

用户数据采集与关联分析

(结课作业)



吴志祥

18205185639

1030624832@qq.com

01 文本数据分析



吴志祥

18205185639

cnwzx2012@njtech.edu.cn

第三讲 词云与可视化

1. 用任意一款词云工具，制作一个好看的词云（内容合理即可），并对词云图有一段话的解释。

The screenshot shows the Weiciyun.com website interface for creating word clouds. On the left, there's a sidebar with various tools like '内容' (Content), '形状' (Shapes), '配置' (Configuration), '插图' (Illustrations), '字体' (Fonts), '快捷键' (Shortcuts), and '帮助' (Help). The main area has tabs for '导入单词' (Import Words), '取消配置' (Cancel Configuration), '表情' (Emojis), '导出' (Export), and '保存草稿' (Save Draft). A red button says '加载词云' (Load Word Cloud). Below these are buttons for '加入我们, 去水印' (Join Us, Remove Watermark), '视频教程' (Video Tutorials), '下载到本地' (Download to Local), and a user profile icon. The central part of the screen displays a large, colorful word cloud centered around the year '1958'. The words are arranged in a dense, overlapping pattern, with larger words indicating higher frequency. Some prominent words include '制造' (Manufacture), '报纸' (Newspaper), '核潜艇' (Nuclear Submarine), '黄旭华' (Wang Xu Hua), '无私奉献' (Selfless Dedication), '试验' (Experiment), '设计' (Design), '建造' (Construction), '正直' (Integrity), '仁厚' (Kindness), and '中国梦' (Chinese Dream). The background of the word cloud features various news clippings and quotes related to Wang Xu Hua's life and work.

第三讲 词云与可视化

2. 使用Echarts，制作3个以上图，其中一个必须是“关系”，图的概念越明确（可解释，而不是自带的模板）越好。



The screenshot shows the ECharts online editor interface. On the left, there is a code editor with JS and TS tabs, displaying the following code:

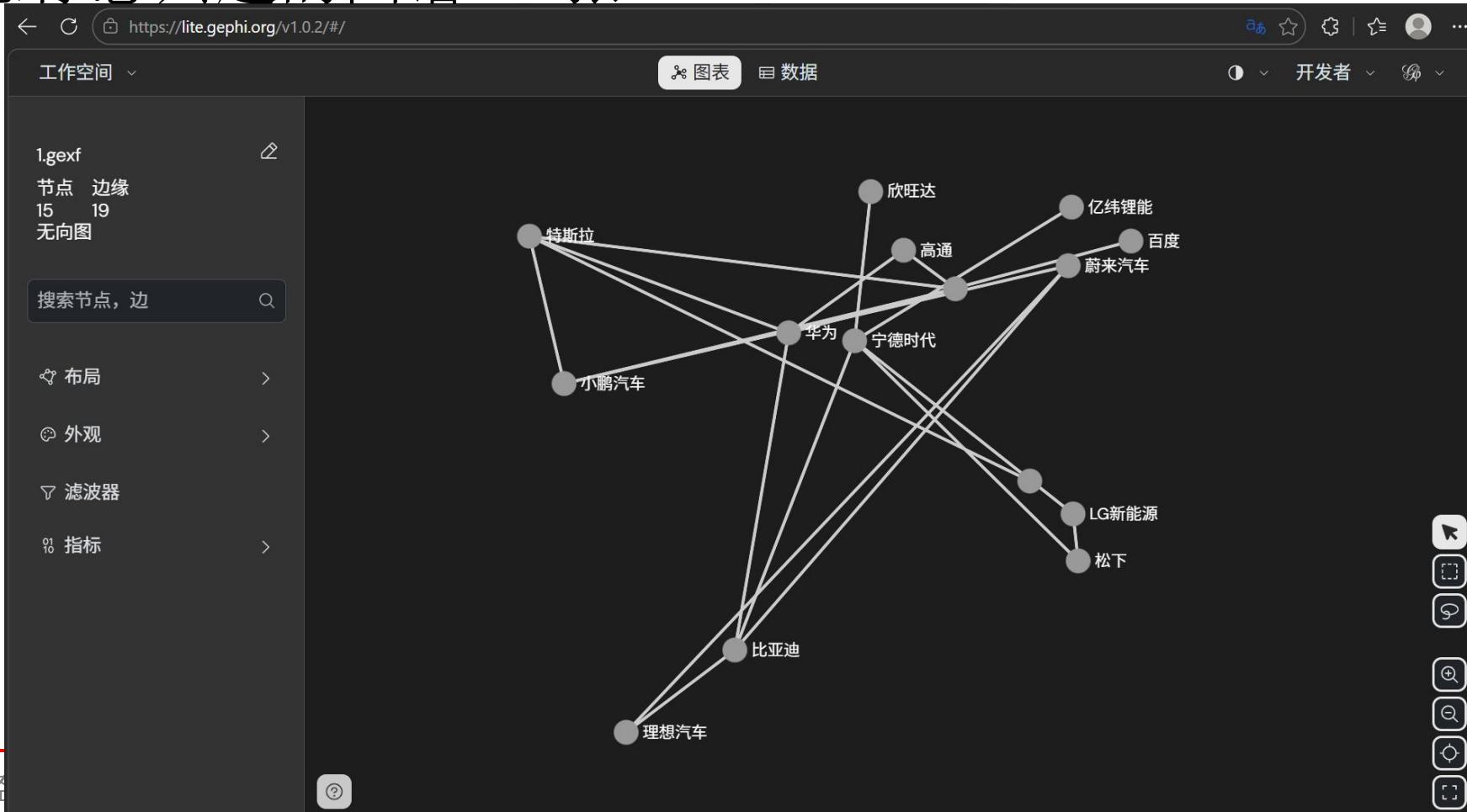
```
<!DOCTYPE html>
<html lang="zh-CN">
<head>
    <meta charset="UTF-8">
    <title>电商用户行为分析仪表盘</title>
    <!-- 1. 引入 ECharts 文件 -->
    <script src="https://cdn.jsdelivr.net/npm/echarts@5.4.3/dist/echarts.min.js"></script>
    <style>
        body {
            font-family: sans-serif;
            display: flex;
            flex-direction: column;
            align-items: center;
            background-color: #f4f4f4;
            margin: 0;
            padding: 20px;
        }
        h1 {
            color: #333;
        }
        .chart-grid {
            display: grid;
            grid-template-columns: 1fr 1fr;
            grid-template-rows: 1fr 1fr;
            gap: 20px;
            width: 95%;
            max-width: 1600px;
            height: 80vh;
        }
        .chart-item {
    
```

