“鹰眼”电商服务问题自动发现平台

操

作

手

册

目录

[**1.** **平台介绍** 1](#_Toc511833761)

[**2.** **数据收集与存储** 2](#_Toc511833762)

[**3.** **特征词提取模块** 2](#_Toc511833763)

[**4.** **文本分类模块** 2](#_Toc511833764)

[**5.** **一级属性发现模块** 2](#_Toc511833765)

[**6.** **二级属性提取模块** 2](#_Toc511833766)

[**7.** **统计与分析模块** 3](#_Toc511833767)

[**8.** **功能展示** 3](#_Toc511833768)

[**8.1.** **登录页** 3](#_Toc511833769)

[**8.2.** **主页** 4](#_Toc511833770)

[**8.3.** **评论标注** 4](#_Toc511833771)

[**8.4.** **质量分析** 5](#_Toc511833772)

1. **平台介绍**

“鹰眼”电商服务问题自动发现平台是一种基于文本情感倾向的电商服务问题自动发现系统，平台包括：特征词提取模块、文本分类模块、一级属性发现模块、二级属性提取模块和统计与分析模块。特征词提取模块负责完成对文本中出现的有意义词的提取。文本分类模块负责对有关电商服务问题的文本进行的初步分类，其分类结果分为物流、产品、售后、营销四类。一级属性发现模块完成基于词性的电商服务问题一级属性关键词的确定，为二级属性的自动发现提供数据基础。二级属性提取模块实现电商服务问题二级属性的自动发现；建立二级属性备选集合；基于情绪词典对二级属性统计与排序，统计在文本中，电商服务问题的分布和消费者的情绪特点，为商家或管理者提供文本概况或画像。统计与分析模块负责对文本分类模块、一级属性发现模块和二级属性提取模块的结果进行汇总分析。本平台适用于对有关电商服务问题的文本进行分类统计、语义分析，并基于文本的情感倾向自动发现电商服务中可能存在的缺陷。系统结构如图1所示。

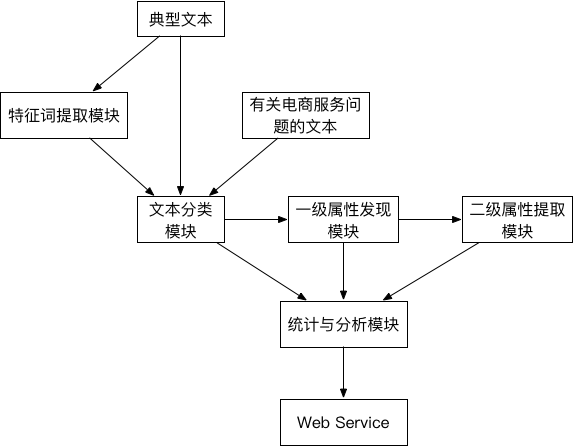


图1系统架构图

1. **数据收集与存储**

通过爬虫获取京东商城差评数据，数据以文件的形式存储。

1. **特征词提取模块**

特征词提取模块负责对文本中出现的有意义词的提取；对已经标记的典型文本进行分词并去除停用词之后，根据词语出现频次排序，得到特征词。

1. **文本分类模块**

文本分类模块，对单条文本的初步分类。具体实现过程如下：

* 1. 手工标记部分典型的文本，作为该文本的类别标注；
  2. 将特征词提取模块得到的特征词，作为特征集；
  3. 从爬取的数据中去掉重复和无效的数据，作为训练数据；
  4. 利用特征集将已标注的数据、训练数据转为向量，并以此训练逻辑回归分类器；
  5. 使用所训练的逻辑回归分类器对文本进行分类。

1. **一级属性发现模块**

一级属性发现模块实现的是基于词性的问题一级属性关键词确定，具体实现方法如下:  
 先对不同类别（物流、产品、售后、营销）相关文本进行分词处理，过滤停用词，筛选其中词性为名词的词语，统计频次并按照降序排序，作为关键词集合，由此得到不同类别商品问题的一级属性及其对应关键词；通常需要辅以管理者等人工经验，对待筛选集合中的词汇进行筛选和组合，以提高所得关键词的可用性。

1. **二级属性提取模块**

二级属性提取模块实现了电商服务问题二级属性的自动发现，首先建立二级属性备选集合，然后基于情绪词典对二级属性统计与排序，统计在文本中。

建立二级属性备选集合时将文本分类模块获得的不同类别（物流、产品、售后、营销）相关文本下属的一级属性关键词作为二级属性备选集合建立模块的输入，使用分词处理和滑动窗口方法建立二级属性备选集合。

基于情绪词典对二级属性统计与排序时，根据所使用文本的特点进行不同处理，文本长度较长、语义密度大，采用文本情感倾向度统计方法；文本长度短，语义密度低，则使用情绪词典区分情感词统计方法。最后输出二级属性集合。

1. **统计与分析模块**

统计与分析模块，利用python对经过文本分类模块的有关电商服务问题的文本以及一级属性发现模块和二级属性提取模块处理的结果进行统计与处理，生成绘图引擎能够直接使用的数据格式。

