

## 面向网络舆情的关联度分析

**摘要：**网络舆情事件处理不当，会引发用户的过激行为。因此通过事件找出有关用户就可以起到一定的监督作用。本文先对用户提供的数据（用户信息和事件）进行预处理，对于用户信息，处理身份证号，还原成：性别，发证地，出生年月日，然后重复事件去除，用 replace pioneer 软件对 html 文件查找含 QQ，手机，作者等关键字，然后手动添加进 excel 里。对 txt 文件中的事件标题进行中文分词处理。接着抽取含有用户信息各属性的事件，转化为“用户-事件-用户”的形式，运用 clementine 软件，分析表的置信度和网络图直观性，得出余晓明和丁羽心和马小龙是关联的等类似结果。但是发现模型的不足，对于属性没有权重，同时也发现 html 文件中没有 MSN 信息，而且 QQ 信息没有能和事件信息匹配上，因此将用户信息分为：姓名，地址，性别（用户的地址是将原来的地址加身份证地址，姓名是加上关键字）3 个属性，将整个事件转换成 3 个指标，地址，姓名，性别。其中姓名是由 2 部分组成：html 中提取的作者，爆料人等，和题分词后的人名；性别由 html 中提取的性别，所以有大量缺失值；地址由两部分组成：根据网源得到的地方网址和事件标题分词后的地址，将处理后的用户和事件匹配相同的字数，利用 C++ 软件，求出结果，一个三维向量，通过欧氏距离，将结果排序。取出每个人距离最短的事件，发现钟建国和胡万林等人事存在关系的。通过这一算法，找到了用户和事件的关系，从而通过事件连接 2 个用户，找出用户之间的关系。再次对模型修改，距离标准化，计算用户与用户之间的关系，得出更清晰的效果。对于网络舆情来说，就可以从事件中找到相关人物以及和人物相关的其他人，那么这样就起到监督作用。

**关键词：**网络舆情 replace pioneer 中文分词 Clementine 关联度 欧氏距离

## Correlation analysis of network - oriented public opinion

**Abstract:** Importantly handled network public opinion events will cause bad users behaviors. Therefore, it is important to find out the concerned users. Firstly, we clarify data(user information and events):for the ID card information of the users', we convert it into gender, certification released department, and date of birth and then remove the repeated events and users. Secondly fill the missing values: search for keywords such as QQ, mobile phone, authors in HTML files with software Replace pioneer. Divide the titles in form of txt into words of different characteristics. Then we analyze the relativity: take out the valid data and then transform it into "users-events-users' form. Use Clementine SPSS to analyze the validity and diagram visualizability. One of the results indicates a strong relevance among Yu Xiaoming, Ding Yunxin and Ma Xiaolong. Nevertheless, due to reveled shortage of this mode, which are weighting little on users 'varieties and incapability to find out the related QQ and mobile phone information in HTML files, three new characters are employed: name(adding key words), address(the original address plus the ID card address) and gender. To be more exact, the name of the user consist of two elements, firstly the referred author and the information leaker and secondly the names after the characterized titles. Genders are extracted from HTML files, therefore lacking in a great amount of data. The address too, consists of two parts: the address from the network resources and characterized title. So the forth step is using three properties roundly to get the Euclidean distances between every event and user with VC2010. According to this algorithm relativity between users and events are found, thus two users are connected and the relativity between the users comes out along. In the end, we modify the model again. Clearer effects can be obtained after standardizing the Euclidean distances, and calculating the relationship among the users. In summary to network public opinion, if the relativity among concerned users can be found, supervisory effects can be applied.

**Key words:** Replace Pioneer Chinese Word Segmentation Relationship  
Clementine SPSS Euclidean Distance

## 目录

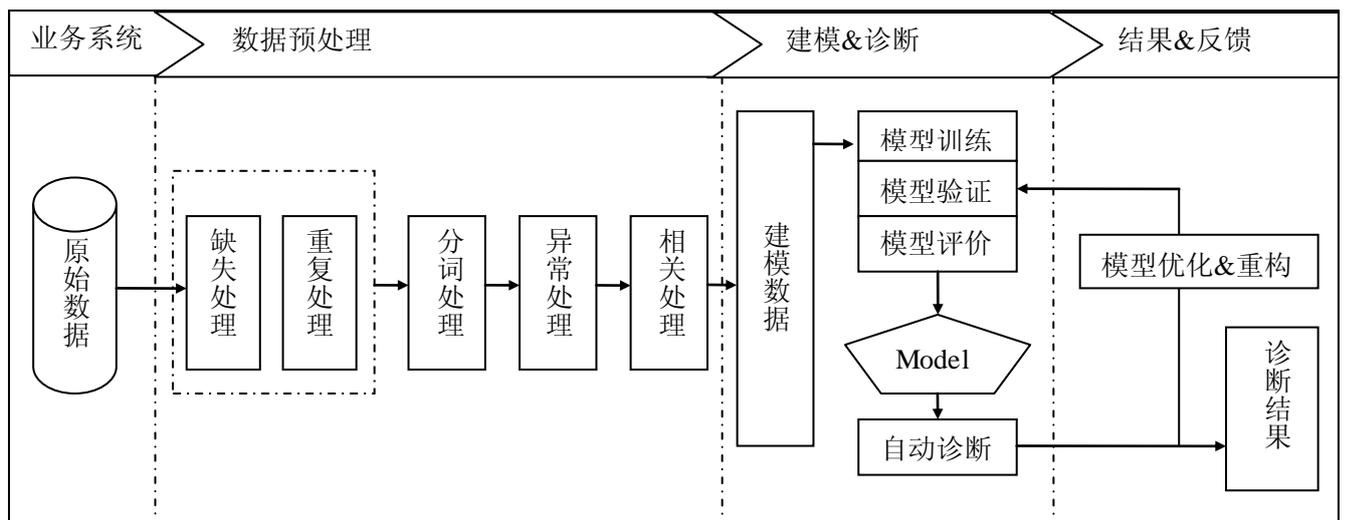
<b>1. 研究目标</b> .....	<b>4</b>
<b>2. 分析方法与过程</b> .....	<b>4</b>
2.1.总体流程.....	4
2.2.具体步骤.....	4
2.3.结果分析.....	8
<b>3. 结论</b> .....	<b>15</b>
<b>4. 参考文献</b> .....	<b>17</b>

