

# 基于系统聚类的警情分析及应用 \*

张立<sup>1</sup> 闵国诚 唐斌 范馨月

1. 贵州大学数学与统计学院

2. 贵州省贵阳市公安局

**摘要：**根据公安部门 110 报警平台报警数据，利用统计和系统聚类法对贵阳市警情进行分析，特别对两类高发案件，按照报警案件发生的分局、时间进行定量分析。通过对不同类别警情的空间分布和时间分布，借助谱系聚类图将警情类别进行合理分类，帮助相关管理部门及时发现并调整当前工作重心，合理部署警力等相关资源。

**关键词：**110 报警平台 系统聚类 警情分析 时间分布

## 引言

近年来，公安信息化建设一直保持着快速平稳的发展态势。随着各级公安部门信息化建设的深入开展，可利用的数据急剧增大，若能够对这些海量数据与信息进行快速深入的分析和处理，从中找出规律，可以帮助公安管理部门更好地进行运筹与决策。

随着大数据时代的快速发展，逐渐应用一些警情研判系统，来宏观观测警情发生的总体情况。其中大部分观测属于定性的分析，也有部分定量分析，研究了关于一个犯罪点的短期数量增长情况，但之前的研究较少结合不同时间段和不同地点的影响进行警情分析<sup>[1]</sup>。

本文选取贵阳市公安部门 110 报警平台接处警实际数据，从多个不同类别的案件报警中选取了两类高发报警类别。利用系统聚类法对这两类高发案件报警情况进行分析处理，按照不同类别的热点分局、时间进行详细分析，找出它们之间的关系。为了更好地观测样本点的分布情况，尝试画出两类案件报警情况数目的三维观测图，通过对不同类别警情的分布进行可视化。再借助谱系聚类图将它们进行合理分类，帮助相关部门及时发现并调整当前工作重心，合理对警力资源进行部署。

## 一、数据处理

本文采用聚类分析对实际案件数据进行检验分析，案件数据来源于贵阳市的公安 110 报警平台报警记录，时段为 2018 年 1 月至 10 月。在本阶段，先将这些大量的数据进行整理分析，根据 2018 年 1 月到 10 月的 110 报警数据显示，这 10 个月一共发生 720623 起报警事件，将每个月分为上半月和下半月，得到 20 组每半月全市报警总数，观察其分布（图 1）。由图 1 可以看出，每半月的报警总数整体是随着时间的推移而明显增加。

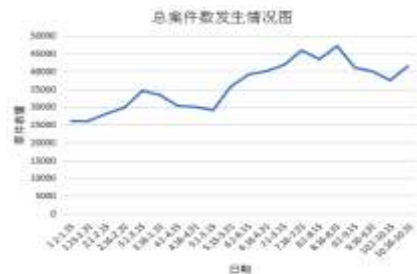


图 1 报警数时序图

通过数据整理发现，贵阳市 110 报警平台原数据记录中包含报警细则 221 种，报警类型 139 种，报警类别 27 种，所属分局 18 处。为了后期对数据更好地叙述及处理，现将报警类别和所属分局进行分类标号（见表 1、表 2）。

表 1 警情类别分类标号

报警类别	类别分类	报警类别	类别分类
安全事故	第 1 类	群众求助	第 15 类
城市管理类	第 2 类	骚扰电话	第 16 类
火灾事故	第 3 类	社会联动	第 17 类
火灾事故类	第 4 类	事件	第 18 类
交通类警情	第 5 类	通知	第 19 类
交通事故类	第 6 类	突发性事件	第 20 类
金融报警	第 7 类	无效警情	第 21 类
纠纷	第 8 类	刑事案件	第 22 类
举报投诉	第 9 类	行政（治安）案件	第 23 类
举报线索类别	第 10 类	应急联动事件	第 24 类
其他职能	第 11 类	治安案件	第 25 类
其它警情	第 12 类	重复报警	第 26 类
求助咨询类	第 13 类	自然灾害	第 27 类
群体性事件类	第 14 类		