1. **功能展示**
   1. **登录页**

如图2所示， 为保护网站的数据，网站设置了一定的用户权限，因此首先进行的时登录操作。登录界面十分简洁。

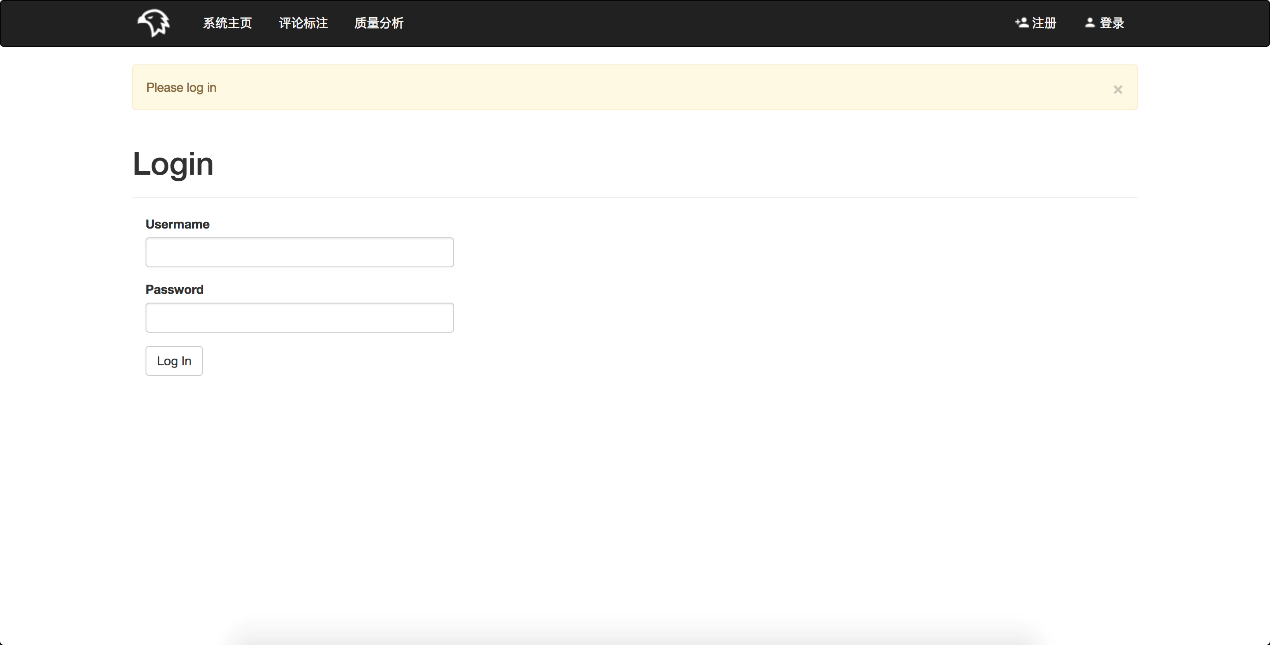


图2“鹰眼”电商服务问题自动发现平台登录页

* 1. **主页**

如图3所示，主页用来提示是否存在未标注的评论和简单的日期提示信息。

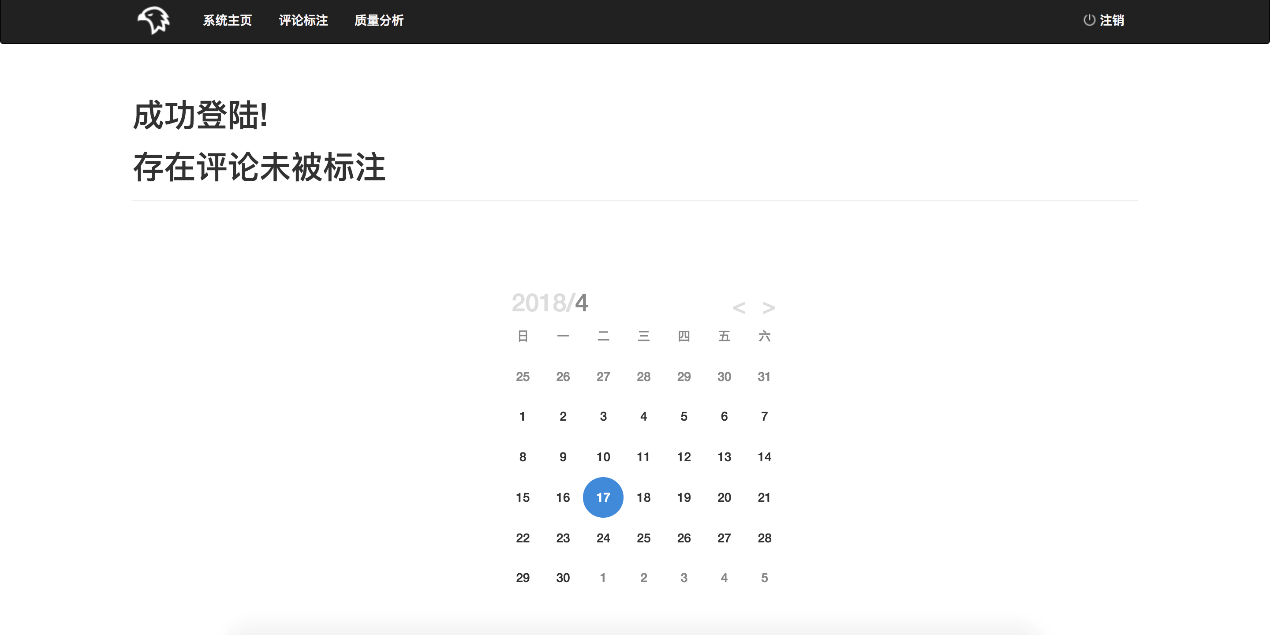


图3“鹰眼”电商服务问题自动发现平台主页

* 1. **评论标注**

如图4所示，该页用来标注用于训练文本分类模块使用的训练集数据。

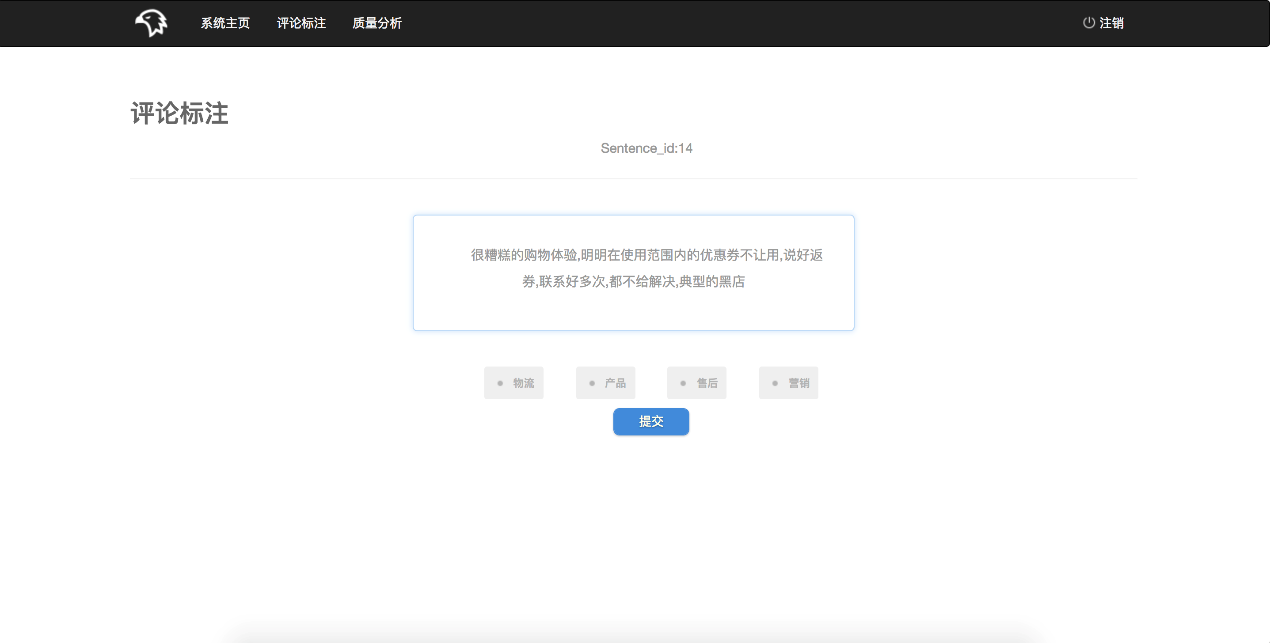


图4评论标注

* 1. **质量分析**

质量分析页面，包含了若干的网页，以对电商服务从行业、品牌到产品的问题进行发现和分析。横向导航栏标出了不同的行业的质量分析。点击图例可以用来筛选图像上出现的数据；将鼠标指针置于图的区域中会以提示栏的形式显示具体的数据信息。详细内容包括：

1）从横向的导航栏点击某个行业（以手机行业为例）将出现某个行业的总 体差评因素分布图，，如图5所示。

2）从右侧浮动导航栏点击总体差评时间分布将显示手机行业每月用户差评的数量堆叠折线图，以颜色区别不同类型的用户。如图6所示。图像标题上方的选项框可以选择按日期或者按小时显示差评的时间分布，如图7，图8所示。

3) 从右侧浮动导航栏点击差评用户分布将显示给予差评的用户分布图，同时出现了左侧的浮动导航栏，如图9所示。从左侧的浮动导航栏可分别点击各原因用户占比、各用户差评占比、用户差评热力图、用户主观倾向、用户主观倾向箱线图查看单原因用户分布图（如图10所示）、各类型用户差评的占比情况图（如图11所示）、用户类型与差评原因热力图（如图12所示）、用户主观倾向图（如图13所示）和用户主观箱线图（如图14所示）

4）从右侧浮动导航栏点击问题分析将显示产品原因树状图，从左侧的浮动导航栏可以选择一级属性按名词或形容词处理的结果，如图15、图16所示。

5）从右侧浮动导航栏点击各品牌差评占比将显示行业总体个品牌差评占比图，如图17所示。点击占比最多的前四个品牌的对应扇形区域将显示对应品牌的差评因素占比图，如图18所示。同时，右侧浮动导航栏变为对应品牌分析。从右侧浮动导航栏点击各型号占比可查看对应品牌下差评数量较多的产品的差评因素累计占比图，如图19所示，以及差评数量统计图，如图20所示。从右侧浮动导航栏点击差评的时间分布将显示对应品牌的差评时间分布图，如图21所示，标题上方选项框中选择典型产品的差评时间分布图，如图22所示。从右侧浮动导航栏点击差评的用户分布，将显示对应品牌差评的用户分布图，如图23所示。



