

# 城市盗窃案件的时空一体化预测研究\*

——以城市两轮机动车失窃案件为例

朱瑞, 汤国安, 吴文博, 房彩申, 王丹妮

(南京师范大学地理科学学院, 南京 210046)

**摘要:** 城市盗窃犯罪同其他地理现象一样, 在一定的时间范围内有其特定的空间分布规律。本文首先通过某地的摩托车盗窃案件数据, 构建起相应的时间序列预测模型。其次, 将研究区域规则格网化, 综合与盗窃案件密切相关的地理因子, 对该类案件的不同因子进行打分, 构建出空间预测模型。然后, 将两个模型相融合, 得到了时空一体化的犯罪预测模型。

**关键词:** 犯罪预测; 时间序列模型; 空间预测模型; 时空一体化

## 1 引言

城市是人类文明发展到一定程度的产物。一方面, 城市是一定地域范围内的经济、政治和文化中心; 另一方面, 城市也是各类犯罪的滋生地和聚集地。改革开放后, 随着我国城市化进程的加快和社会转型期的到来, 城市犯罪迅速增长, 给社会的稳定和人民的生命、财产安全带来了巨大的危害。犯罪预测可以为相关部门提供多方面的决策支持, 有助于引导市民的日常行为, 进而有效地进行犯罪防控。国内外的地理学家和犯罪学家对于该课题都做了相关研究。刘小娟、韦立华等人从地理建模的角度, 各自进行了犯罪的计量和推测<sup>[1-2]</sup>。金光使用数据挖掘技术, 建立了犯罪风险预测模型<sup>[3]</sup>。李其富通过概率统计进行了相关预测<sup>[4]</sup>。然而, 他们的研究将时间和空间割裂开, 只单方面地关注时间预测或者空间预测, 这在实际应用中造成了一定的弊端。Shane D. Johnson 从犯罪制图(Crime Mapping)角度, 通过犯罪预测制图(Predictive Mapping of Crime), 进行了犯罪时空一体化预测的研究<sup>[5]</sup>, 取得了较好的效果。本文作者则从地理建模的角度出发, 通过某类城市犯罪数据, 创新性地将时间预测模型和空间预测模型相结合, 构建出了时空一体化预测模型, 并对该类案件进行了有效预测。

## 2 思路

### 2.1 时间预测模型的构建思路

对基于时间的案发统计数据进行正态性和稳定性检验。经检验, 当数据满足实验要求后, 通过线性回归模型去除统计曲线中的趋势, 得到具有周期性的统计曲线。而周期的意义, 对发掘案发时空规律, 指定相关决策, 有重要意义<sup>[6-8]</sup>。经过分析, 预测对象随时间变化呈现某种有规律的变化, 但这种变化既不符合常见的指数模型、灰色模型, 也不符合对数曲线模型。于是考虑使用多项式曲线外推法模型进行预测。首先构建概念模型, 加法模型, 用周期性叠加趋势模型描述:

$$Y_t = G_t + Z_t + E_t \quad (1)$$

收稿日期: 2009-01-05; 修订日期: \*\*\*

项目基金: 2008 年南京师范大学本科生创新实践训练项目(1812200013)。

作者简介: 朱瑞, 男, 在读本科生, 主要从事 GIS 空间分析研究。E-mail: 911zhurui@163.com  
汤国安(1961-), 通讯作者。

式中  $Y_t$  代表  $t$  月案发量观测值,  $G_t$  代表  $t$  月案发量观测值的趋势分量;  $Z_t$  代表  $t$  月案发量观测值的周期分量;  $E_t$  代表随机噪声; 该模型的一般形式为:

$$Y = b_0 + b_1 t^1 + b_2 t^2 \dots + b_n t^n \quad (2)$$

最后将回归模型的一次函数和趋势线拟合的多项式函数代数相加, 得到基于整个研究区域的时间预测模型。

## 2.2 空间预测模型的构建思路

借助美国控制论专家L. A. 查德(Zaden)于1965年首先提出的“模糊集合”的概念, 通过“模糊集合综合评价法”<sup>[9]</sup>进行空间建模。设对某一类对象(两轮机动车失窃)进行等级评定, 有 $P$ 个评价因素 $F_1, F_2, \dots, F_p$ , 每一个评价因素都有一个离散化的状态集, 记为 $V_1, V_2, \dots, V_p$ , 其中 $V_i = \{V_{i1}, V_{i2}, \dots, V_{in}\}$  ( $i=1, 2, \dots, p$ )。值得注意到是,  $V_{ij}$ 不一定是数量。评价等级, 就是要区分该该类案件在每个评价单位内发生的危险系数:

$$P = \sum_{i=1}^p F_i \quad (3)$$

表 3 案发模型因子

Tab.3 Model Factors of the Crime Cases

诱发发案的 地理因子	因子得分(一个评价单 位)
饭店( $F_1$ )	中档 3 分; 低档 1 分; 高档 1 分
超市( $F_2$ )	大中型 3 分; 小型 1 分
网吧( $F_3$ )	2 分
银行( $F_4$ )	1 分; ATM 机 0 分
医院( $F_5$ )	2 分; 分院 1 分
道路类型 ( $F_6$ )	高速主干道、无路 0 分; 步行街、商业区道路 1 分
距派出所距 离( $F_7$ )	所在评价单元-1 分

通过空间分析可以得出, 两轮机动车失窃案发点主要沿城市街道两侧分布, 且主要集中于城市主干道和商业集中区。经过专家分析, 与失窃案件紧密相关的商业集中区主要包括: 饭店、超市、网吧、银行、医院和银行等。这些城市建筑的功能, 决定了经常有大量车辆长时间停留, 且处于无人看管状态。

高档饭店周围的停车多为轿车, 而非两轮机动车, 且都有专门人员负责看管。低档的餐馆门口的停车多在消费者视线以内, 且就餐时间短。而对于中档饭店, 前去就餐的消费者多数持有两轮机动车, 饭店缺少对车辆的监管, 且就餐时间长。由此, 经过听取专家意见, 确立了两轮机动车失窃的模型因子得分, 如表 3 所示。

## 2.3 时空一体化预测模型的构建思路

时间序列预测模型是在案件数量上, 对整个研究区域的统计。它可以预测在单位时间内、整个研究区域中、某类犯罪的犯罪总数。而空间预测模型则是在地理空间上, 给出在某一段时间内、整个研究区域中、某类犯罪的空间分布情况。具体的表现形式就是第  $i$  行、第  $j$  列的格网内发案危险系数为  $P_{ij}$ , 那么它在整个研究区域的比重就是:

$$\alpha_{ij} = \frac{P_{ij}}{\sum_{j=1, i=1}^{15, 15} P_{ij}} \quad (4)$$

那么, 在第  $i$  行、第  $j$  列的评价单元内, 该类案件在第  $t$  月中可预测的案件发生数目为:

$$case\_number(t) = y(t) \times \alpha_{ij} \quad (5)$$

## 3 数据

本文使用某市某区2006和2007两年的盗窃类型数据进行建模。原始盗窃数据的类型为电瓶车失窃案件和摩托车失窃案件。为了更好地进行研究，根据案件发生的时空性质相同的理论，将电瓶车失窃案件数据和摩托车失窃案件数据进行合并，称为两轮机动车失窃(motor)，如表1所示。每个数据都有与之对应的相关属性：案发时间，案发地点。并根据市区地图，在ArcMap9.2中构建起如图1所示的，研究必须的基础地理底图数据和与犯罪密切相关的地理要素数据。

表1 时间序列模型统计数据

Tab.1 The Statistic Data for Modeling Via Time Series

	Jan	Feb	Mar	May	Apr	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2006	13	12	17	30	31	32	27	28	20	44	47	36
2007	25	20	33	28	20	14	43	50	54	39	29	31

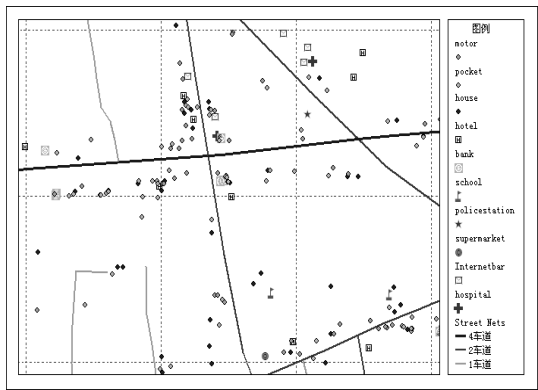


图1 空间预测模型基础地理数据

Fig.1 The Basic Geographic Data for Building  
Spatial Predictive Model

## 4 模型的构建

### 4.1 时间预测模型的建立

1. 对 motor 数据进行检验。根据图 2 和图 3 可知，该数据满足正态分布特征，来自同一总体。根据图 4 可知，样本序列数据的自相关数在某一固定水平线附近摆动，且按周期性逐渐衰减，所以该时间序列基本是平稳的。

2. 根据回归方程去除趋势。为了便于分析、保证去除趋势后的所有数据非负，使用公式  $Y = (Y_t + 22) - (0.843X + 20.783)$  进行趋势去除。