表 2 所属分局分类标号

所属分局	类别分类	所属分局	类别分类
白云公安分局	第 1 类	市交警支队	第 10 类
便衣分局	第 2 类	双龙分局	第 11 类
公交分局	第 3 类	塔山公安分局	第 12 类
观山湖公安分局	第 4 类	乌当公安分局	第 13 类
花溪公安分局	第 5 类	息烽县公安局	第 14 类
经开公安分局	第 6 类	修文县公安局	第 15 类
开阳县公安局	第 7 类	云岩公安分局	第 16 类
南明公安分局	第 8 类	指挥中心	第 17 类
清镇市公安局	第 9 类	中心公安分局	第 18 类

为了较好地对警情进行分析，选取 27 种报警类别进行观测。如图 2 所示，交通事故类（第 6 类）案件报警总数

\*基金项目：贵州省大数据重点实验室开放课题（编号：2017BDKFJ012）；贵州大学省级本科教学工程项目（编号：2017520015）；贵州大学博士基金项目（编号：贵大人基合字 2012015）号；贵州大学“本科教学工程”建设项目（编号：G201723）

数、求助咨询类（第 13 类）案件报警总件数分别占有所有类别总件数的 25.447% 和 41.798%。

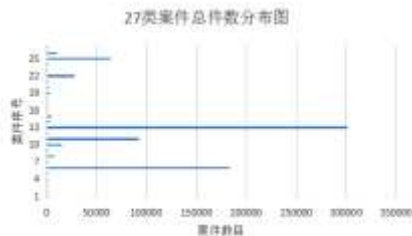


图 2 7 类案件总数分布图

将 27 类警情进行逐一分析，需要整合的数据复杂，不易找出其中联系。因此将这 10 个月中报警类别总件数小于 20 件的案件类别剔除，剩余 20 类报警案件类别。将这 20 类的报警案件类别按时间进行排列，得到 10 个月内不同类警情发生数目的时序图（图 3）

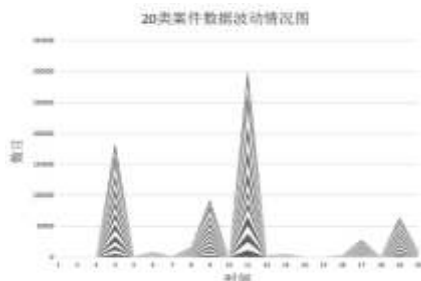


图 3 20 类案件时序图

这 20 类数据随时间推移，整体波动情况相似。所以可以考虑选择前两个高发案件报警类别进行分析，即交通事故类（第 6 类）案件和求助咨询类（第 13 类）案件。

将交通事故类和求助咨询类的报警案件数目按照每半月为一个单位，将信息整理出来得到了两组数据，每组包含 20 个（表 3），以及这两类案件报警数目在 17 个不同分局的分布情况。绘制这两类高发案件时序图（图 4），其中交通事故类案件增长情况较为稳定，求助咨询类案件数目波动幅度较大，在 5~9（3 月上半月~5 月上半月）处于一个递减的状态。从整体上看，每半月这两类案件报警数目都呈现

表 3 两类案件报警数目表

月份	交通事故类		求助咨询类	
	上半月	下半月	上半月	下半月
一月	5568	5598	11595	12995
二月	6543	6988	13380	13245
三月	9184	9728	14670	12217
四月	9393	9125	9924	9452
五月	8919	9305	9367	13558
六月	9628	9524	16660	17510
七月	10169	11288	18688	20275
八月	10710	11431	18615	19977
九月	9664	9951	17782	17511
十月	9369	11292	16442	17343
总计	183377		301206	

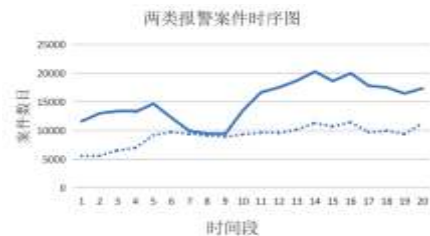


图 4 两类高发报警案件时序图

一个递增的趋势，后面将利用这两类案件报警数据情况进行聚类分析。

## 二、聚类分析

### （一）系统聚类分析

系统聚类法是一种使用较多的聚类方法，这种方法先将所需要聚类的样本，变量分别看成一个群体，确定不同类别之间的相似统计量，选择最接近的两类或若干个类重新组合成为一个新类，接着计算新类与其它各类之间的相似性，选择最近的两群或若干群重新组合成一个新类，直到所有的样本或变量都合并成一类为止<sup>[2]</sup>。

