

# “正大杯”第十二届 全国市场调查与分析大赛公益培训

## 2.6 数据建模与分析（六）

第十二届  
(Credamo)

# 数据建模与分析（六）

## 主要内容

1

文本分析

2

图片分析



## 文本分析

第十二届市调大赛公益培训课件  
(Credamo 见数 版权所有)

# 文本分析

- **文本分析** (text analysis) 是文本挖掘、信息检索的一个基本问题，它把从文本中抽取出的特征词进行量化来表示文本信息。
- 当今我们能获取的很多数据，包括录音、报道、网页数据等都以文本形式储存，需要将文本从一个无结构的原始数据转化为结构化的计算机可以识别处理的信息
- **用途：** 文本分析可找出一段文本出现频次最多的关键词；还可以分析文本的情感；可以推断文本提供者的意图和目的等

# 文本分析

## 基本原理

- **数据采集：** 文本分析的第一步是要从收集要分析的文本开始 – 定义、筛选、获取和存储原始数据，这些数据可以包含文本文档，网页（博客，新闻等）和在线评论等等。
- **准备数据：** 获取到数据后，需要做好进行分析的准备，采用适当的形式来整理数据，以便配合机器学习模型使用，包括文本清理、分词、标注等。
- **数据分析：** 数据分析是一个分析经预处理后文本数据的过程，通过机器学习等方法，可用于分析庞大的数据集。
- **数据可视化：** 可视化是将数据转换为有深层次价值信息的过程，以图形，表格和其他直观的表达形式表示数据的过程。

# 文本分析

## 功能演示——录音转文字①

- 当使用录音题收集到录音文件，并采纳答卷后，可通过语音识别应用，将录音文件智能识别为文字，可用于分析或下载。
- 在应用列表找到语音识别后点击使用，在已有的包含录音题的问卷或新创建的问卷中建立新项目。

The screenshot displays the '数据模型' (Data Model) section of the Credamo application. At the top, there is a navigation bar with links: '个人资料', '我的标签', '账户信息', '交易记录', '问卷记录', '订单管理', '发票信息', '应用列表', and '账户升级'. Below this, the '数据模型' section features three main cards, each labeled '高级账户专享' (Advanced Account Exclusive):

- 结构方程模型** (Structural Equation Model): 使用结构方程模型完成多元数据分析 (Use structural equation model to complete multivariate data analysis). Includes '查看帮助' (View Help) and '立即使用' (Use Immediately) buttons.
- 联合分析** (Conjoint Analysis): 基于CBC联合分析精准 规划营销方案 (Based on CBC joint analysis for precise marketing planning). Includes '查看帮助' (View Help) and '立即使用' (Use Immediately) buttons.
- 语音识别** (Voice Recognition): 将录音文件智能识别为文字 (Intelligently convert audio files to text). Includes '查看帮助' (View Help) and '立即使用' (Use Immediately) buttons. The '立即使用' button is highlighted with a red box.

Below the cards, there is a '基本设置' (Basic Settings) section with a sidebar containing '项目标题' (Project Title) and '内容简介' (Content Introduction). The '项目标题' field contains '手机广告测试' (Mobile Ad Test), and the '内容简介' field contains '手机广告录音转文字功能展示' (Mobile Ad Audio-to-Text Feature Demo). Below this, there is a '问卷' (Survey) section with a list of surveys. The survey '手机广告效果研究V5.0' (Mobile Ad Effect Research V5.0) is selected, and the '选择问卷' (Select Survey) button is highlighted with a red box. To the right of the survey list is a '+ 创建新问卷' (Create New Survey) button. At the bottom, there are '确定创建' (Confirm Creation) and '取消创建' (Cancel Creation) buttons.