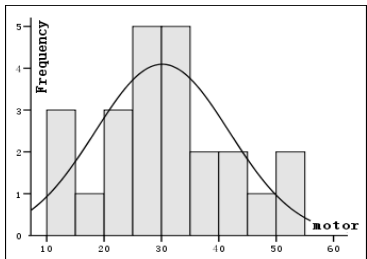


图2 数据正态性检验

Fig.2 The Check of Data's  
Std .Dev

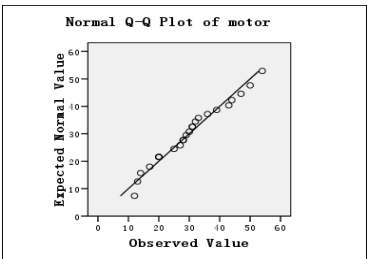


图3 研究数据 Q-Q 图

Fig.3 The Q-Q Map Based on the  
Research Data

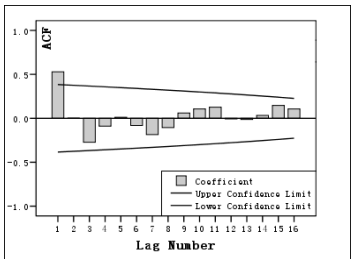


图4 数据稳定性检验

Fig.4 The Check of Data's  
Autocorrelations

3. 进行噪声去除和周期的调整。对一定的时间段内的数据进行移动平均处理，以减少孤立点和离群点，以突出总体趋势。简单移动平均的各元素的权重都相等，其计算公式如下：

$$F_t = (A_{t-1} + A_{t-2} + \dots + A_{t-n}) / n \quad (6)$$

式中  $F_t$  代表对下一期的预测值； $n$  代表移动平均的时期个数； $A_{t-1}$  代表前期实际值； $A_{t-2}$ ， $A_{t-3}$  和  $A_{t-n}$  分别表示前两期、前三期直至前  $n$  期的实际值。由于当年 12 月到次年 1 月该类案件案发机理相同，特征曲线连续，于是调整曲线的起始点。

4. 将数据分为两段，分别进行多项式拟合，建立趋势模型。得到当年 7 月到次年 1 月的

数据曲线(如图 7 所示), 以及当年 2 月到当年 6 月的数据曲线(如图 8 所示)。

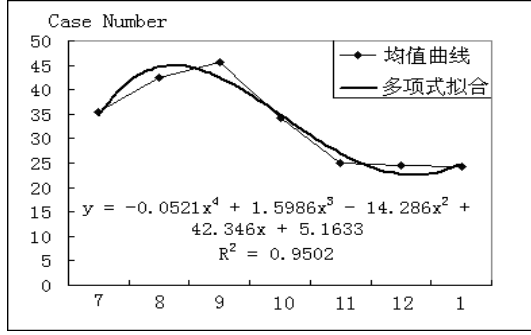


图7 7月到次年1月趋势拟合线

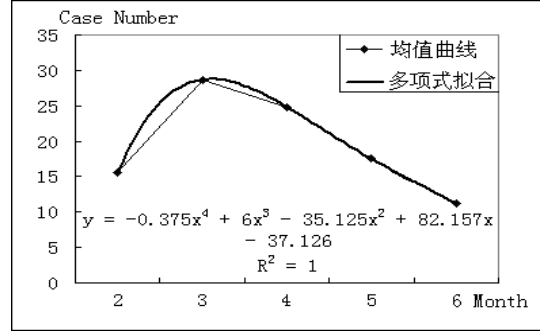


图8 2月到6月趋势拟合线

Fig.2 A Line Based on a Statistical Trend From Jul to Next Jan

Fig.2 A Line Based on a Statistical Trend From Feb to Jun

5. 进行反算, 构建预测模型。结合其物理意义, 可得分段函数:

①当  $T = 1 + 12k$  ( $k=1, 2, 3, \dots$ ) 时

$$Y_t = 25.4737 + 0.834X + 1.217 \quad (7)$$

③当  $2 \leq T \leq 6$ ;  $x$  为以“月”为单位的时间序列时

$$Y_t = -0.375(T-1)^4 + 6(T-1)^3 - 35.125(T-1)^2 + 82.157(T-1) + 0.834X - 38.343 \quad (8)$$

②当  $7 \leq T \leq 12$ ;  $x$  为以“月”为单位的时间序列时

$$Y_t = -0.0521(T-6)^4 + 1.5986(T-6)^3 - 14.286(T-6)^2 + 43.18(T-6) + 0.834X + 3.9436 \quad (9)$$