# 1. 挖掘目标

本次建模的目标是利用客户提供的 2013 年热点事件和用户信息表，采用中文分词技术，建立用户和事件的相互关系，通过计算人物属性（姓名，性别，住址）和每个事件中相对应的分词结果中相同的字数，得出用户和事件之间的关联度，通过 clementine 软件，得出用户和用户的关联度。从而可以知道一个事件背后，这个人有多大关系，同时还能看出，这个人和其他人是否有关系。从而在一件事中，找出幕后操纵者和同伙。

# 2. 分析方法与过程

## 2.1. 总体流程



本次建模主要包括以下步骤：

步骤一：数据预处理

步骤二：建模和诊断

步骤三：模型优化

## 2.2. 具体步骤

**步骤 1：数据预处理**

◇ 缺失值处理

在用户信息表中，一些用户的身份证是错误的，无法修正，当成缺失值，因此该用户的身份证这一项不列入用户属性中。在提取 html 文件中，不一定能够把所需要的属性（如：

性别，地址）提取出来，若不能根据网址和标题分词得到的地址对地址进行填补，计算时当缺失值处理。

#### ◇ 重复值处理

在原始数据中，同样的事件可能会出现很多次，而经过访问，这些事件大多是抓取时间不同，代表了网站有更新，即事件的更新度，该事件的频率可以作为一个热度进行考虑，但在本次挖掘中，我们是研究用户与用户之间的关系，一个事件可能关系着几个用户，那么如果本事件重复出现，就会使这 2 个用户的关联更大，影响着最后结果的正确性。因此把重复事件全都去掉，只保留第一次出现的事件，同时提取了重复事件频率，方便研究事件的热度以及用户和事件频率的关系。

#### ◇ 分词处理

运用中科院的分词软件，将每个 txt 文本中事件标题进行分词，词性标注，以方便提取各个属性的词语。

#### ◇ 异常值处理

在分词后，由于分词软件的词库是有限大的，因此有些词语是识别不了。例如：奥巴马，会被自动分成 3 个单独的名词：奥，巴，马。因此，对于这些分词异常的词语，要进行人工处理，修正。分词后数据，见附件 1。

在用户表中，有身份证一列，而在 html 文件中，几乎没出现过身份证号，为了能充分体现着一项数据，因此可以把身份证转换：性别，出生年月日，发证地。以及一些错误身份证号码的修正。用户信息处理后，见附件 2。

#### ◇ 相关处理

为了找出用户与用户之间的关系，需要事件去连接起来。因此，各事件中找出具有用户任一属性（例如：姓名，住址，关键字）的事件，见附件 3，然后转化成“用户-事件-用户”这样结构的一张表，见附件 4。

## 步骤 2：建模与诊断

#### ◇ 用户与用户关联度

为了能让 Clementine 识别附件 4，因此把附件 4 转换为附件 5，以矩阵的形式存储数据。建立以下流程（附件中记为流 1）：

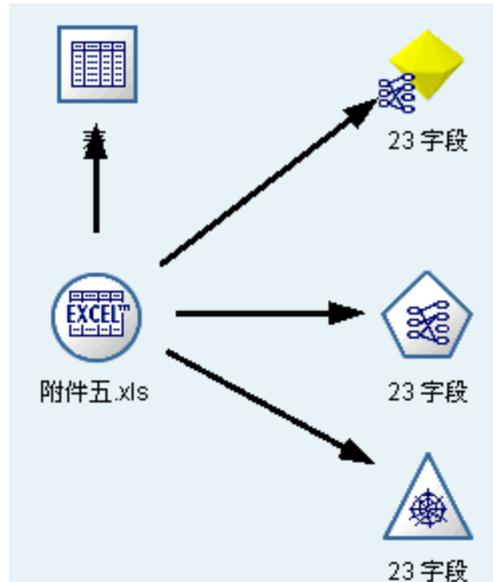


图 1

◇ 用户与事件关联度

1、因为编号以 1031 开头的事件均为新闻，有效属性值太少，故主要研究编号以 1083 开头的事件。通过 html 文件，通过对 html 文件进行查找所有可能跟用户有关的 QQ，手机，住址，MSN 等信息。因为帖子中一般发帖人不将个人信息公布，如果不进入网站查看该贴主的个人资料，从帖子中很少有内容是直接包含个人信息。但在一些招聘以及买卖场合，贴主会公布个人信息，如手机，QQ，性别等，且含个人信息的网址主要是西祠胡同 ([www.xici.net](http://www.xici.net))。故特通过 replace pioneer 文本处理软件，按照合适的正则表达式来提取含贴主个人信息的事件。并将提取得到的 QQ 手机性别等信息作为与该事件的重要属性。

另外，住址主要通过地方网站，如西子论坛 ([bbs.xizi.com](http://bbs.xizi.com))，而地方网站能很好的代表了贴主的住址地或者相关的地址。为了跟用户表的属性相对应，我们将此作为贴主最可能的居住地。毕竟新闻网站占了多数，造成大量与事件有关的居住地的缺失。当然对于新闻来说，事件标题所含的地址不一定能代表贴主的居住地，但仍以最大匹配为原则，根据事件标题分词得到地点对缺失值进行补充。见附件 11，12，13。