# 文本分析

## 功能演示——录音转文字②

- 项目创建成功后，来到总览页，在这里可以看到基本的问卷作答情况。点击右上角“查看报告”，进入报告页。
- 点击左上角“添加分析”，在弹出的配置框中，选择需要分析的录音题和目标语言，点击“确定”。（当前仅支持普通话，未来将逐步开放常用方言和外语）





# 文本分析

## 功能演示——录音转文字③

- 点击右侧“导出”可将分析结果导出到数据清理页。您可以点击左上角“数据清理”，在新窗口打开的数据清理页确认导出结果。
- 在数据清理界面，将滑栏拖至最右端可以看见录音转文字后的文本结果。

The screenshot displays the Credamo data cleaning interface. At the top, there are buttons for '+ 添加分析' and '数据清理'. Below is a table with columns: 标题, 类型, 状态, 任务数, 进度, 更新时间, and 操作. A row shows a survey titled '请简要表达你看完该手机及广告后的感受' with a status of '已完成' and a progress bar at 100%. A red box highlights the '导出' button in the '操作' column.

Below the table, there are tabs for '问卷设计', '问卷发布', '数据清理', and '建模分析'. The '数据清理' tab is active. Underneath, there are options for '数据概览', '变量设置', and '下载管理'. A table of survey results is shown, with a red box highlighting a specific row. A red arrow points from this row to a large black box containing the transcribed text: '我知道了是华为手机，而且是徕卡三摄呢，四千二百毫安安的。快充电池，然后是曲面屏平面。屏大概占百分之九十以上。' Below this text, the score '7.0' is visible, and another score '5.0' is shown next to the truncated text '我知道了是华为...'.

# 文本分析

## 功能演示——词云图①

- 当问卷的回答中出现非结构化文本时，可使用词云图对文本进行词频分析，并形成自定义的词云展示分析结果
- 在“自定义词语”框中，可以输入类似人名、地名的自定义词语，提升分词的准确率。在“停用词”框中，可以输入一些词语来避免它们出现在分词结果中。

分析方法

词云图 [修改](#)

分析名称

词云图1

分析配置

选择分词变量

<input type="checkbox"/> 作答ID	<input type="checkbox"/> 用户ID
<input type="checkbox"/> 用户名	<input type="checkbox"/> 开始时间
<input type="checkbox"/> 结束时间	<input type="checkbox"/> 作答渠道
<input type="checkbox"/> 发布ID	<input type="checkbox"/> 问卷发布名称
<input type="checkbox"/> IP	<input type="checkbox"/> 经度
<input type="checkbox"/> 纬度	<input type="checkbox"/> 省份
<input type="checkbox"/> 城市	<input type="checkbox"/> 设备类型
<input type="checkbox"/> 操作系统类型	<input type="checkbox"/> 浏览器类型
<input type="checkbox"/> 屏幕分辨率	<input type="checkbox"/> 随机元素
<input type="checkbox"/> 性别	<input type="checkbox"/> Q5
<input type="checkbox"/> Q6	<input type="checkbox"/> 信息型
<input type="checkbox"/> 情感型	<input type="checkbox"/> Q9
<input type="checkbox"/> Q10	<input type="checkbox"/> Q13
<input type="checkbox"/> Q17	<input type="checkbox"/> Q19
<input type="checkbox"/> Q23	<input type="checkbox"/> Q28
<input type="checkbox"/> Q29_1	<input type="checkbox"/> Q29_2
<input type="checkbox"/> Q33	<input checked="" type="checkbox"/> 看完手机广告后的感受

自定义词语

添加自定义词语 (如人名、地名等)，提升分词准确率。词语之间请使用空格隔开

0/500

停用词

添加停用词，停用词不会在分词结果中出现。词语之间请使用空格隔开

0/500

同时使用默认停用词  ?

默认停用词包含约4500个使用频繁但实际意义较低的中英文单词及标点符号

www.credamo.com

# 文本分析

## 功能演示——词云图②

- 在“看完手机广告后的感受”分词下，形成的词频分析如图所示，也可根据需求删除对应词语，或重置为最初词频表。
- 在词云图配置中可更改词云图的字体范围，旋转角度，旋转角度间隔以及模板样式

分词结果 ⓘ

<input type="checkbox"/>		词语	数量 ↕
<input type="checkbox"/>	1	手机	230
<input type="checkbox"/>	2	感觉	82
<input type="checkbox"/>	3	广告	73
<input type="checkbox"/>	4	摄像头	43
<input type="checkbox"/>	5	款	42
<input type="checkbox"/>	6	喜欢	36
<input type="checkbox"/>	7	功能	32

