# 消费者运营分析

11.4.1评价词频分析&11.4.2 评价情感分析











# PART ONE 评价词频分析



词频分析是文本分析的基础,基于词频可以了解大量文本信息的内容,既可以分析自己的商品 也可以分析竞品。

接口说明如下。

- 通过基于机器学习算法分词的开源接口,可查阅接口的相关说明。
- 接口: http://api.pullword.com/get.php?source= &param1=&param2=。
- 接口参数如下。
- "Source=":要分词的文本。
- "Param1=": 返回分词结果的概率,最小为0,最大为1。
- "Param2=": 是否为调试模式,调试模式可以看到分词结果的概率。0为非调试模式,1为调试模



例1: http://api.pullword.com/get.php?source=清华大学是好学校&param1=0&param2=1。

返回:清华:0.604942 清华大学:1 华大:0.068537 大学:0.937618 好学:0.730473 好学校:0.699564 学校

例2: http://api.pullword.com/get.php?source=清华大学是好学校&param1=0.8&param2=1。返回:清华大学:1 大学:0.949906 学校:0.936925

例3: http://api.pullword.com/get.php?source=清华大学是好学校 &param1=0&param2=0 返回: 清华 清华大学 华大 大学 好学 学校



例4: 图11-78所示为采集某竞品的消费者评价数据,分析评价的词频。 数据采集路径: 【淘宝商品】-【评价详情页】。

解:选中数据,如图11-79所示,在【数据】选项卡中单击【自表格/区域】选项,将评价导入Power Query编辑器。

评价
今天收到了,今天就开始喝了,味道还不错,和 在成都买的一个味道,正品无疑!
这次的茶叶质量没有上次好,一粒粒都是扁扁的, 颜色也有带点黄,味道是没变,希望卖家给个解 释,找客服也找不到回应的声音。
感觉确是春茶,茶香四溢!6袋都是一样的,这个 价位很值!
非常好!一定再次购买 之前在成都喝过,很喜欢。这次买回来,感觉不 错。



, 如图11-80所示, 在Power Query编辑器中筛选数据, 搜索"此用户", 将无效的评价"此用 户没有填写评价。"过滤掉。 如图11-81所示, 在【添加列】选项卡中, 单击【自定义列】功能键。



	_				
文件	开始	转	奂	添加列	
	び 自定 义列 元	月大 调用自 注义函数		条件列 索引列 ▼ 重复列	-
		常规			,

图11-81

由于是访问互联网数据需要设置隐私级别,为了方便练习,如图11-82所示,可直接勾选【忽略此文件的隐私级别检查·····】,老版本没有这个选项,选择公共权限。

#### 如图11-83所示,键入公式"

Csv. Document (Web. Contents

("http://api.pullword.com/get.php?source="&Uri.

EscapeDataString

([评价]) & "& param1=0.8& param2=0"))"。

M函数说明:

Web. Contents: 是以二进制文件的格式下载网页。

Csv. Document: 是将二进制文件转换成Csv格式的表格。

<u>角秋</u> 级9		
られの別田	了。 于确保数据合并而不产生不必要的数据传输。不正确的隐私级别可能	导致敏感数据泄露到
モイムタスカリノロ	The second	and the second
言任范围之	外。有关隐私级别的更多信息,请参阅此处。	
言任范围之 2 忽略此文	外。有关隐私级别的更多信息,请参阅 此处。 件的隐私级别检查。忽略隐私级别可能会向未经授权的用户公开敏感数据回	成机密数据。
言任范围之 2 忽略此文 3 当前工作	外。有关隐私级别的更多信息,请参阅此处。 件的隐私级别检查。忽略隐私级别可能会向未经授权的用户公开敏感数据# F簿 ①	<b>龙机密数据。</b> ~~

#### 自定义列

	新列名		
	分词		
	自定义列公式:	可用列:	
	<pre>= Csv.Document(Web.Contents("http://api.pullword.com/get.php? source="&amp; Uri.EscapeDataString([评价]) &amp;"&amp;param1=0.8&amp;param2=0"))</pre>	评价	
	了解 Power Query 公式		
B	✔ 未检测到语法错误。	确定 取消	ſ

#### 消费者复购率计算与分析

获取数据后,如图11-84所示,勾选【展开】和【Column1】字段,展开后将字段 "Column1"重命名为"分词"。展开后,按照图11-85所示内容,筛选掉空白行。

A <sup>B</sup> c 评价	-	▲BC 分词 新設
很好喝的绿茶,	搜索要扩展的列	Ą
拿到茶叶就迫不!		2.1
特别愉快的体验。		
不怎么样,茶叶枝	✓ (选择所有列)	
茶叶的质量很好,	✓ Column1	
小杯子有点小		
今天收到了,今天		
这次的茶叶质量》		
感觉确是春茶,	🗌 使用原始列名作为前缀	
非常好!一定再)		
之前在成都喝过,	/ 列表可能不完整。	加载更多
茶叶非常香。应ì		确定取消
茶叶很好喝。独主		

A↓	升序排序	
Z↓	降序排序	
	清除排序	
$\mathbb{Y}_{\times}$	清除筛选器	
	删除空	
	文本筛选器	•
	搜索	
	■ (全选)	
	(空白)	
	✓ 1000	
	☑ 1000元	~
	✓ 4克	
	1 列表可能不完整。	加载更多
L		确定取消

图11-84 图11-85

#### 消费者复购率计算与分析

х.

如图11-86所示,在【开始】选项卡中单击【分组依据】选项。

如图11-87所示,基于分词进行分组。



分组依据			
● 基本 ○ 高级			
指定分组所依据的列以及所需的	喻出。		
分组依据			
分词 🔻			
新列名	操作	柱	
计数	对行进行计数 🔹		~
			确定取消

-

.

图11-87

以上内容仅为本文档的试下载部分,为可阅读页数的一半内容。如要下载或阅读全文,请访问: <u>https://d.book118.com/997123154034006166</u>