The main area displays a dashboard titled "电商平台用户购买行为分析仪表盘". It contains four charts:

- 用户年龄分布**: A donut chart showing the distribution of user ages. Legend: 18-24岁 (blue), 25-34岁 (green), 35-44岁 (yellow), 45-54岁 (red), 55岁以上 (light blue).
- 各年龄段消费金额 (单位: 万元)**: A bar chart showing consumption amount by age group. Data (approximate): 18-24岁 (~100), 25-34岁 (~180), 35-44岁 (~150), 45-54岁 (~120), 55岁以上 (~80).
- 不同年龄段用户购买偏好**: A radar chart showing purchase preferences across five categories: 电子广品, 服装鞋帽, 食品生鲜, 家居用品, 美妆护肤.
- 商品类目共同购买关系**: A bubble chart showing the relationship between product categories. Categories include 服装, 鞋履, 美妆-护肤品, 食品-生鲜, 家居用品, 电子广品, 美妆-护肤品.

第三讲 词云与可视化

3. 使用Gehpi、VOSViewer、CiteSpace…其中任意一款工具，绘制任意你感兴趣的图谱1-2张。



第三讲 词云

4. 采用给的文本的词云图绘制，越清晰越好 (生成的词云)

```
1.py <input>
1   from PIL import Image
2
3   # --- 1. 准备文本数据 ---
4   # 这是一段关于“人工智能”的中文文本
5   text = """
6       人工智能 (Artificial Intelligence, 简称AI) 是一门让计算机能够执行通常需要人类智能才能完成的任务的科学与工
7       它涵盖了机器学习、深度学习、自然语言处理、计算机视觉、专家系统等多个领域。
8
9       机器学习是人工智能的一个核心分支，它研究如何使计算机能够从数据中学习并改进，而无需显式编程。
10      深度学习作为机器学习的一个重要子集，通过构建深层神经网络模型，在图像识别、语音识别等领域取得了突破性进展。
11      自然语言处理致力于使计算机能够理解、解释和生成人类语言，为智能客服、机器翻译等应用提供了技术支持。
12      随着大数据和计算能力的飞速发展，人工智能技术正深刻影响着医疗、交通、金融、教育等各行各业，展现出巨大的发展潜力和
13      应用前景。
14
15      ...
16
17      # --- 2. 准备自定义形状的蒙版 (Mask) ---
18      # 读取蒙版图片
19      try:
20          mask_image = np.array(Image.open("mask.png")) # 替换成你的蒙版图片文件名
21      except FileNotFoundError:
22          print("未找到蒙版图片 'mask.png'，将生成默认矩形词云。")
23          mask_image = None # 如果找不到图片，则不使用蒙版
24
25      # --- 3. 配置并生成词云 ---
26      # 设置中文字体路径（请根据你的操作系统和字体位置进行修改）
27      # Windows 示例: 'C:/Windows/Fonts/simhei.ttf' (黑体)
28      # macOS 示例: '/System/Library/Fonts/PingFang.ttc' (苹方)
29      # Linux 示例: '/usr/share/fonts/truetype/wqy/wqy-zenhei.ttc' (文泉驿正黑)
30      font_path = 'C:/Windows/Fonts/simhei.ttf'
31
32      # 创建 WordCloud 对象
33      wc = WordCloud(
34          font_path=font_path,                      # 设置字体，解决中文显示问题
35          background_color="white",                  # 设置背景颜色为白色
36          max_words=200,                           # 设置最大显示的词数
37          mask=mask_image,                         # 设置蒙版，使词云形状与蒙版一致
38          contour_width=3,                          # 如果使用蒙版，可以设置轮廓宽度
39          contour_color='steelblue',                # 设置轮廓颜色
40          max_font_size=100,                         # 设置最大字体大小
41          random_state=42,                          # 设置随机种子，使每次生成的布局一致
42          colormap='viridis'                       # 设置颜色方案
43      )
44
45      # 从文本生成词云
46      wc.generate(text)
47
48      # --- 4. 显示和保存词云图 ---
49      # 使用 matplotlib 显示词云图
50      plt.figure(figsize=(10, 10)) # 设置图像大小
51      plt.imshow(wc, interpolation='bilinear')
52      plt.axis("off") # 关闭坐标轴
53      plt.show()
54
55      # 保存词云图到文件
56      wc.to_file("ai_wordcloud.png")
57      print("词云图已保存为 'ai_wordcloud.png'")
```

文本的词云图绘制，越清晰越好