图5行业总体差评因素分布图

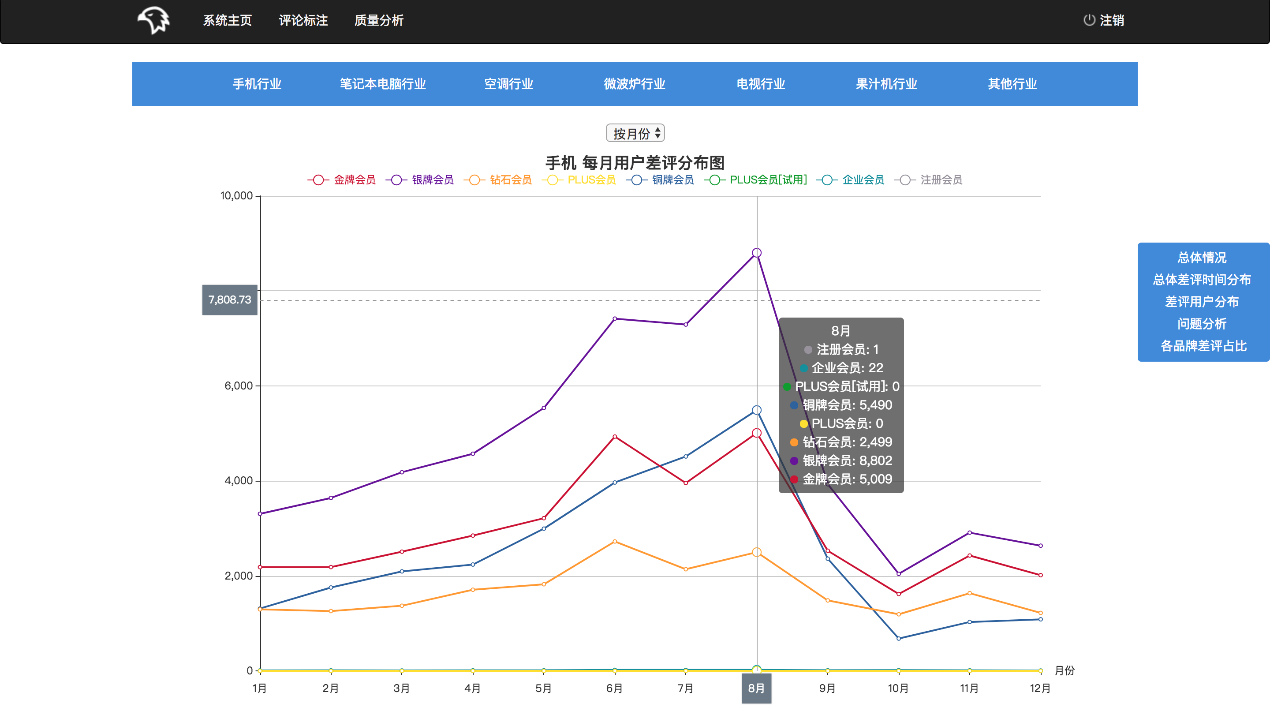


图6行业总体差评时间分布\_按月份

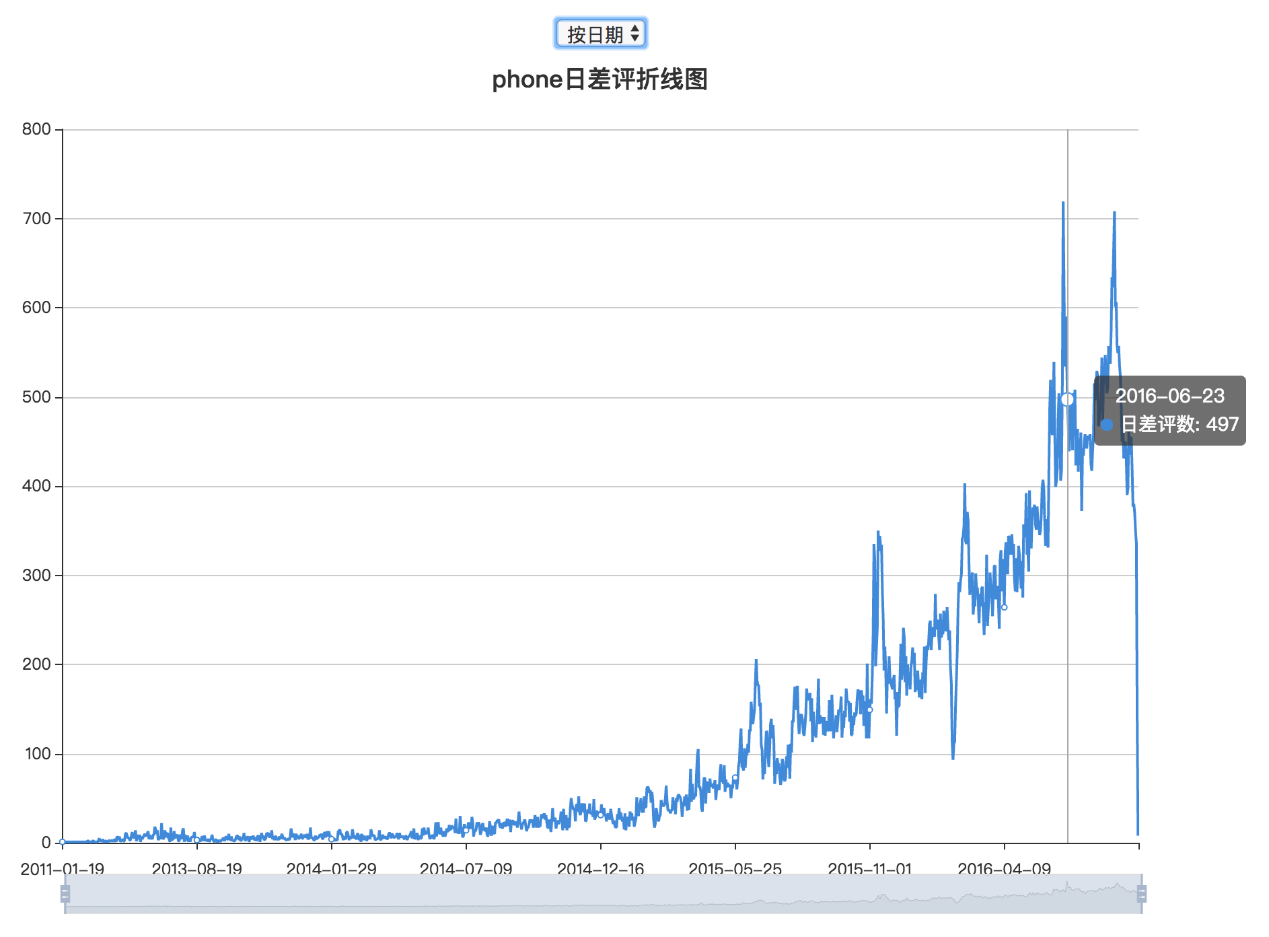


图7行业总体差评时间分布\_按日期

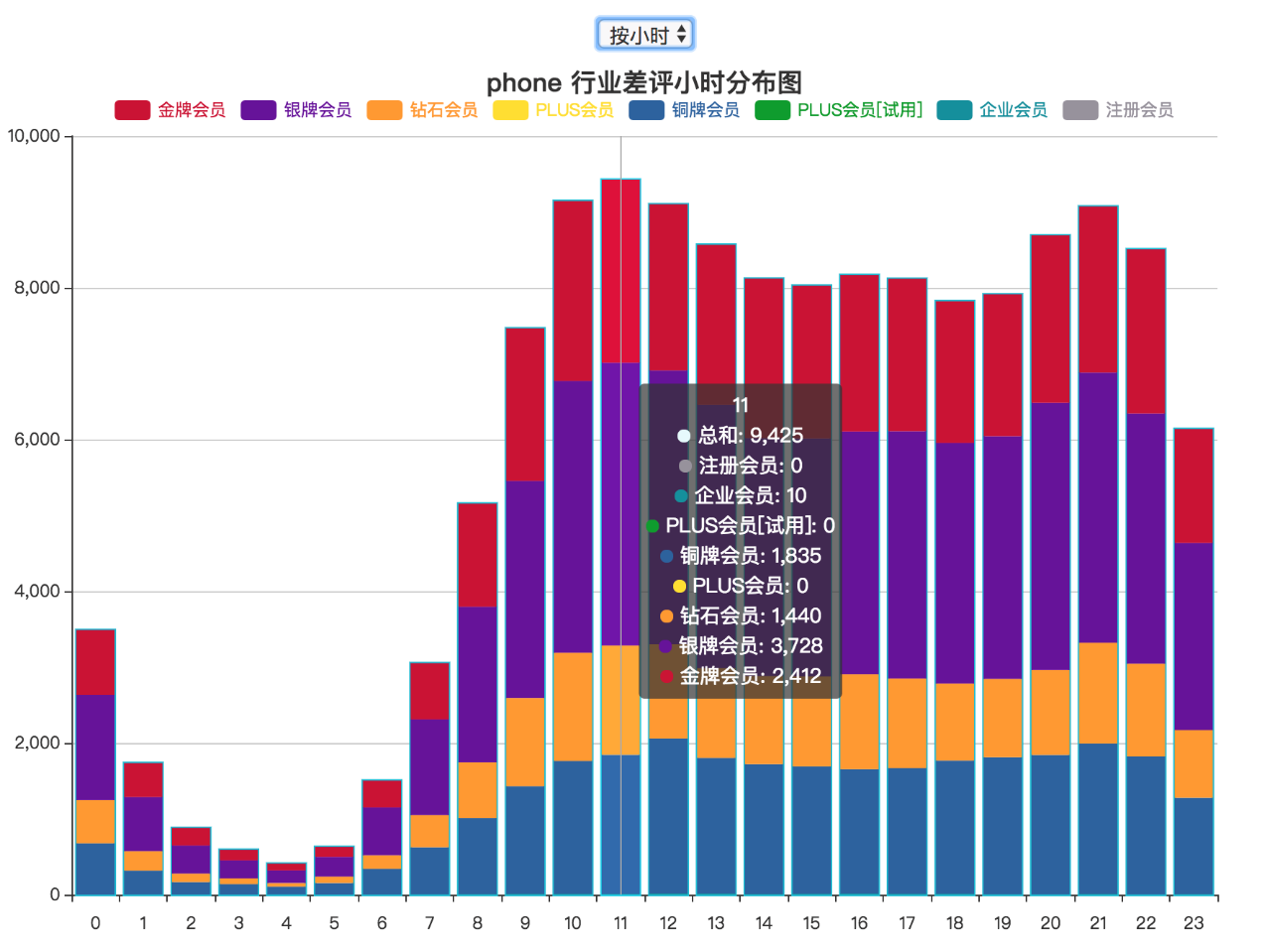