而对于用户表中的居住地，因为要跟事件进行关联，我们将该居住地进行扩充。利用身份证得到一个地址，如果两个地址相同，则不对地址进行扩充。如王林，用户表中住址为江西萍乡人，根据他的身份证，查得发证地也为江西萍乡。而对于两个地址不同的用户，如李天，用户表为北京市海淀区，而发证地为辽宁省抚顺市，故李天很可能关注的事件不仅仅是北京，还可能是辽宁，甚至仍使用辽宁的地方网。故我们得到李天相关的地址为北京海淀，辽宁抚顺。

因为 QQ，手机，MSN 等属性具有唯一性，故先对这些属性与用户的 QQ，手机，MSN 进行查找，除了余晓明附加关键字中出现的一个手机号码 18924889850 外，可发现里面未从出现出用户表里的 QQ，MSN，邮箱，手机。而再从 html 查找 18924889850，可知这是顺德 BBS 网站下的

一个热线电话，如果这个电话是与余晓明相关的，那么我们初步可以得到余晓明跟含顺德 BBS 有关，再通过他的另外一个附加关键字，汽车销售，从提取含 18924889850 的 html 提取出有关汽车销售的，发现均为二手汽车销售有关，则可以认为余晓明跟事 10831435, 10831436, 10831437, 1831438, 10838241, 10838242, 10838243 有关。除此之外，找不到用户跟 html 中的 QQ, MSN, 手机有相同的，因此把用户的姓名，地址，性别作为用户的 3 个主要属性，通过这 3 个属性，去比较每个事件，把事件变成一个 3\*1 的属性集。通过 C++ 软件（代码见附件 7）分别得到各属性的相同字数，组成一个 3 维向量，如事件 10834508-再曝萍乡巨贪姚波无法无天。因其对应的是新闻网，且相应的 html 没有性别，住址等信息，则根据标题分词得到 姚波, 江西萍乡#, 而每个用户的向量是根据每个属性所含的字数多少进行设置，如王林, 男, 江西萍乡#, 则对应向量为 (2, 1, 4), 则上述两个向量的关联向量是根据对应属性字数相同来得到，为 (0, 0, 4), 再通过 (2, 1, 4) 和 (0, 0, 4) 求欧式距离得到一个距离，如此，可以得到每个事件跟每个用户的欧氏距离，得到附件 6。通过描绘每个人对编号的散点图，如图 2.22。取最下面 2 个点。按照此做法，应用到每个人。

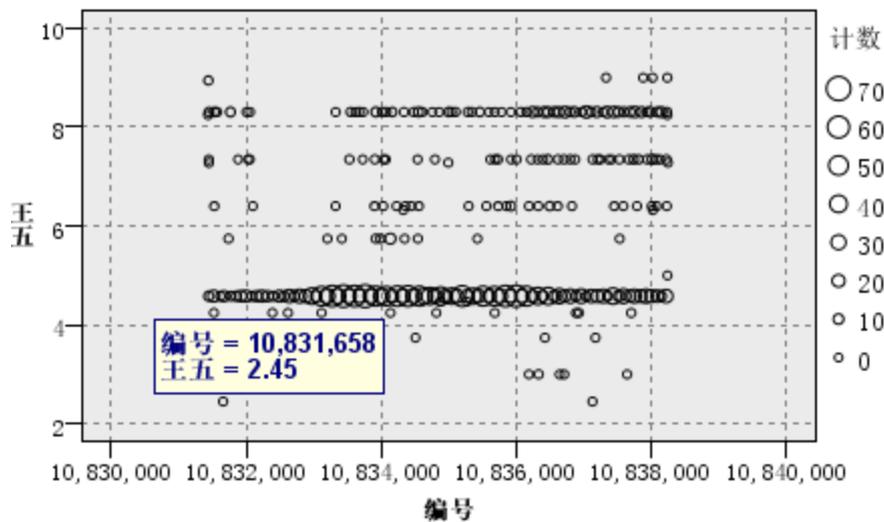


图 2

◇ 模型诊断

以上模型，我们可以容易发现某些用户对所有事件的距离始终保持很小，因为该用户的自身向量长度比较小，就算这个用户与事件没有关联，其自身向量- (0, 0, 0) 求得的欧式距离仍比较小。所以只能比较用户和事件，不能比较用户和用户的相关度。为消除量纲，我们将距离标准化得到附件 8。再从附件 8 中，对每个用户取出距离最短的 5 个事件，然后通过 Clementine 软件求出结果。流程如图 3，用户.txt 见附件 9。