假设现在已知  $m$  个样本点，每个样本点需要考虑到多项指标（记有  $k$  项指标），得到一个样本数矩阵：

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mk} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} \quad (1)$$

这里， $a_{ij}$  表示的是第  $i$  个样本第  $j$  项指标的值， $A_i (i=1, 2, \dots, m)$  表示的是第  $i$  个样品，用  $d_{ij}$  表示样本  $i$  和样本  $j$  之间的距离。

根据这种思想，可以利用系统聚类将现有样本点划分为一类，接着把其中一些最为相似的样本点聚集在一类里，将已经聚合的类别按各自之间的相似性再次聚合。这样随着相似性的递减，最后观察进行判别，确定所划分的大类，组成一个依照相似性的大小聚集而成的谱系聚类图。

系统聚类法的聚类情况取决于样本点的相似系数和不同类别间的距离（类间距离）。关于类间距离常见的有：最短距离法、最长距离法、中间距离法、可变距离法等八种方法。本文考虑到需要利用已知的样本点来判断不同时间地点案件报警的一些最为相似的分布情况，采用的是最短距离法：

$$D_{pq} = \min_{i \in G_p, j \in G_q} d_{ij} \quad (2)$$

其中， $G_p$ 、 $G_q$  分别表示的是两个类别，各自包括了  $m_p$ 、 $m_q$  个样本， $D_{pq}$  表示的是两类别样本之间的距离最短情

况，将它视作两类间的最短距离。

系统聚类法的基本步骤（基于matlab下的算法）：

Step1 输入得到的样本点，建立 $m \times k$ 阶矩阵；

Step2 计算各样品各自之间的距离，并利用（2）得到 $G_p$ 、 $G_q$ 两类样本之间的最短距离 $D_{pq}$ ；

Step3 合并符合类别距离要求的两类作为一个新类；

Step4 继续计算新类与现有类之间的距离，继续合并最短距离的类为新类；

Step5 循环 Step1 到 Step4，直至类个数为1；

Step6 画出谱系聚类图；

Step7 决定分类个数，得出各类的样本。

### （二）样本情况

在数据的处理部分，已经确定了两类案件报警数目在1到10月的数目情况，得到这两类案件报警数目在17个不同分局的分布情况。在进一步整理小样本数据的时候，发现交通事故类案件报警在第2、3、11、12、13、16、17、18分局发生概率极小，求助咨询类案件报警在第2、10、11、12、18分局发生概率极小。考虑到18类不同分局进行逐个分析，需要整合的数据较多，不易找出其中关系，将10个月内所属分局总件数小于260件的分局剔除。

其中，交通事故类案件报警剩余10处不同分局的分布情况，求助咨询类案件报警剩余13处不同分局的分布情况。继续将这两类案件报警所包含的剩余分局重新进行一个分类标号（见表4、表5），绘出交通事故类和求助咨询类案件报警情况数目的三维观测图（图5、图6）。

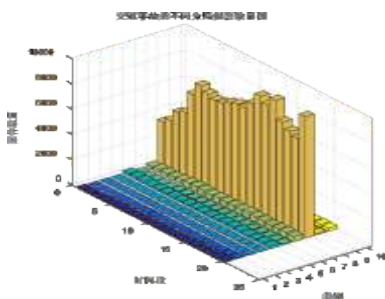


图5 交通事故类不同分局报警数目图

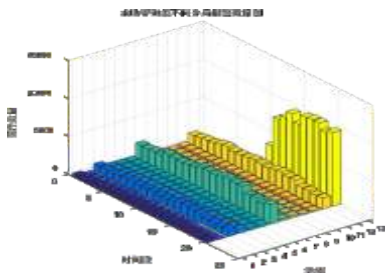


图6 求助咨询类不同分局报警数目图

表4 交通事故类的所属分局分类标号表

属分局所	类别分类
云公安分局	第1类
山湖公安分局	第2类
溪公安分局	第3类
开公安分局	第4类
阳县公安分局	第5类
明公安分局	第6类
镇市公安分局	第7类
交警支队	第8类
烽县公安局	第9类
文县公安局	第10类