图8行业总体差评时间分布\_按小时



图9差评用户分布

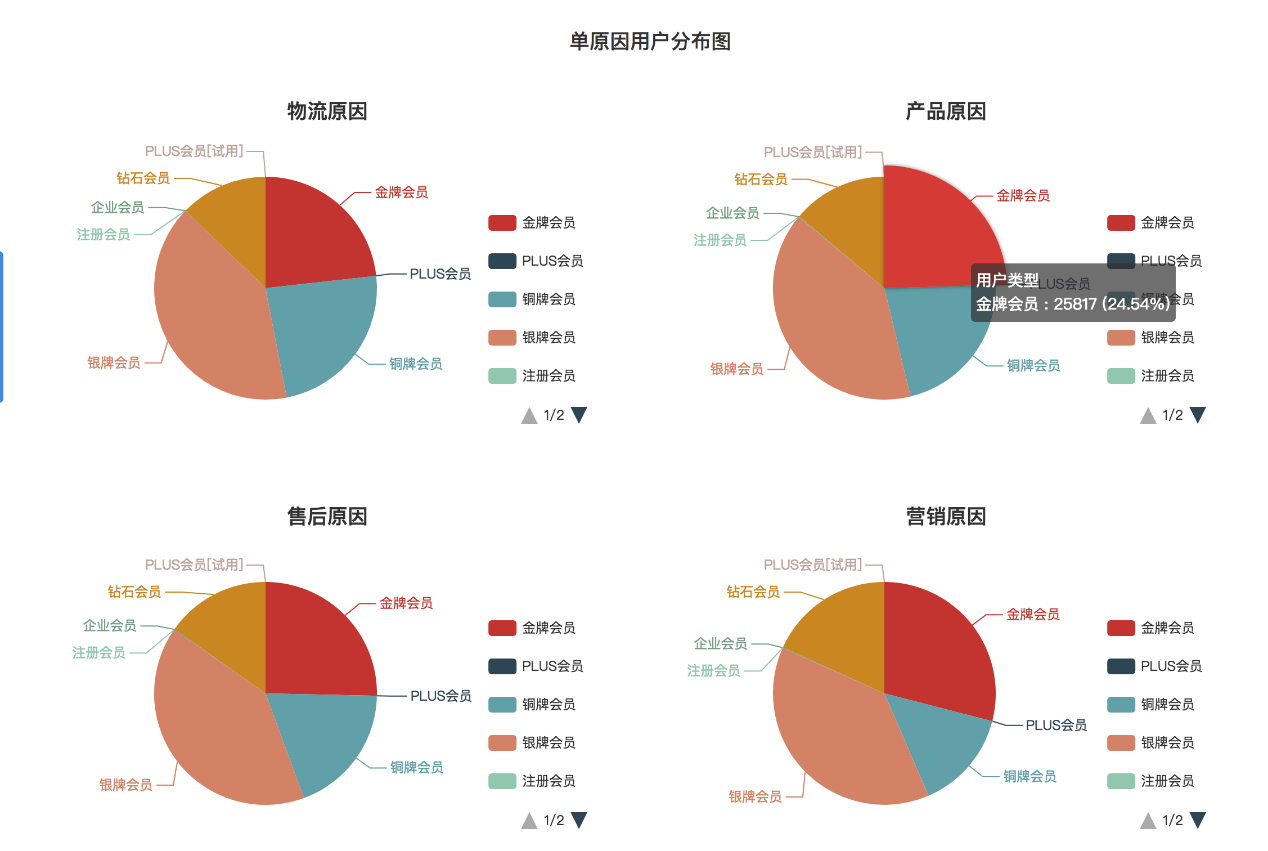


图10单原因用户分布图

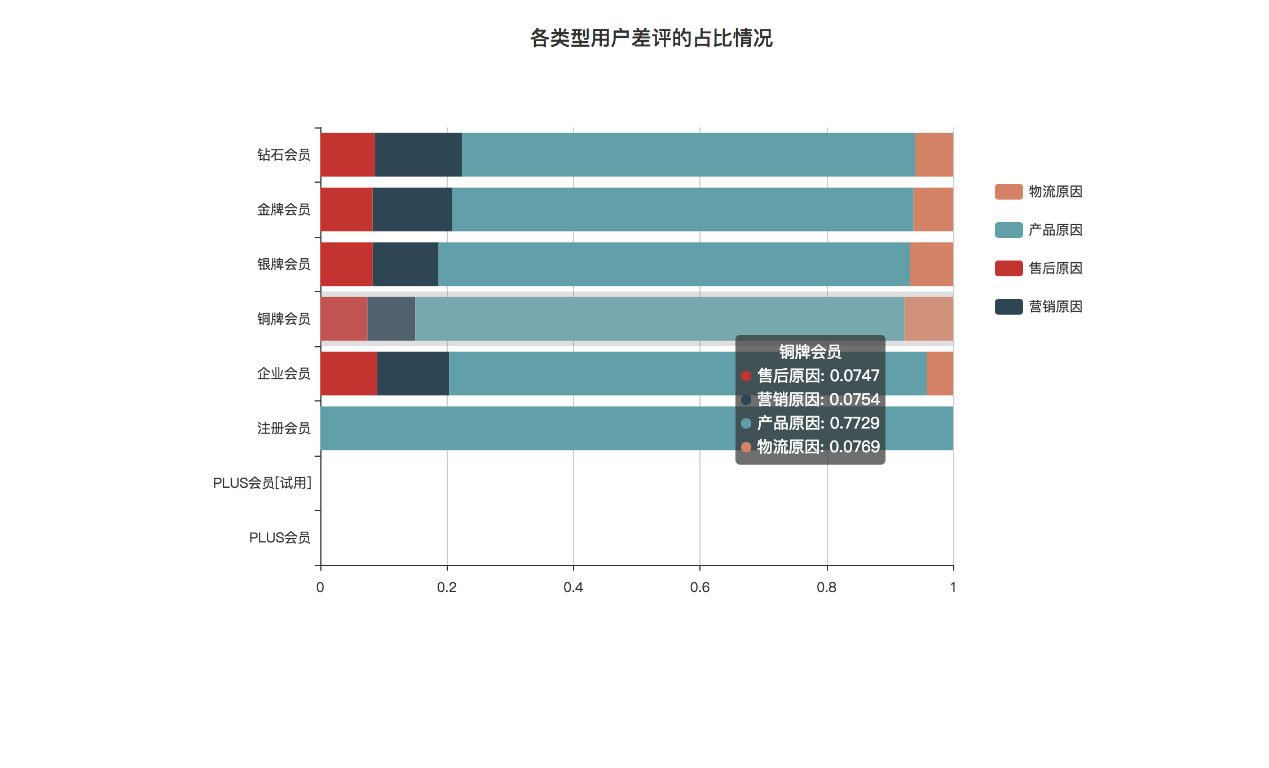


图11各类型用户差评的占比情况图