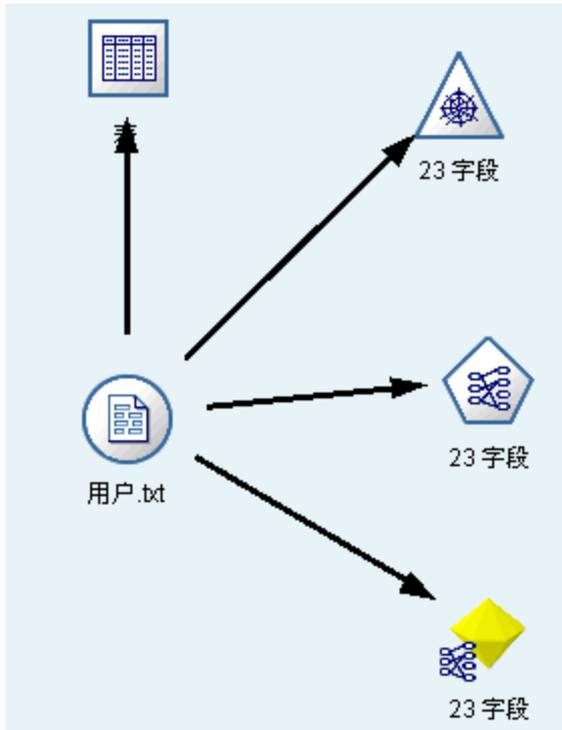


图 3

### 步骤 3: 模型分析和优化

#### ◇ 模型缺点

- 1、鉴于 clementine 的输入结构局限性，模型中，性别，姓名，地名设置了同样的权重，所以匹配的数量会非常多，实际上，如果设置好各项的权重，那么区分细度会更大。
- 2、很容易发现，几乎每个人之间都会有直接或间接联系，当匹配数取相对较小时，所有人都成为了一个集体。

#### ◇ 模型优点

- 1、利用 clementine 软件，过程容易操作，而且结果直观。
- 2、结果可靠，运用置信度，提升度分析。

#### ◇ 模型改进

- 1、先算用户-事件的关联度，利用欧氏距离计算用户与事件之间的距离，距离越小代表用户与事件的关联度越大。
- 2、通过 1 的结果，联系到用户-事件-用户，从而找出用户-用户之间的关联度。
- 3、在整个模型改进中，将距离标准化。这样就可以消除量纲不同带来的影响，从而使数据更加的准确。

## 2.3. 结果分析

1、用户与用户

(1) 通过 clementine 软件，实行以上流程，得到下图 4:

按以下内容进行排序: 置信度 %								
后项	前项	规则 ID	实例	支持度 %	置信度 %	规则支持 %	提升	部署能力
余晓明	李天	41	26	5.628	92.308	5.195	1.324	0.433
余晓明	周茂名	30	25	5.411	88.0	4.762	1.263	0.649
余晓明	张望	26	19	4.113	84.211	3.463	1.208	0.649
余晓明	王五	3	10	2.165	80.0	1.732	1.148	0.433
余晓明	王力宏	18	20	4.329	75.0	3.247	1.076	1.082
余晓明	张秋白	20	23	4.978	73.913	3.68	1.06	1.299
余晓明	高连岳	12	11	2.381	72.727	1.732	1.043	0.649
余晓明	郑玉龙	19	21	4.545	71.429	3.247	1.025	1.299
张望	方小明	72	5	1.082	40.0	0.433	9.726	0.649
黄浩	方小明	85	5	1.082	40.0	0.433	6.844	0.649
胡万林	黄明	79	6	1.299	33.333	0.433	1.949	0.866
张望	黄浩	68	7	1.515	28.571	0.433	6.947	1.082
方小明	黄浩	84	7	1.515	28.571	0.433	3.474	1.082
胡万林	黄浩	90	7	1.515	28.571	0.433	1.671	1.082
丁羽心	黄浩	37	27	5.844	25.926	1.515	1.361	4.329
李江	周世涛	6	9	1.948	22.222	0.433	2.702	1.515
黄明	胡万林	80	9	1.948	22.222	0.433	3.667	1.515
黄浩	胡万林	91	9	1.948	22.222	0.433	3.802	1.515
丁羽心	陈志祥	49	65	14.069	21.538	3.03	1.131	11.039
丁羽心	黄明	32	28	6.061	21.429	1.299	1.125	4.762

图 4

由于数据集中 0 居多，该事务级较稀疏，因此支持度较不可信，所以本文作者主要从置信度和提升度两方面来分析。

由于提升度一般都是大于 1 才有效，因此设置了筛选条件为：提升度>1。另外，还设置了置信度>20%。上图的结果就是满足设置条件，并且按照置信度从大到小排序的。从图中可以看到各个用户之间的关联，其中，多个用户与余晓明都有关联。再结合附件二中的用户信息表内容，发现余晓明这一用户的多个属性都与其他用户有相同之处。这说明此次关联是合理的，是和用户信息表紧密联系的。

(2) 同时也得出了网络图如图 5 所示:

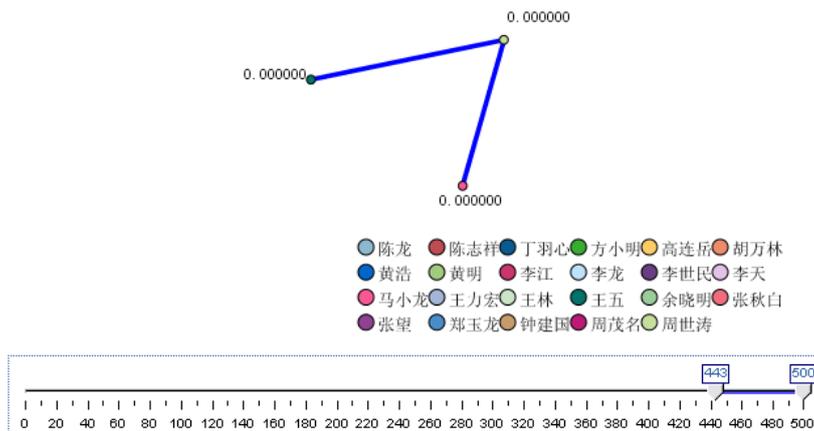


图 5

从图 4 中可见，移动范围到 443-500，余晓明与丁羽心、马小龙之间有关联，这亦可以从关联分析的结果中可以看出。说明关联规则和图形展示的结论一致。将匹配次数分等级，见表 1。等级越高说明相关性越大。根据这个等级划分，可以得出用户间的相关等级表如表 2 所示。