表5 求助咨询类的所属分局分类标号表

属分局	类别分类
云公安分局	第1类
交分局	第2类
山湖公安分局	第3类
溪公安分局	第4类
开公安分局	第5类
阳县公安分局	第6类
明公安分局	第7类
镇市公安分局	第8类
当公安分局	第9类
烽县公安分局	第10类
文县公安分局	第11类
岩公安分局	第12类
挥中庭	第13类

图5绘出了交通事故类案件报警在10处分局发生的分布情况，可以十分明显地看到交通事故类在市交警支队（第8类）发生是最频繁的，其他9处分局发生情况一般。图6绘出了求助咨询类案件报警在13处分局发生的分布情况，可以十分明显地看到在指挥中心（第13类），从6月到10月求助咨询类案件报警的发生是最频繁的，而南明公安分局（第7类）和云岩公安分局（第12类）求助咨询类案件报警的发生是较为频繁的，其他10处分局求助咨询类案件报警的发生情况一般。

于是选取交通事故类案件报警的10组数据和求助咨询类案件报警的13组数据进行分析，见表6和表7。

表6 交通事故类案件报警的部分数据

时间段	开阳县公安局	清镇市公安局	市交警支队	修文县公安局
1.1 ~ 1.15	216	210	3458	185
1.15 ~ 1.31	260	287	3864	203
2.1 ~ 2.15	224	281	4784	197
2.16 ~ 2.31	144	268	5110	131
3.1 ~ 3.15	155	345	6837	187
3.16 ~ 3.31	170	423	7824	211
4.1 ~ 4.15	188	443	7478	182
4.16 ~ 4.31	152	425	7256	195
5.1 ~ 5.15	158	372	7354	146
5.15 ~ 5.31	190	442	7647	206
6.1 ~ 6.15	176	423	7875	199
6.16 ~ 6.31	152	446	7899	166
7.1 ~ 7.15	208	415	8524	164
7.16 ~ 7.31	240	511	9413	268
8.1 ~ 8.15	231	444	9000	259
8.16 ~ 8.31	229	441	9697	238
9.1 ~ 9.15	180	380	8288	220
9.16 ~ 9.31	189	424	7829	228
10.1 ~ 10.15	220	452	7586	262
10.16 ~ 10.31	197	406	9605	238

表 7 求助咨询类案件报警的部分数据

时间段	观山湖公安	南明公安	云岩公安	指挥中心
1.1~ 1.15	869	2001	1259	60
1.15~ 1.31	870	1946	1368	62
2.1~ 2.15	663	2085	1488	56
2.16~ 2.31	944	2054	1590	66
3.1~ 3.15	988	2536	1857	64
3.16~ 1.31	1092	2490	1981	76
4.1~ 4.15	949	2393	1777	80
4.16~ 4.31	960	2548	1789	60
5.1~ 5.15	1019	2647	1876	56
5.15~ 5.31	1050	3017	2054	57
6.1~ 6.15	997	3046	1968	3942
6.16~ 6.31	1062	2977	1831	7864
7.1~ 7.15	1358	3205	2046	8516
7.16~ 7.31	1401	3347	2220	9516
8.1~ 8.15	1339	3173	1979	8548
8.16~ 8.31	1251	3265	1996	9907
9.1~ 9.15	951	2516	1494	10061
9.16~ 9.31	995	2546	1533	10065
10.1~ 10.15	855	2308	1367	9292
10.16~ 10.31	936	2612	1434	9697

### (三) 聚类

根据上述两组案件报警数据按照最短距离法做聚类分析,并绘出交通事故类案件报警和求助咨询类案件报警类的谱系聚类图<sup>[1]</sup>(图 7、图 8)