同时考虑到实际意义,  $X$  的取值为 1 到 12 的整数, 每个分段函数  $Y_t$  的值也必须取正整数。

## 4.2 空间预测模型的建立

1. 利用 ArcMap9.2, 将整个研究区域, 按东西向和南北向, 规则格网化为  $15 \times 15$ , 共 225 个矩形格网。每个格网作为一个评价单位。

2. 借助叠置分析的思想, 为每个评价因素建立一个图层<sup>[10-11]</sup>。每个图层的图幅与研究区完全重合, 分辨率与建立好的矩形格网匹配。并按照表 3 中对案发模型因子的打分规则, 计算出每个格网(评价单元)的  $N = \sum_{i=1}^p F_i$ 。计算出的结果如图 11 所示。

3. 根据如图 12 所示的数学统计, 将每个评价单元内的案发量分等级。每个评价单元内案发统计量为 0-3 的权重  $W(\text{weight})$  设为 0.1, 4-14 的权重设为 0.3, 14 起以上的权重设为 0.6。由此得到了如图 13 所示的各个评价单元的权重  $W_{ij}$ 。

4. 最后计算第  $i$  行、第  $j$  列的评价单元内案发危险系数:

$$P_{ij} = W_{ij} \times \sum_{i=1}^p F_i \quad (10)$$

需要说明的是, 计算出的  $P_{ij}$  是一个案发系数, 非是预测出的评价单元内案发的数目。

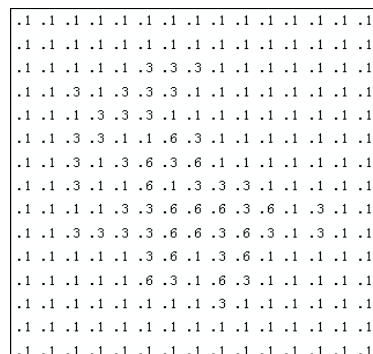


图 13 评价单元权重

Fig.13 Weight for Every Grid

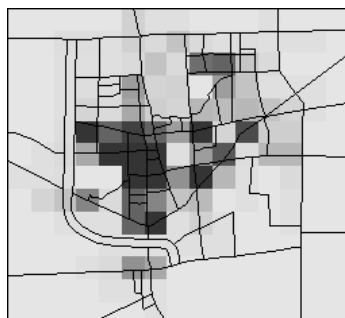


Fig.15 Crime Predicted Number

- [1] 刘小娟,高连生.灰色系统理论在犯罪动态预测中的应用.中国人民公安大学学报,2005,113:44-48.
- [2] 韦立华,朱德林.犯罪预测动态回归分析方法.江苏警官学院学报,2004,19(2):24-27.
- [3] 金光,钱家麒,钱江波,黄蔚民.基于数据挖掘决策树的犯罪风险预测模型.计算机工程.2003,29(9):183-185.
- [4] 李其富,最优组合预测方法及其在犯罪预测中的应用.四川省公安管理干部学院学报,1999:26-27.
- [5] Shane D. Johnson, Kate J. Bowers, Dan J. Birks, Ken Pease. Predictive Mapping of Crime by ProMap: Accuracy, Units of Analysis, and the Environmental Backcloth. Putting Crime in its Place:171-198.
- [6] 徐海量,叶茂,宋郁东.塔里木河流域水资源变化的周期性分析及预测.干旱区研究,2005,22(4):454-457.
- [7] 韦玉春,陈锁忠.地理建模原理与方法.北京:科学出版社,2005,218-230.
- [8] 张国忠,杨涛,聂文霞,叶霞,胡益民,陈宇峰.发电量季节周期性预测模型的探讨.湖北电力.2001,25(2):9-10.
- [9] 辜寄蓉,陈晓,彭东.九寨沟地质灾害预测的空间分析模型.中国地质,2002,29(1):109-112.
- [10] 陈丽霞,殷坤龙,张桂荣.浙江省磐安县滑坡灾害空间预测.地球与环境,2005,33:359-363.
- [11] 关国杰,米吉提阿布里米提,艾飞.层次分析法在新疆民丰县地质灾害空间预测中的应用.安全与环境工程,2008,15(2):81-85.
- [12] 张磊.城市犯罪的原因与控制.法制与社会.2008,2:203.
- [13] 王仲伸.未来五年崇文区犯罪增长趋势预测和行动对策.北京人民警察学院学报,2006,1:18-26.
- [14] 捍卫平.城市巡逻防控工作研究——以浙江省为例.浙江警官学院学报.2008,3:30-33.