匹配次数	等级
430~432	D
433~439	C
440~441	B
442~443	A

表 1

关系用户	等级
周世涛-王五	A
周世涛-高连岳	A
高连岳-王五	B
周世涛-李世明	B
王五-李世明	B
高连岳-李世明	C
周世涛-张望	C
王力宏-周世涛	C
张望-王五	C

表 2

## 2、用户与事件

在建模过程中，附件就是用户与事件之间的关联。只是，该结果只知道用户与某个事件是否有关联，并不知道用户与各个事件的关联度排序，见附件 3。因此，在模型改进中，本文作者将解决该问题。

## 3、模型诊断

(1) 用户与事件之间的关联关系如附件 9 所示。此处就不再累述。

(2) 用户与用户之间的关联利用 Clementine 软件来进行分析。由图 3 得到结果如图 6 (部分结果) 和图 7。

后项	前项	规则 ID	实例	支持度 %	置信度 %	规则支持 %	提升
马小龙	王五	121	5	8.929	100.0	8.929	11.2
王五	马小龙	122	5	8.929	100.0	8.929	11.2
马小龙	胡万林	406	4	7.143	100.0	7.143	11.2
王五	胡万林	407	4	7.143	100.0	7.143	11.2
张秋白	陈龙	182	3	5.357	100.0	5.357	11.2
陈龙	张秋白	183	3	5.357	100.0	5.357	11.2
张秋白	陈龙	185	3	5.357	100.0	5.357	11.2
陈龙	张秋白	186	3	5.357	100.0	5.357	11.2
马小龙	陈龙	211	3	5.357	100.0	5.357	11.2
王五	陈龙	212	3	5.357	100.0	5.357	11.2
马小龙	张秋白	241	3	5.357	100.0	5.357	11.2
王五	张秋白	242	3	5.357	100.0	5.357	11.2
郑玉龙	王林	253	3	5.357	100.0	5.357	11.2
王力宏	王林	254	3	5.357	100.0	5.357	11.2

- 后项 包括所有值
- 前项 包括所有值
- 置信度 以上 90.0
- 条件支持度 以上 5.0
- 提升 大于 10.0

图 6

从图 5 中可以看出，此次模型检验虽然采用了关联分析。但由于此次的关联分析的事务数较少，支持度具有较大的说服力。所以从支持度、置信度、提升度来分析。从图中可以看出，王五、马小龙、胡万林、陈龙、张秋白几个用户之间具有较大的关联。这和建模过程得到了结果相差较多。由于，此次关联分析的置信度设置了大于 90%，支持度也大于 5%，另外，提升度也设置了大于 10.由此可见，此次的分析结果比建模时候的结果要精确的多。查看用户信息表（附件 2），发现，王五、马小龙、胡万林、陈龙都是广东广州地区的人，彼此的关联度确实比其他用户要大。这也反应了此次的关联分析的结果是合理、准确的。

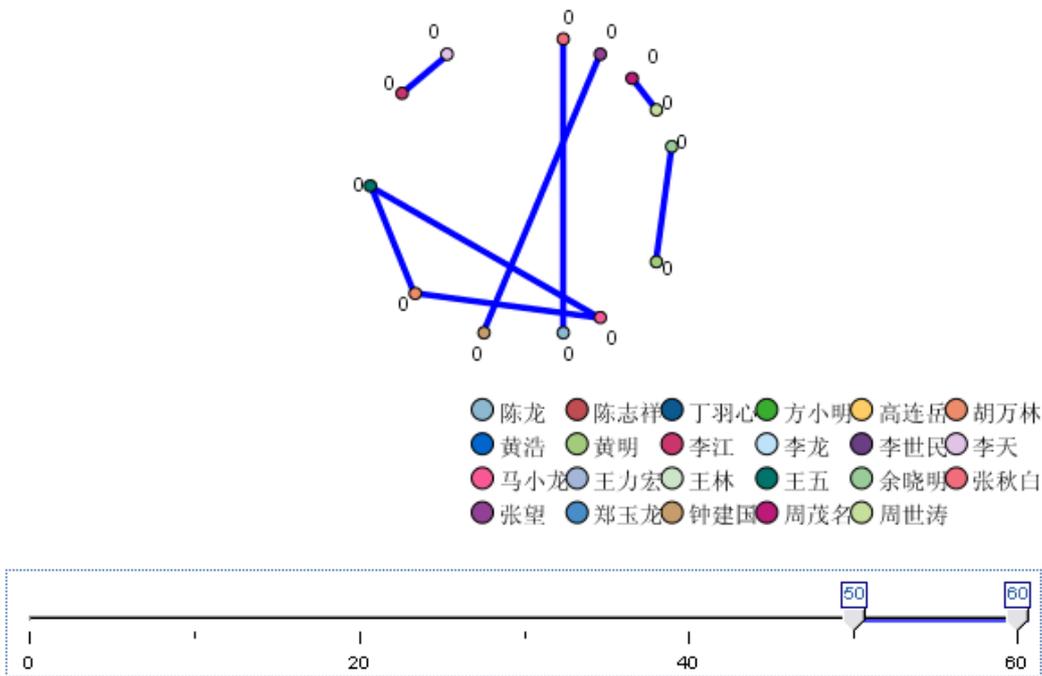


图 7

图 6 是移动图下方的滑块到频数较高区域后得到的网状图。从图中可以看到，此时，有关联的用户有：李江和李天、张秋白和陈龙、张望和丁羽心、周茂名和周世涛、余晓明和黄明、王五和马小龙以及胡万林。这和关联规则分析的结果一致。说明此次的关联分析是成功的。