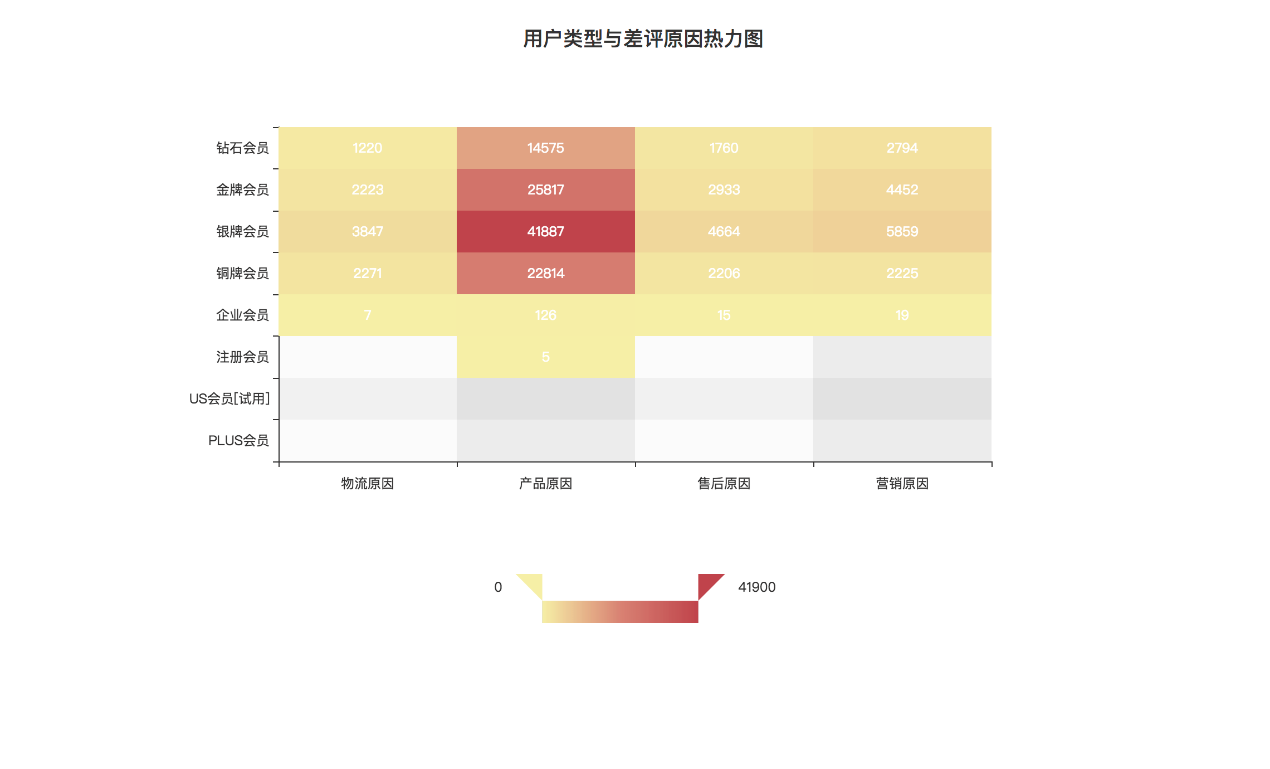


图12用户类型与差评原因热力图

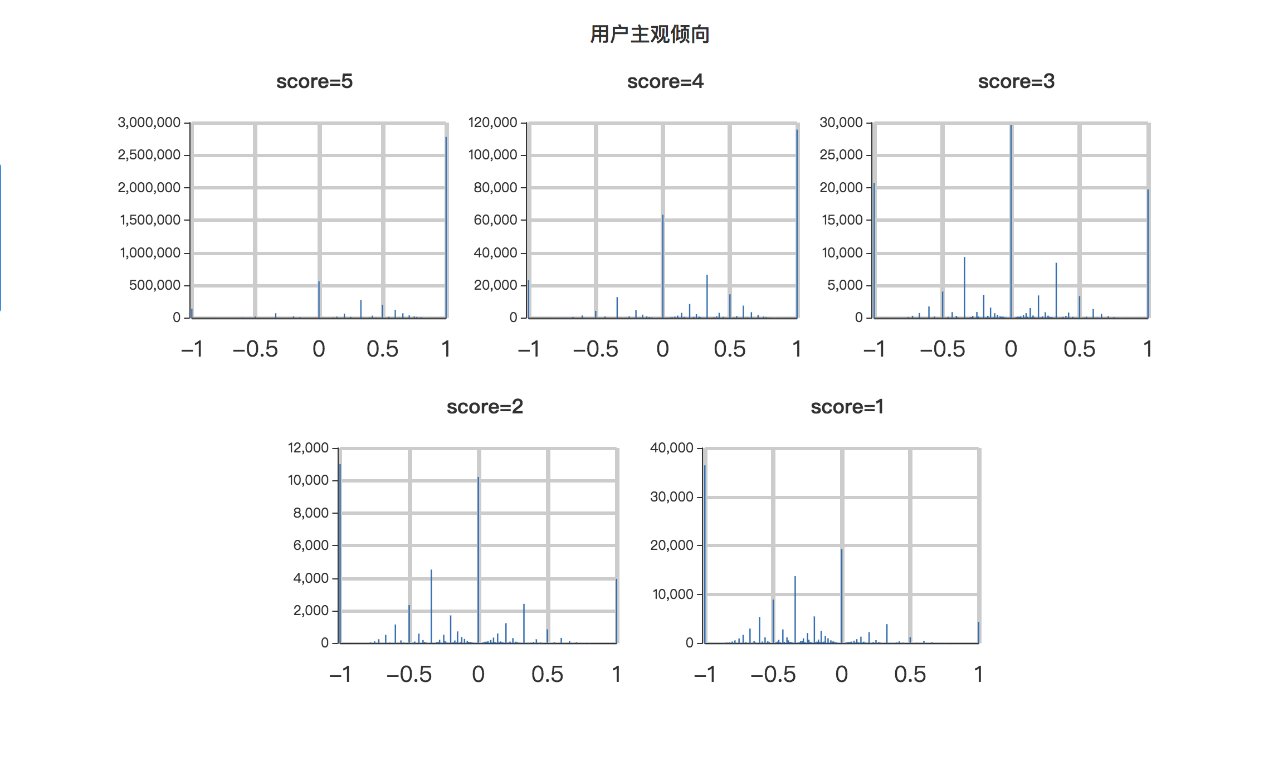


图13用户主观倾向图

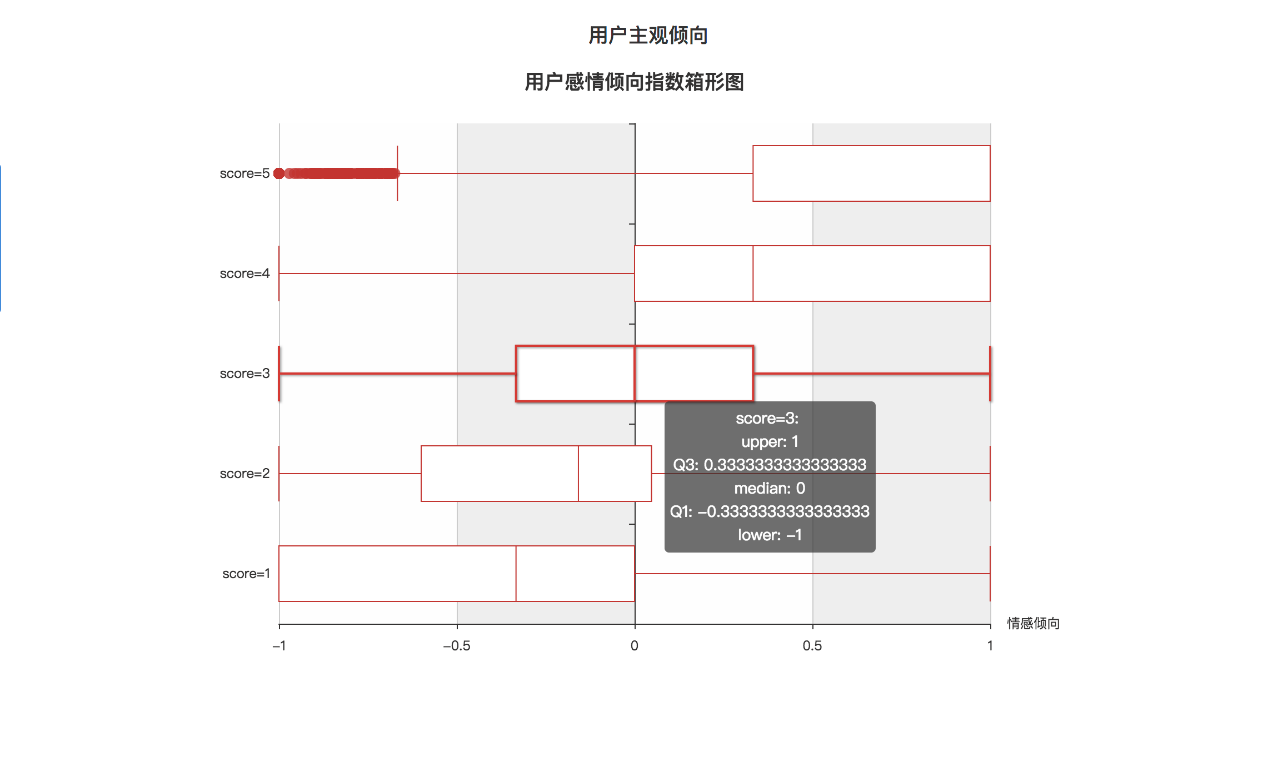