词云图配置

字体范围: 25 ~ 40

旋转角度范围: -90° ~ 90°

旋转角度间隔: - 45 +

模板: 无

取消

确定

# 文本分析

## 功能演示——词云图③

- 右图为该配置下所生成的词云图，也可继续点击“生成词云图”再次更换词云图图示，以选择满意的展现形式。

词云图配置

字体范围: 20 ~ 60

旋转角度范围: -15° ~ 15°

旋转角度间隔: - 45 +

模板: 中国

取消 确定



# 文本分析

## 功能演示——情感分析①

- 情感分析，简单而言，是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。
- 在credamo平台中，可选择填空题变量来实现情感分析。

分析方法

情感分析

[修改](#)

分析名称

情感分析1

分析配置

文本变量

看完手机广告后的感受

取消

确定

# 文本分析

## 功能演示——情感分析②

- “总体情感倾向”图展示了所有作答者针对“深海牡蛎精”的情感倾向。“情感倾向概率汇总表”中，可以查看各个文本和具体的正负向情感概率。点击表格右上角“导出变量”，可将分析生成的正、负向情感概率导出到数据清理，用于进一步分析或下载。





## 图片分析

第十二届市桥大赛公益培训课件  
(Credamo 见数 版权所有)

# 图片分析

- 图片分析 (image analysis) 侧重于研究图像的内容, 包括但不限于使用图片处理的各种技术, 它更倾向于对图片内容的分析、解释和识别。

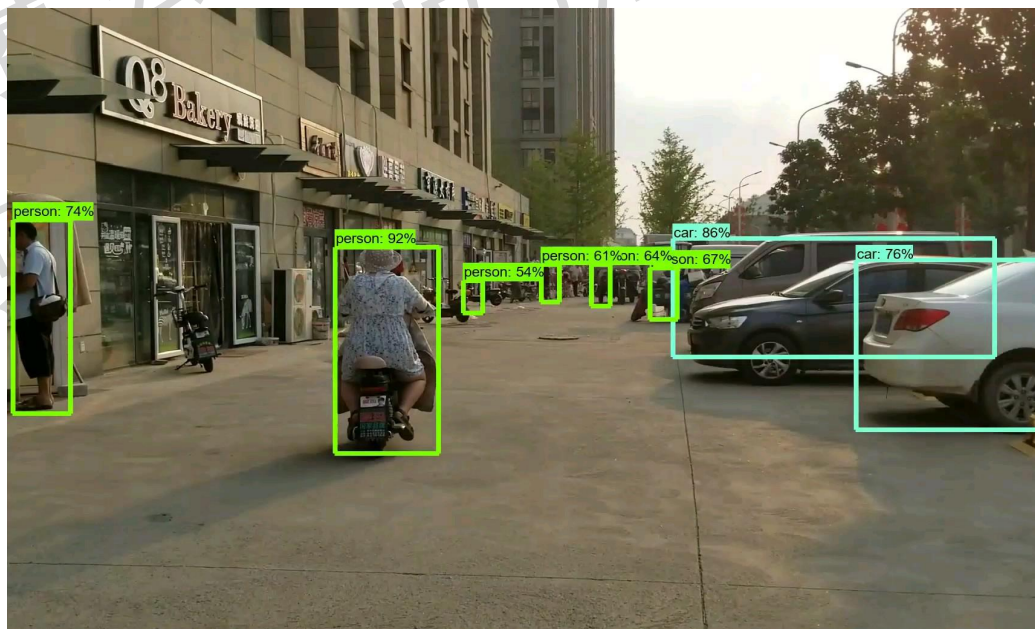
✓ 人脸识别

✓ 表情识别

✓ 物体识别

✓ 文字识别

.....





