1. 5维动态网络可视化方法

主流网络可视化方法：

中小型网络：基于力导向的图布局算法

### 二、影响力图谱可视化

对很大的图进行聚类，看流动的关系

周志光：地理空间数据可视化

#### 一、简介

具有严格地理空间意义的对象描述，包括真实的物理生存空间数据，以及在这个物理空间中人类活动所产生的人文数据。



**地理信息系统 (GIS)**：用于输入、存储、查询、分析和显示地理数据的计算机系统，是一门综合性学科，结合地理学与地图学以及遥感和计算机科学，广泛用于不同行业的领域。

二、

地图投影：

等角度：投影前后任何位置的局部切向和法向方向组成的角度保持不变。

缺点：面积改变

圆柱投影



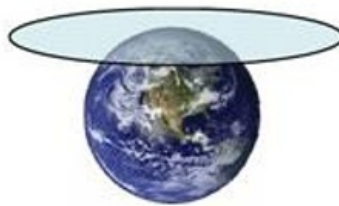
等面积：投影前后的相应图形面积大小保持不变。

圆锥投影



等距离：在标准经纬线的长度保持不变。

方位投影



地理数据集：百度地图……

地理空间可视化元素：

分类的(what/where)、定量/定序的、关系的

点数据可视化：

- ①点密度估计 (KDE)
- ②多类蓝噪点采样
- ③层次采样一致性 (层次化采样)
- ④多类散点图可视化 (颜色设计)

线数据可视化：

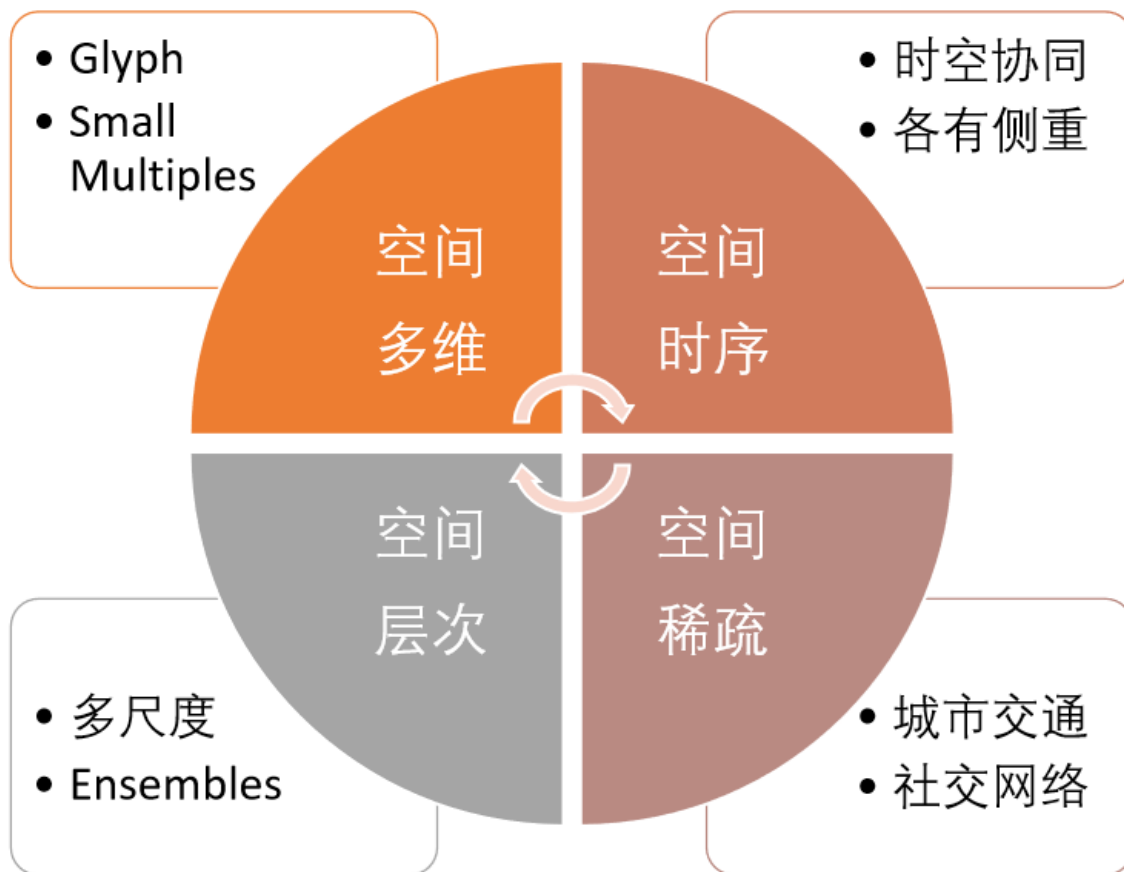
线数据：线数据通常指连接两个或更多地点的线段或者路径

线数据具有长度属性，即所经过的地理距离

三、地理空间数据组织形式

地理空间数据可视化元素：（包含四点）

# 地理空间数据可视化元素



地理空间多维属性数据

地理空间的时序数据：包含时间变量和空间位置的数据，相当于在空间数据上增加了时间维度。

- 强调时序可视化
- 时空协同可视化
- 时空动态可视化
- 突出地理空间可视化

地理空间层次数据

地理空间稀疏数据：稀疏特点—— 数据稀疏特点

采样时间

采样位置

例子：社交媒体、GPS信息

马显欣老师：数据挖掘——从数据中提取知识

任务分类：1. 描述性任务

2. 预测性任务

方法介绍：

描述性任务：对类或概念进行描述

1. 数据特征化

2. 数据区分

- 与OLAP密切相关

关联分析：

发掘频繁模式：频繁项集，频繁子序列，频繁子结构

聚类：最大化类内相似性，最小化类内相似性

例子：K均值聚类

离群点分析：

**发掘罕见对象**：与数据或模型的一般行为不一致

分类：用来描述某一个特征

- 导出能够描述和区分类别的模型
- 对类别未知的数据进行类别预测

回归：

目标输出是连续值

数据挖掘+可视化——

方法：

可视化增强的数据挖掘方法

数据挖掘方法在可视化中的应用

预测式可视分析

1. 决策树

2. 聚类：DICON

支持自主探索高维数据的聚类模式

基于图标，紧凑、可被嵌入

3. 文本分析：

主题模型Topic——一篇文章里的关键词

4. 深度神经网络：

分析神经元对给定数据的激活状态

对相似的神经元进行聚类

应用：降维

5. 预测式可视分析

面向预测性分析任务

自动分析方法采用预测模型

6. 监控视频分析

任务：识别监控视频中的人

基于主动学习策略；融合用户标记和分类器输出结果

7. 短文本分析

任务：过滤与目标事件相关的社会文本信息

结合人工筛选与自动文本分类筛选

筛选结果融合分析

**总结：**

**策略不同**

**技术互补**

**面向任务结合**