图14用户主观箱线图

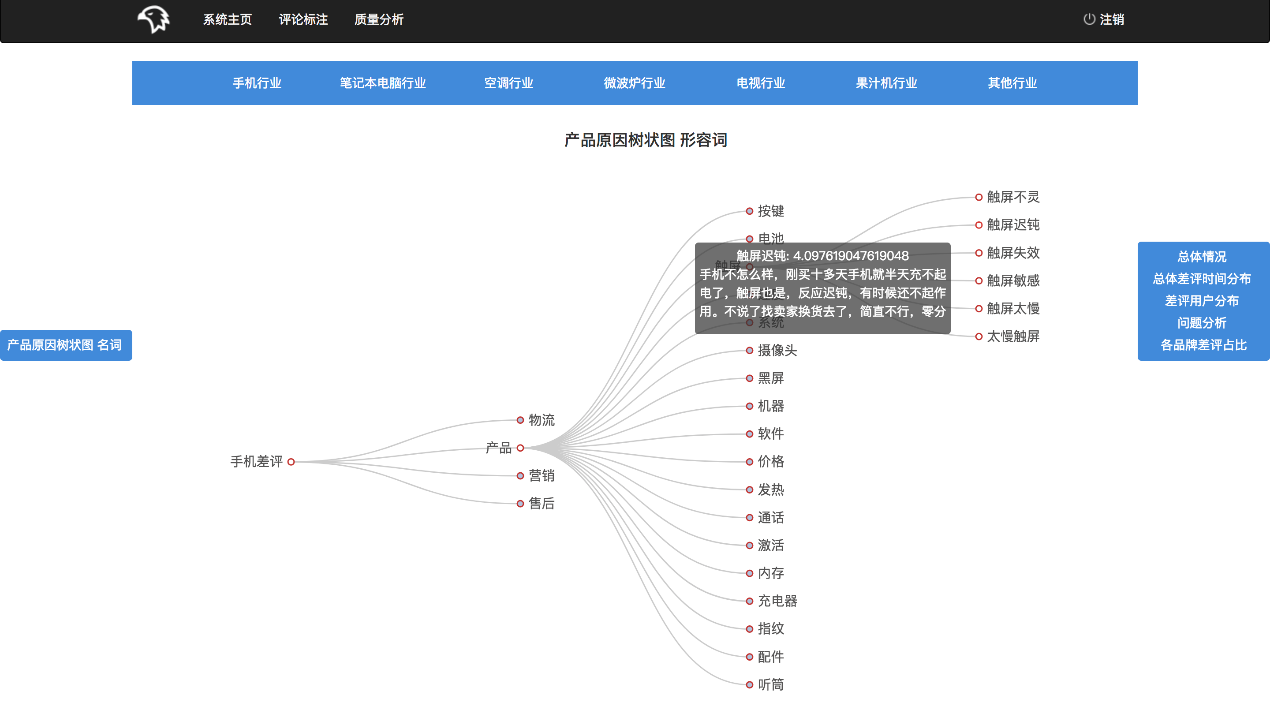


图15产品原因树状图\_形容词

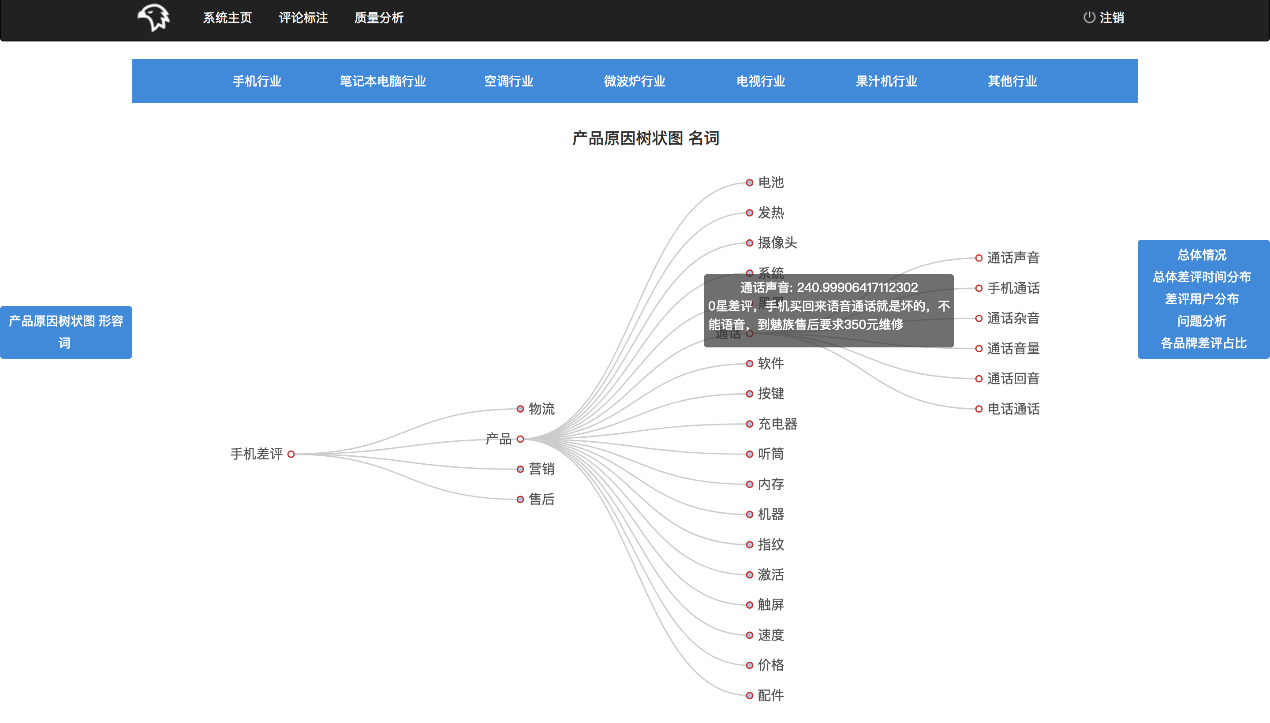


图16产品原因树状图\_名词

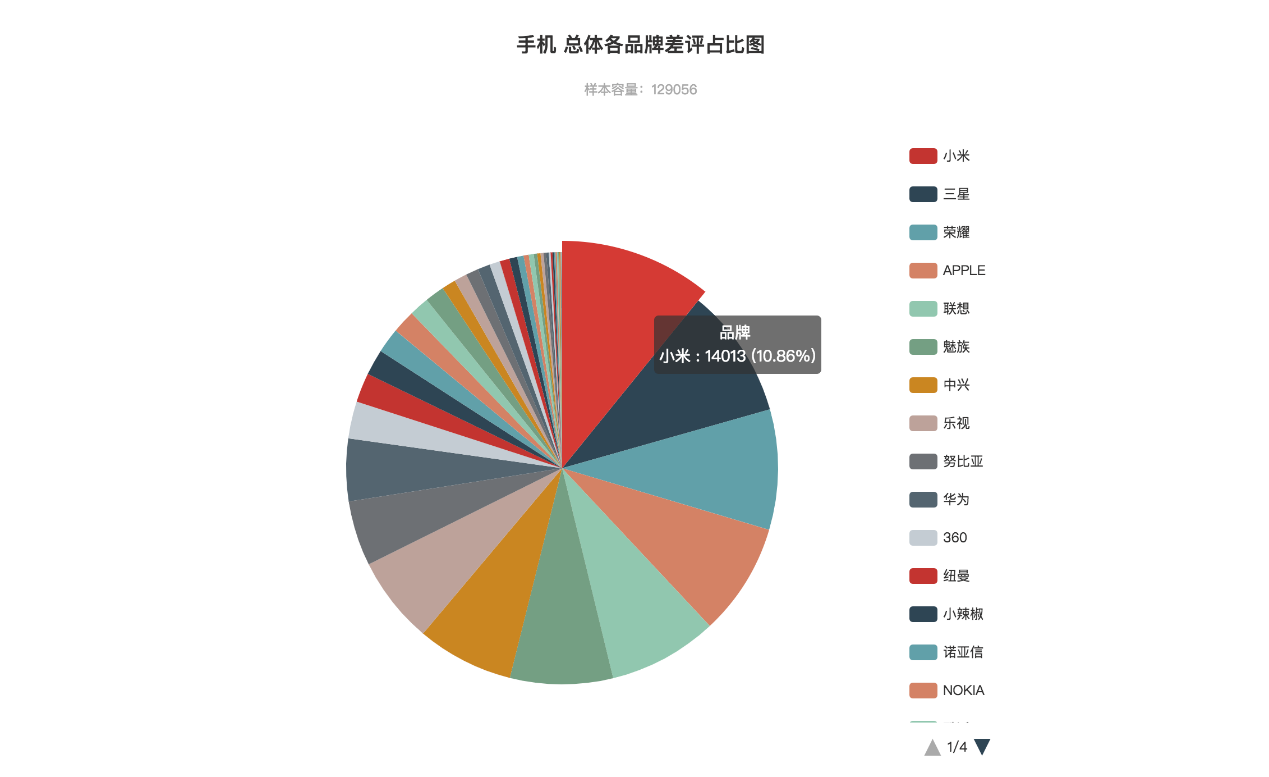