# 图片分析

## 分析步骤:

1. 图像信息的获取: 获取并存储所需要处理的图片;
2. 预处理: 图象的平滑, 变换, 增强, 恢复, 滤波等;
3. 特征抽取和选择: 在模式识别中, 需要进行特征的抽取和选择, 例如, 一幅64x64的图象可以得到4096个数据, 这种在测量空间的原始数据通过变换获得在特征空间最能反映分类本质的特征;
4. 分类器设计: 分类器设计的主要功能是通过训练确定判决规则, 使按此类判决规则分类时, 错误率最低。
5. 分类决策: 在特征空间中对被识别对象进行分类

# 图片分析

## 功能演示——图片文字分析①

- 图片文字识别功能，顾名思义，能智能识别图片中所含有的文本。
- 在应用列表中进入图片文字识别后，创建该项目。

The screenshot displays the '数据模型' (Data Model) section of the Credamo application. It features a navigation bar at the top with options like '个人资料', '我的标签', '账户信息', '交易记录', '问卷记录', '订单管理', '发票信息', '应用列表', and '账户升级'. Below the navigation bar, there are four main application cards, each labeled '高级账户专享' (Advanced Account Exclusive). The cards are: '结构方程模型' (Structural Equation Model), '联合分析' (Conjoint Analysis), '语音识别' (Voice Recognition), and '图片文字识别' (Image Text Recognition). The '图片文字识别' card is highlighted with a red border. Below these cards, there is a '基本设置' (Basic Settings) section with three tabs: '基本设置', '项目标题', and '内容简介'. The '项目标题' tab is active, showing a text input field with '手机广告' (Mobile Advertising) and a larger text area with '图片转文字功能测试' (Image to Text Function Test). Below the text area, there is a '问卷' (Survey) section with a dropdown menu showing '手机广告效果研究V5.0' and a '选择问卷' (Select Survey) button. At the bottom, there are two buttons: '确定创建' (Confirm Creation) and '取消创建' (Cancel Creation).

# 图片分析

## 功能演示——图片文字分析②

- 点击“添加分析”，选择需要分析的文件上传题或签名题。目前支持的格式包括：jpg、jpeg、png、bmg。
- 成功添加分析后，状态栏将展示预计耗时，进度条将展示当前分析进度。分析全部完成后，点击右侧“导出”可将分析结果导出到数据清理页。可以点击左上角“数据清理”，在新窗口打开的数据清理页确认导出结果。

The screenshot shows the 'Image Text Analysis' configuration window and the results table. The configuration window has a title '分析配置' and a close button 'X'. It contains a text input field with the placeholder '请上传你准备好的小物件的图片，尽量保持背景清晰整洁 (示 ?)' and two buttons: '取消' (Cancel) and '确定' (Confirm). Below the configuration window is a table with columns: '标题' (Title), '类型' (Type), '状态' (Status), '任务数' (Task Count), '进度' (Progress), '更新时间' (Update Time), and '操作' (Action). The table contains one row with the following data: Title: '请上传你准备好的小物件的图片，尽量保持背景清晰整洁 (示例如下图)', Type: '通用', Status: '已完成', Task Count: '199', Progress: '100%', Update Time: '2021-10-07 21:02:35', and Action: '导出' (highlighted with a red box), '更新', '删除'.

标题	类型	状态	任务数	进度	更新时间	操作
请上传你准备好的小物件的图片，尽量保持背景清晰整洁 (示例如下图)	通用	已完成	199	100%	2021-10-07 21:02:35	导出 更新 删除

# 图片分析

## 功能演示——图片文字分析③

- 解析效果如图所示。



请上传你准备好...	请上传你准备好...
0.712603	●● SHOT ON MI10 5G
0.879875	●● SHOT ON MI...