(3) 因为表格之间的数值难以直接体现事件与用户之间的距离差，但因数据量过大，只选取用户和事件的距离进行进一步分析。为消除量纲，将c++程序得到的距离进行标准化形成 distance1 表格。使用 SAS 画出相应的散点图或雷达图。一个事件与各个用户：

1. 根据事件 10831430 与各个用户的距离得到的散点图分别如下：

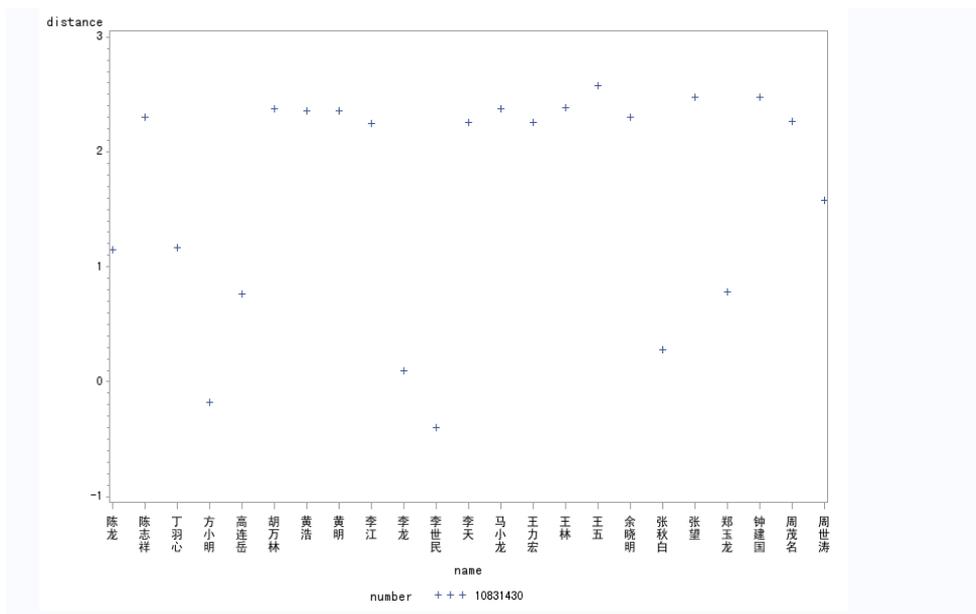


图8

可见 10831430 事件与用户李世民和方小明的距离较小，也就是关联比较大,并且较多的用户跟 10831430 事件的距离还是大的，而且基本与最大距离持平，也就是这些用户跟事件基本没有关系。而 10831430 具体的事件是：深圳 10 单位涉假劣药品被查，该事件去重后，计算出频率为 1372，是所有事件中频率最高的事件。而频率也代表了它的受关注程度，对于该事件所关联的用户应该得到重视。因此选取了与该事件距离近的用户资料进行比对：男, 广州, 李龙；男, 广州, 方小明；男, 广州, 马六李世民；男, 深圳广州, 陈龙；男, 深圳广州, 李四张秋白。发现距离最近的四个：方小明，李龙，李世民，张秋白竟然均在广州，而在深圳的陈龙，与事件 10831430 却不见得很近。代码见附件 11。

于是特意选取了在广州发生的事件10831537与各个用户进行分析，得到散点图如下：

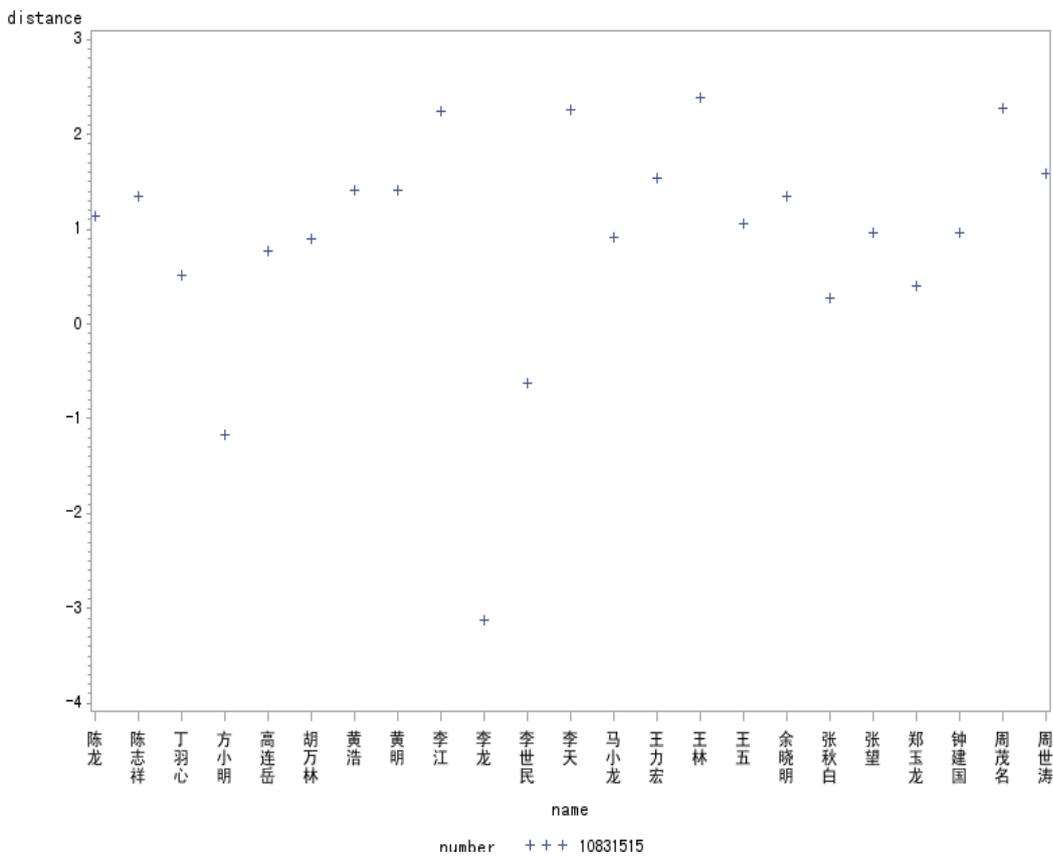


图9

这时李龙与10831515的距离明显最近，接着是方小明，李世民，这样看来，对于事件与用户的距离小于-1（标准化后）比较有可信度，而对于距离>0的用户和事件可以说几乎没关系。