图17行业总体个品牌差评占比图



图18对应品牌的差评因素占比图

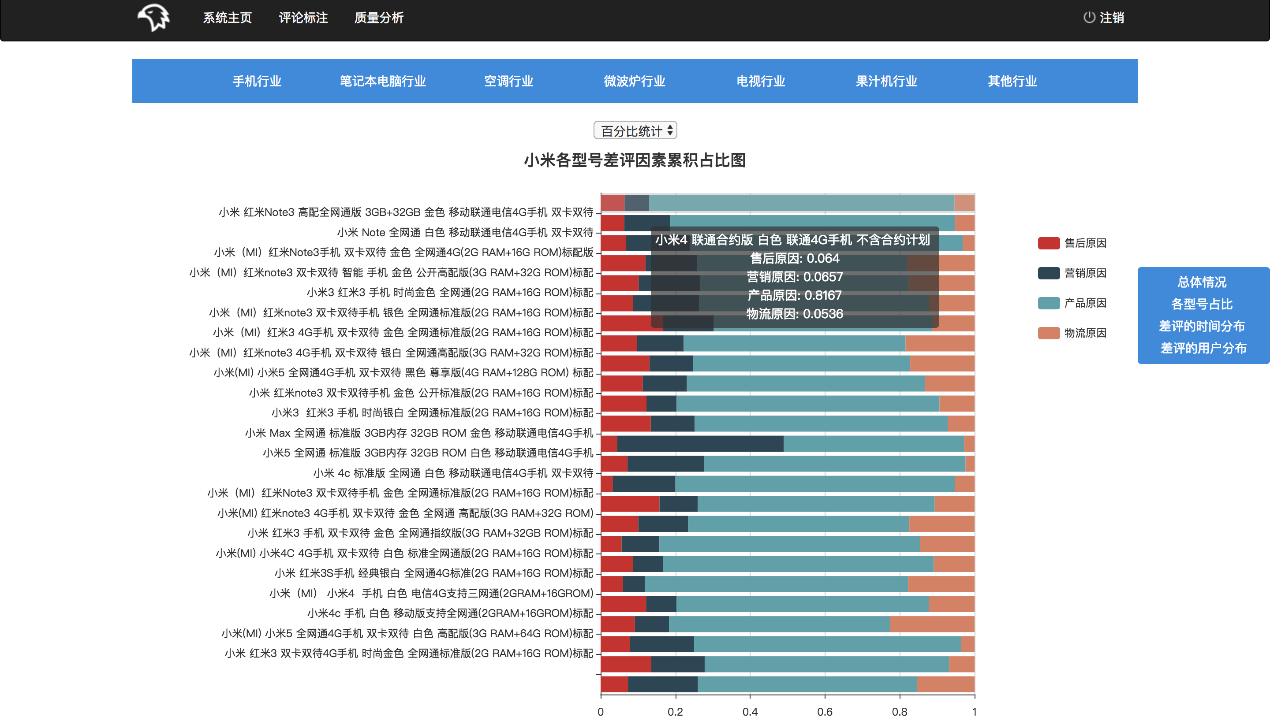


图19对应品牌各产品差评因素累计占比图

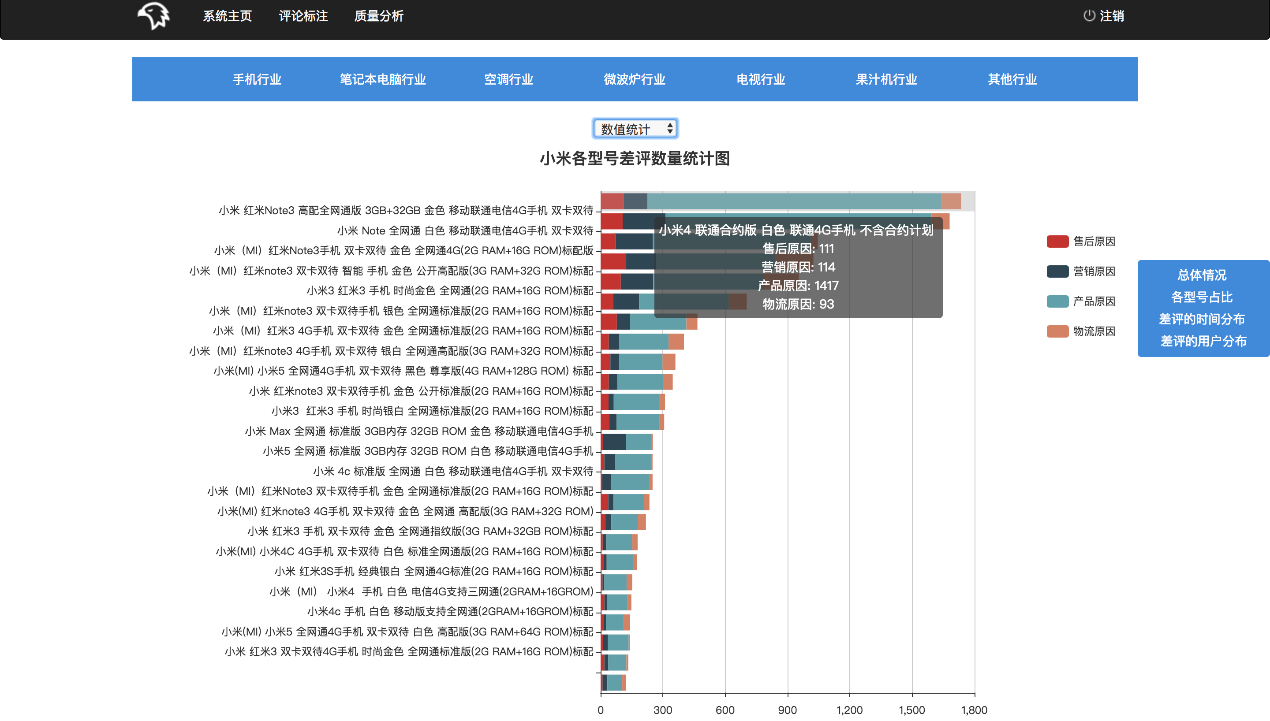