# 图片分析

## 功能演示——图片物体分析①

- 图片物体识别功能，顾名思义，能智能识别图片中所含有的物体。
- 在应用列表中进行图片物体识别后，创建该项目。前两步操作与图片文字识别一致

个人资料 我的标签 账户信息 交易记录 问卷记录 订单管理 发票信息 应用列表 账户升级

### 数据模型

高级账户专享

**结构方程模型**  
使用结构方程模型完成多元数据分析

查看帮助 立即使用

高级账户专享

**联合分析**  
基于CBC联合分析精准规划营销方案

查看帮助 立即使用

高级账户专享

**语音识别**  
将录音文件智能识别为文字

查看帮助 立即使用

高级账户专享

**图片文字识别**  
智能识别图片中的文字

查看帮助 立即使用

高级账户专享

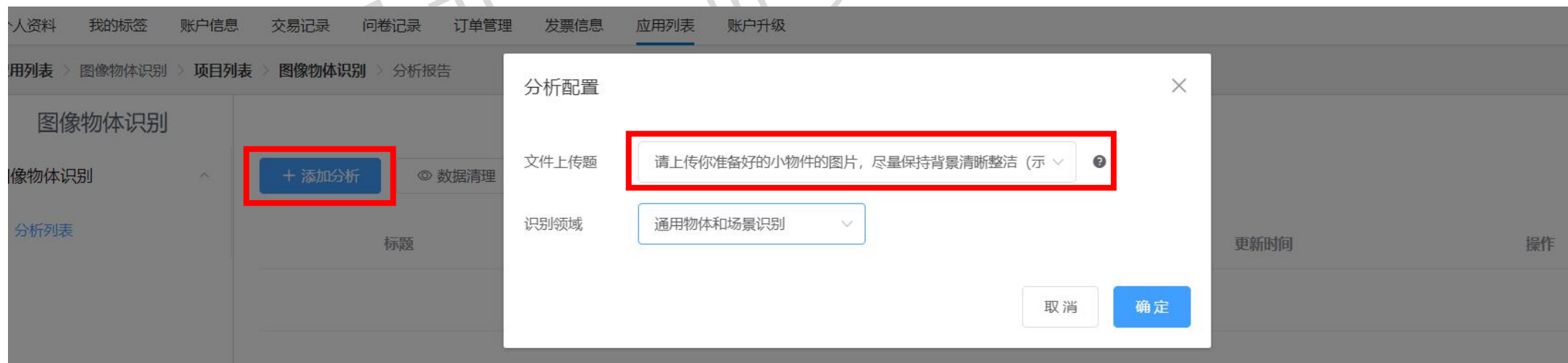
**图像物体识别**  
智能识别图像中的物体或场景，支持品牌logo、车型等多种细分领域

查看帮助 立即使用

# 图片分析

## 功能演示——图片物体分析②

- 点击“添加分析”，选择需要分析的文件上传题和识别领域
- 目前支持的格式包括：jpg、jpeg、png、bmp。
- 目前支持的识别领域包括：通用物体和场景识别（识别动物、植物、商品、建筑、风景、动漫、食材、公众人物等10万个常见物体及场景，给出图中物体的名称）；logo识别（识别超过2万类商品logo，给出图片中品牌logo的名称）；车型识别（识别3000款常见车型（小汽车为主），给出车辆品牌型号（如宝马X3））。



# 图片分析

## 功能演示——图片物体分析③

- 可以点击左上角“数据清理”，在新窗口打开的数据清理页确认导出的分析结构。分析结果包括识别结果和置信度（0~1之间的数值，数值越大表示系统对识别结果越有把握）。



The screenshot shows a data management interface with a table of tasks. The '导出' (Export) button is highlighted with a red box.

标题	类型	状态	任务数	进度	更新时间	操作
请上传你准备好的小物件的图片，尽量保持背景清晰整洁（示例如下图）	通用	已完成	199	100%	2021-10-07 20:47:31	<a href="#">导出</a> <a href="#">更新</a> <a href="#">删除</a>