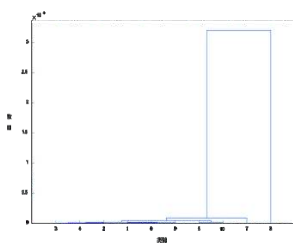


图 7 通事故类案件报警谱系聚类图

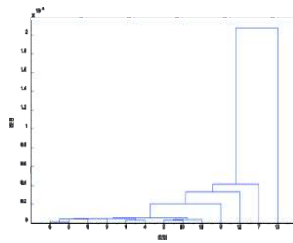


图 8 求助咨询类案件报警谱系聚类图

根据案件报警谱系聚类图,可以将这两类案件报警情况分为几个大类,见表 8、表 9。

表 8 通事故类案件报警的分类情况表

类	分局	分布情况
1	1~7、9、10 类	各分局发生数目相近,都较少发生。
2	市交警支队(第 8 类)	发生数目相对很高,占该组数据总样本的 87.38%。

表 9 求助咨询类案件报警的分类情况表

类	分局	分布情况
1	云岩公安(第 12 类)	一般发生。
2	1~6、8~11 类	各分局发生数目相近,都较少发生。
3	南明公安(第 7 类)	一般发生,比云岩公安分局更普遍发生。
4	指挥中心(第 13 类)	发生数目相对较高,占该组数据总样本的 33.76%。

在表 8 中,将交通事故类案件报警情况依照不同的分局分为两类。第一大类是第 1~7、9、10 分局,它们各分局在相同的时间段里发生的数目是相近的,而且都较小。第二大类是市交警支队,发生的数目相对于其他分局而言较高,占本组 10 个分局样本数目的 87.38%。而且从图 5 也可以清晰地看出,交通事故类案件报警在市交警支队最

易发生,其他 9 处分局发生的情况一般。

发现这个分布情况后,公安相关部门应加强交通事故类案件报警情况在市交警支队的管理力度,考虑将其他分局的剩余资源适当转移到市交警支队。

表 9 将求助咨询类案件报警情况依照不同的分局分为四类,第一大类是云岩公安分局(第 12 类),它在这 10 个月发生的情况相对稳定。第二大类是第 1~6、8~11 分局,它们在相同的时间段里发生的数目相近,而且都较小。第三大类是南明公安分局(第 7 类),它在这 10 个月发生的情况稳定,较云岩公安分局更为普遍。第四大类是指挥中心(第 13 类),它的表现较为突出,发生的数目相对于其他分局较高,占本组 13 个分局样本数目的 33.76%。而且从图 6 也可以清晰地看出,求助咨询类案件报警在指挥中心最易发生。

发现这个分布情况后,公安相关部门应加强求助咨询类(第 13 类)案件报警情况在第 7、12、13 分局的管理力度。再结合求助咨询类案件报警情况数目的三维观测图,另外还可以发现在下半年的时间里,求助咨询类案件报警数目迅速增加,建议公安相关管理部门在下半年密切关注这类求助咨询的问题,增设部分求助咨询点,更好地服务群众。

## 三、结语

本文结合公安部门 110 平台的报警情况,分别从案件的时间分布、空间分布进行警情分析。借助系统聚类对贵阳市不同报警类型、发生分局、发生时间的历史数据进行分析对比,能够较好地找出各类警情特点。

鉴于警情的发生与社会各项事件关系复杂,本文从月份、分局、报警类型等影响较大的关联因素入手进行警情分析。帮助相关部门能够在全局观测社会情况,理性分析警情分布情况,及时发现并调整当前工作重心,合理利用部署有限资源<sup>[4]</sup>。

## 参考文献

- [1]朱盛泉. 警预警和重点对象查控平台的设计与实现[D]. 江西财经大学, 2017.
- [2]沈大庆主编, 黎荣、马吉臣、张红宁副主编. 警建模. 国防工业出版社[M], 0102.
- [3]李家. 警分析谱系图二叉树的建立及应用[J]. 辽宁大学学报(自然科学版), 99(02):157-162.
- [4]陈春东, 程. 基于时间序列分析模型的公安警情分析和应用研究[J]. 安全技术与应用, 012(10):44-47.