多个事件多个用户：

2.探寻 10831476, 10831477, 10831478, 10831479, 10831504, 10831512, 10831514, 10831515, 10831517, 10831518, 10831519 事件跟用户高连岳，王力宏，郑玉龙，丁羽心的关系：

散点图：

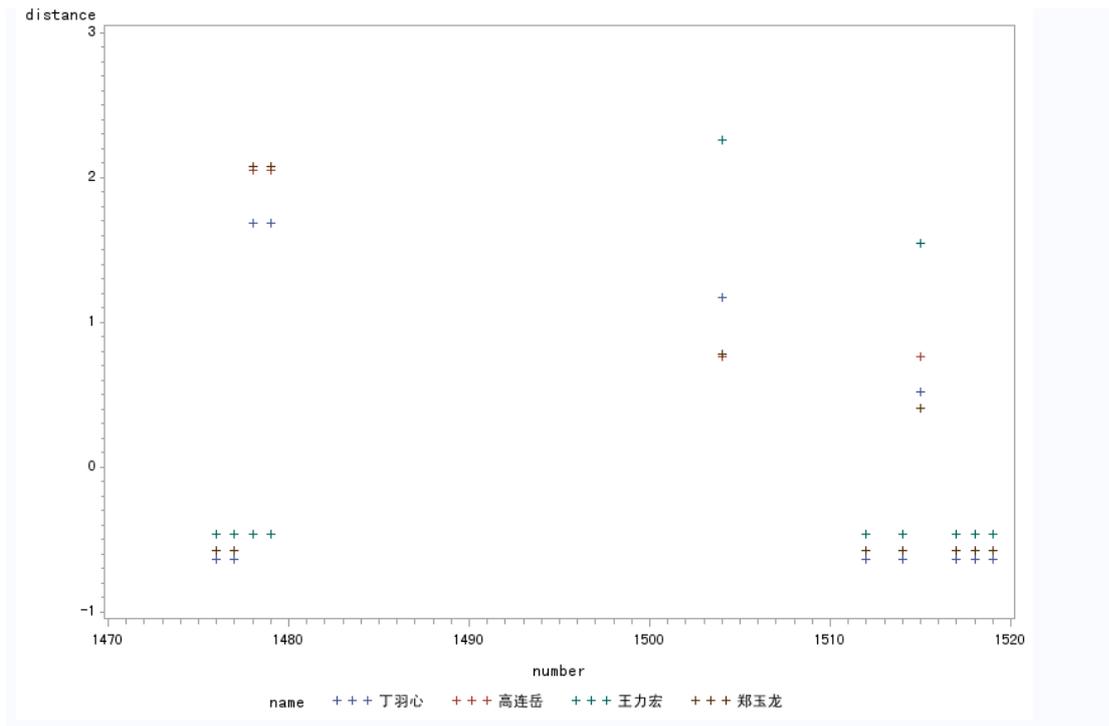


图 10

在上述事件中，丁羽心跟除了跟 10831478,10831479 事件外的关联都比较大。10831047 麦莉-塞勒斯承认与男友分手现在享受单身；10831479 李某某等强奸案 12 月底或将终审因为姓名的原因跟丁羽心找不到关系。10831512, 10831514, 10831515, 10831518, 10831519 跟丁羽心，王力宏，郑玉龙的关联比跟高连岳强。但因为距离都比较靠近，不特别悬殊，因此不能说这些事件跟用户的关系程度。

事件 10831537 10831673 10834508 10836579 事件与李世民，李龙，陈龙，李江进行进一步探究，因为雷达图数据需 $>0$ ；故对上述的距离+4，进行作图，结果更加鲜明：事件 10831673 与李龙特别近，考察事件 10831673：任泉 25 岁女友美艳近照曝光疑为李双江爱徒(组图)

本来标题中含李双江，目测跟用户中的李江是关联比较大的，而且事件中也提取不了地址，故单单从姓名来得到的距离显示，也是李江与该事件联系最大下述结果也比较吻合，而比较其他类型的姓名，李龙，李世民，陈龙，王林距离越来越大。而对于事件 10834508：再曝萍乡巨贪姚波无法无天。其中唯一信息也只有萍乡，而用户王林恰好就是江西萍乡的，于是在下图 11 中，距离明显比其他用户的近。