- 待识别图像： 识别结果（以通用物体和场景识别为例）：



请上传你准备好...	请上传你准备好...
透明胶袋	0.665042
蛋糕	0.732574
杯子	0.730797
瓶子	0.324909

# 图片分析

## 功能演示—— 热力图分析


▼ 图片标记题

Q29 请在图片上标记2处你觉得最吸引你的地方

更换图片 删除图片

新品上市 卓越不凡

6400万像素  
5000毫安大电量  
骁龙870处理器



验证类型

固定个数

2

个

数量范围

标记方式

点选标记

框选标记

标记评论

否  是

是否必答

否  是

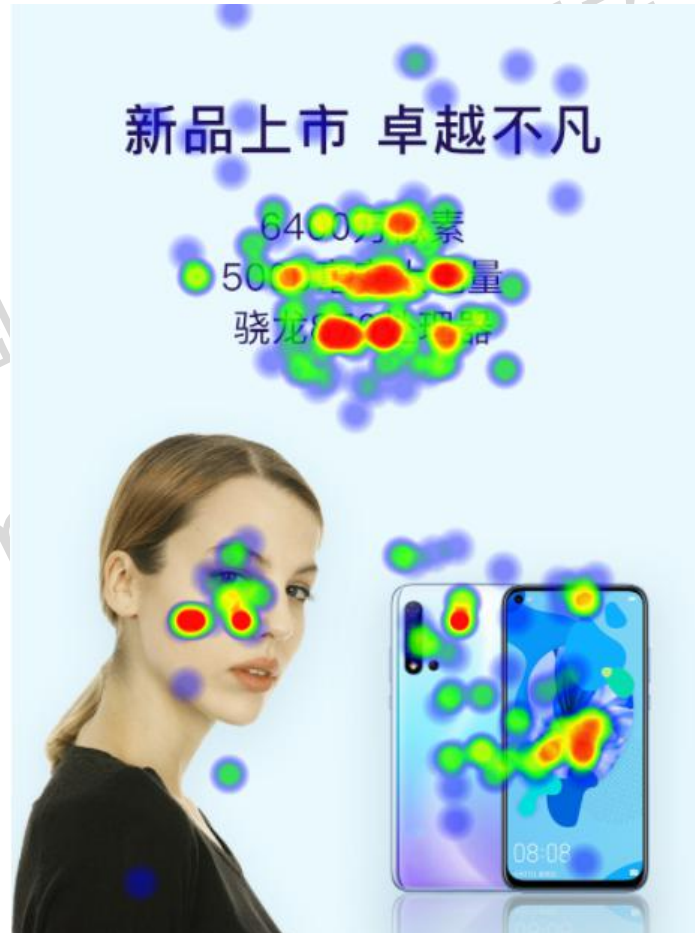
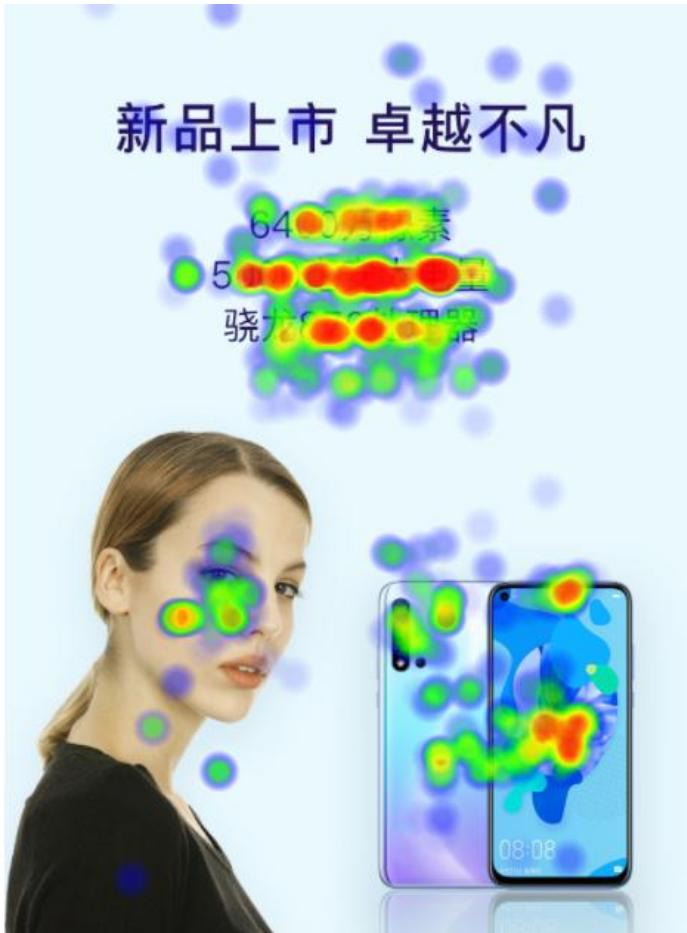


# 图片分析

## 功能演示——热力图分析

男生

女生



# 谢谢大家

关注右边公众号



及时了解大赛资讯和进程

随时学习大赛公益培训

## 中国商业统计学会

官方网址：<http://www.china-cssc.org>

公众号：



## Credamo见数

官方网址：[www.credamo.com](http://www.credamo.com)

公众号：