图20对应品牌各产品差评数量统计图

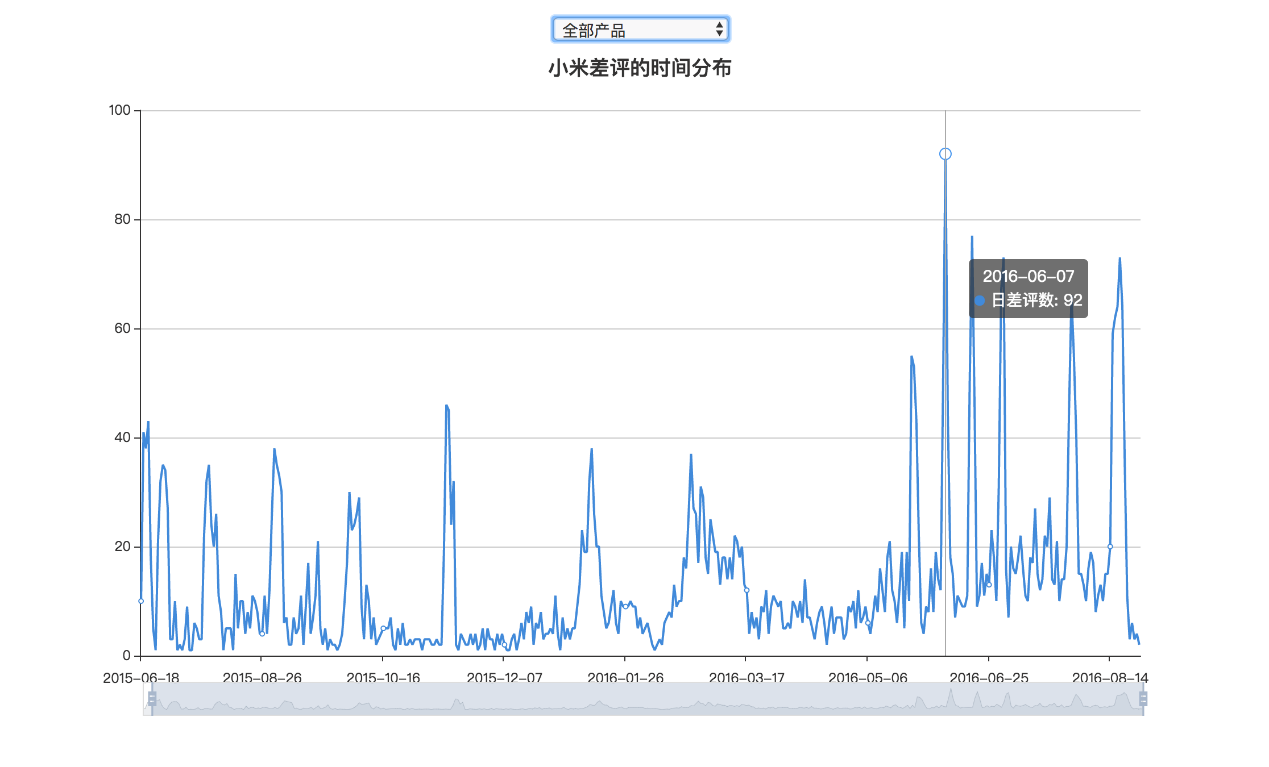


图21对应品牌差评时间分布图

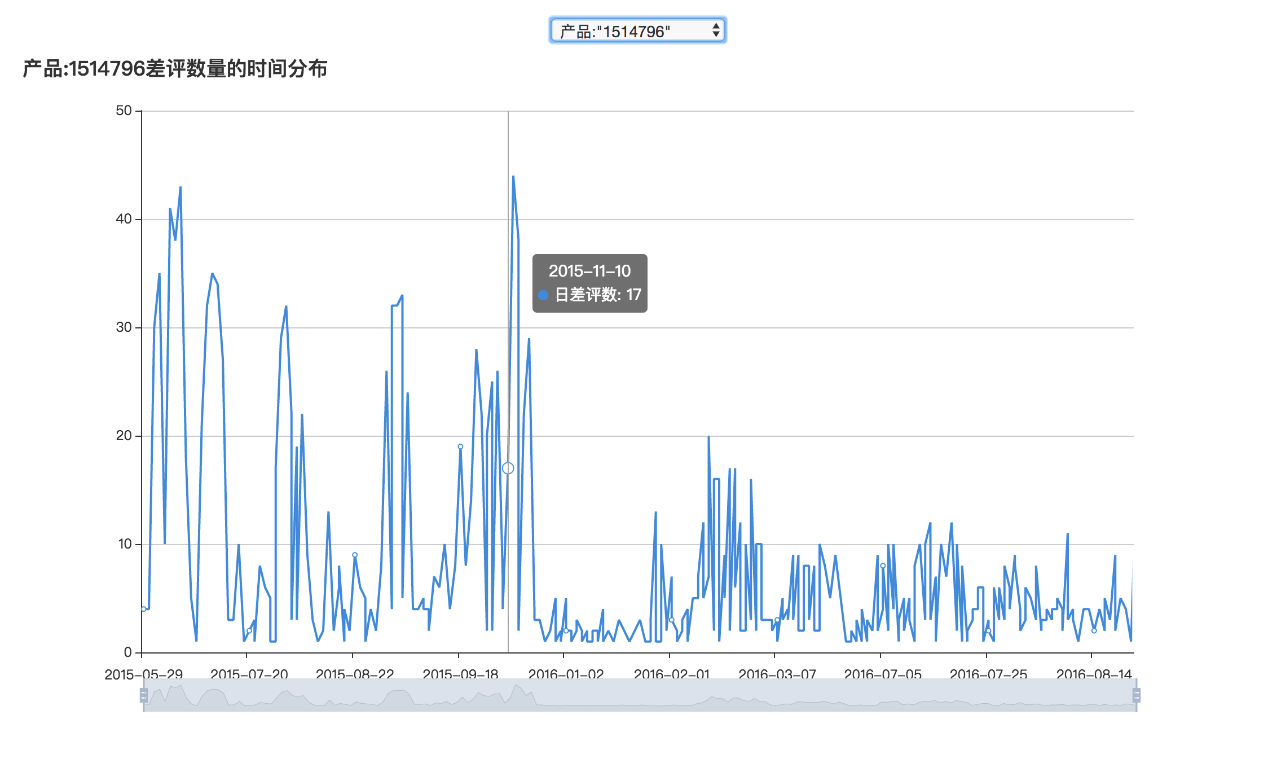


图22对应品牌典型产品的差评时间分布图



图23对应品牌各产品差评数量统计图