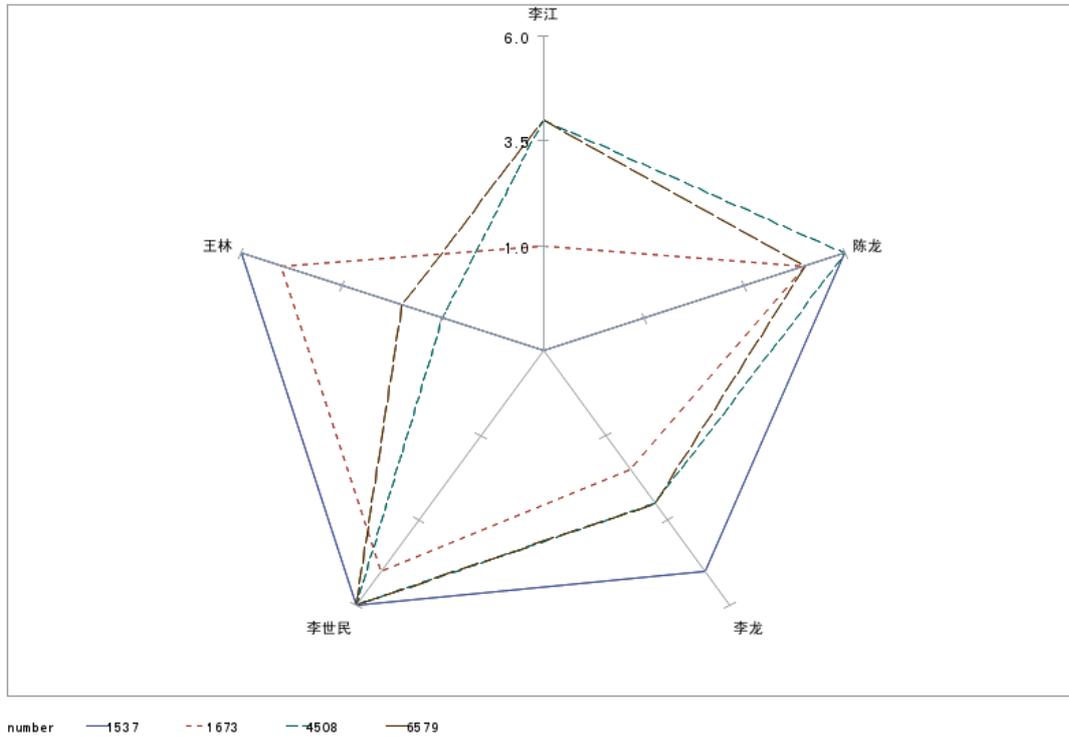


图 11

### 3. 结论

本次建模的目标是建立用户和事件的相互关系，并得出用户和用户的关联度。通过此次的数据挖掘，得到用户与事件的关联如表 3 所示：

用户名	王林	高连岳	王力宏	郑玉龙	丁羽心	胡万林	周茂名	周世涛
事件一	10834508	10831618	10831658	10831658	10831545	10831658	10831709	10831709
事件二	10831658	10836244	10837131	10837131	10835672	10837131	10837644	10837644
事件三	10837131	10837871	10831583	10831583	10835854	10831718	10832544	10832544
事件四	10831583	10838194	10834136	10837631	10834874	10831603	10833204	10833204
事件五	10831627	10831432	10836878	10831603	10835269	10831605	10834259	10832316

用户名	李天	李江	陈志祥	黄明	黄浩	余晓明	张望	方小明
事件一	10832365	10832365	10832365	10835759	10836922	10835759	10831709	10835759
事件二	10833917	10833917	10837631	10835881	10834203	10835881	10837644	10835881
事件三	10834334	10834334	10831562	10836922	10833665	10834156	10831658	10835767
事件四	10834822	10834822	10837856	10834874	10834232	10834874	10837131	10835885
事件五	10835094	10835094	10833307	10835269	10834793	10835269	10831718	10835912

用户名	张秋白	王五	李世民	钟建国	李龙	陈龙	马小龙
事件一	10831628	10831658	10832316	10831709	10836876	10831603	10831658
事件二	10831603	10837131	10831628	10837644	10834326	10831605	10837131
事件三	10831605	10831603	10832365	10831658	10838025	10831647	10831603
事件四	10831647	10831605	10838132	10837131	10831709	10831850	10831605

事件五      10831850      10831647      10837864      10831603      10837644      10832081      10831647

表 3

为了更直观显示事件，于是将事件换另一个编号，得到表 4：

用户名	王林	高连岳	王力宏	郑玉龙	丁羽心	胡万林	周茂名	周世涛	李天	李江	陈志祥
事件一	1	6	1	1	15	3	22	22	28	28	14
事件二	2	7	3	3	16	5	23	23	29	29	28
事件三	3	8	5	5	17	13	24	24	30	30	33
事件四	4	9	11	13	18	20	25	26	31	31	34
事件五	5	10	12	14	19	21	26	27	32	32	35

黄明	黄浩	余晓明	张望	方小明	张秋白	王五	李世民	钟建国	李龙	陈龙	马小龙
16	38	16	3	36	13	3	27	3	22	13	3
17	39	17	5	37	20	5	28	5	26	20	5
36	40	36	21	44	47	13	47	13	52	48	13
37	41	37	22	45	48	20	50	22	53	49	20
38	42	43	26	46	49	48	51	26	54	55	48

表 4

根据上表等级排序，可得出用户与用户之间的关联如下所示：

一级关系：李天-李江

二级关系：周世涛-周茂名，黄明-余晓明，张望-钟建国，王五-马小龙

三级关系：王林-郑玉龙-王力宏，郑玉龙-马小龙-钟建国

## 4. 参考文献

- [1] 林宇等编著. 数据仓库原理与实践. 北京: 人民邮电出版社, 2003
- [2] 朱明, 数据挖掘. 合肥: 中国科技大学出版社 2002, 5
- [3] 陈京民等. 数据仓库与数据挖掘技术[M]. 北京: 电子工业出版社, 2002.
- [4] 毛国君等. 数据挖掘原理与算法[M]. 北京: 清华大学出版社, 2005.
- [5] 陈文伟等. 数据挖掘技术[M]. 北京: 北京工业大学出版社, 2002.