

以故意伤害案件为例的犯罪预测教学探讨

Discussion on Crime Prediction Teaching in Intentional Injury Cases

曹霁

Ji Cao

南京森林警察学院基础部 中国·江苏南京 210000

Basic Department of Nanjing Forest Police College, Nanjing, Jiangsu, 210000, China

摘要: 分别使用移动平均法和灰色预测模型, 在 2010—2019 年公安机关受理故意伤害案件数等历史数据的基础上, 对接下来的犯罪行为进行预测。希望通过这部分的教学, 学生在步入工作岗位之后, 可以在单位实际数据的基础上, 预测未来犯罪走向, 合理分配警力, 从而更加有效地打击犯罪。

Abstract: We use moving average method and grey prediction model respectively to predict the following criminal behaviors based on historical data such as the number of intentional injury cases accepted by public security organs from 2010 to 2019. We hope that through this part of the teaching, students can predict the trend of future crimes based on the actual data of the unit and rationally allocate police force, so as to fight crime more effectively after entering the job.

关键词: 时间序列分析; 移动平均法; 灰色预测; 故意伤害案件

Keywords: time series analysis; moving average method; gray prediction; intentional injury cases

基金项目: 江苏高校哲学社会科学研究项目“基于时间序列分析的公安数据短期预测模型研究”(项目编号: 2021SJA0565)。

DOI: 10.12346/sde.v3i11.4776

1 引言

随着信息技术的快速发展, 数据作为国家基础性战略资源的地位已经成为共识。当我校学生迈出校园, 步入警务实践中的时候, 会接触到大量的犯罪情报数据和犯罪信息。但若信息不能得到有效的处理和分析, 呈现出来的大多是较为杂乱无序的, 对接下来的警务工作的指导意义并没有得到完全的体现。因此, 我们接下来在教学中将加入一些利用统计方法来深入分析犯罪规律, 从而对犯罪行为的发展趋势进行预测的内容^[1]。通过这些数据处理的结果分析, 结合犯罪学等相关理论, 为警务实践中的犯罪防控工作奠定基础。

2 背景

2.1 研究背景

故意伤害罪, 是指针对故意非法损害他人身体健康的行

为而定的罪, 是我国刑法中侵犯公民人身权利、民主权利罪的最常见的一种犯罪。我们在中国统计年鉴、中国社会统计年鉴中可以查到从 2010—2019 年公安机关受理故意伤害案件数的相关历史数据, 然后从数据出发, 选择合适的数据处理方式, 对接下来的犯罪行为进行预测。

2.2 教学背景

我们将要面对的学生已经简单学习了《概率论与数理统计》课程, 但由于课时限制, 课程集中在概率论部分, 后续的数理统计部分几乎没有讲解, 所以我们选择的预测方式一定要适合学生基础, 要易于理解, 便于操作。每年我们将有一部分学生参与到全国大学生数学建模竞赛中去, 对数据处理有一定的基础, 并会使用简单的 MATLAB 软件操作。考虑到这一点, 在数据处理过程我们也选择了 MATLAB 软件。

【作者简介】曹霁 (1979-), 男, 中国江苏南京人, 硕士, 讲师, 从事应用数学研究。

3 移动平均法

移动平均法是一类确定性时间序列分析方法^[2]。所谓的时间序列,就是将预测的对象按照时间的顺序排列起来。时间序列预测模型就是在数据发展有延续性的基础上,观察、研究时间序列,对时间序列的未来的发展趋势进行预测的统计分析模型。当预测目标的发展趋势变化不大时,可以采用移动平均法来做简单的近期预测。由于学生学习过数学期望和方差,对于这种时间序列预测方法接受起来也更为容易一些。

设观察序列为 Y_1, Y_2, \dots, Y_{10} , 取移动平均项数 $N < 10$ 。可用一次移动平均法建立预测模型如下:

$$\hat{y}_{t+1} = \frac{y_t + y_{t-1} + \dots + y_{t-N+1}}{N}, t = N, N+1, \dots, T$$

预测的标准误差为:

$$S = \sqrt{\frac{\sum_{t=N+1}^T (\hat{y}_t - y_t)^2}{T - N}}$$

通过预测公式我们可以发现,所谓移动平均法,也就是将最近的 N 个序列的平均值来作为预测值。随着时间间隔的增长,历史数据对未来值的参考性会减小,因此 N 值并不是越大越好,最佳的 N 值是使预测标准误差最小的值。

以 2010—2019 年公安机关受理故意伤害案件数带入 MATLAB 程序如下:

```

clc,clear
y=[237471, 278482, 306328, 294449, 229277,
216716, 220193, 204054, 176387, 156090];
m=length(y);
n=1: 9;
for i=1: length(n)
yhat{i} = zeros(1, m);
yhat{i}(1: n(i)) = y(1: n(i));
for j=n(i)+1: m
yhat{i}(j)=sum(y(j-n(i): j-1))/n(i);
end
s(i)=sqrt(mean((y(i+1: end)-yhat{i}(i+1: end)).^2));
end
s
    
```

计算结果显示,当 $n=1$ 时,预测的标准误差较小,所以选择 $n=1$ 。此时预测 2020 年公安机关受理故意伤害案件数为 156090。

4 灰色预测

除了移动平均法之外,灰色预测模型也是一种常见的预测模型。它的特点是不直接使用原始数据序列,例如最常用的 GM(1, 1) 模型,就是将原始数据累加得到新的数据序列再进行数据建模^[3]。它的优点在于对数据量的要求不高,并且可以直接套用程序框架。例如以 2010—2019 年公安机关受理故意伤害案件数如表 1 所示。

以此为基础预测 2020—2024 年的数据。步骤如下:

①建立时间序列:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(10)) = (237471, 278482, 306328, 294449, 229277, 216716, 220193, 204054, 176387, 156090)$$

原始数据累加得到新数据序列

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(10)) = (237471, 515953, 822281, 1116730, 1346007, 1562723, 1782916, 1986970, 2163357, 2319447)$$

②构造数据矩阵 B 及数据向量 Y,

$$B = \begin{bmatrix} -0.5(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -0.5(x^{(1)}(9) + x^{(1)}(10)) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(10) \end{bmatrix}$$

$$\text{求解灰参数 } \hat{u} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} 0.0756 \\ 336981 \end{bmatrix}, \text{将灰参}$$

数带入, $\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{\hat{b}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}$, 对 $\hat{x}^{(1)}(k+1)$ 及 $\hat{x}^{(1)}(k+1)$ 作差还原 $x^{(0)}(k+1)$ 序列, 得到原始数据的预测序列

$$\hat{x}^{(0)} = (237471, 307262, 284886, 264140, 244904, 227069, 210533, 195201, 180986, 167806)$$

③模型检验。预测的均方差比 $C=0.365$, 小误差概率 $p=1$, 预测的精度较好。

④利用模型进行预测 2020—2024 年的数据如表 2 所示。

2010—2029 年实际值与预测值之间的差距如图 1 所示。

表 1 2010—2019 年公安机关受理故意伤害案件数表

年份	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
案件数	237471	278482	306328	294449	229277	216716	220193	204054	176387	156090

表 2 利用模型进行预测的 2020—2024 年的数据

年份	2020	2021	2022	2023	2024
预测值	155585	144255	133750	124010	114979

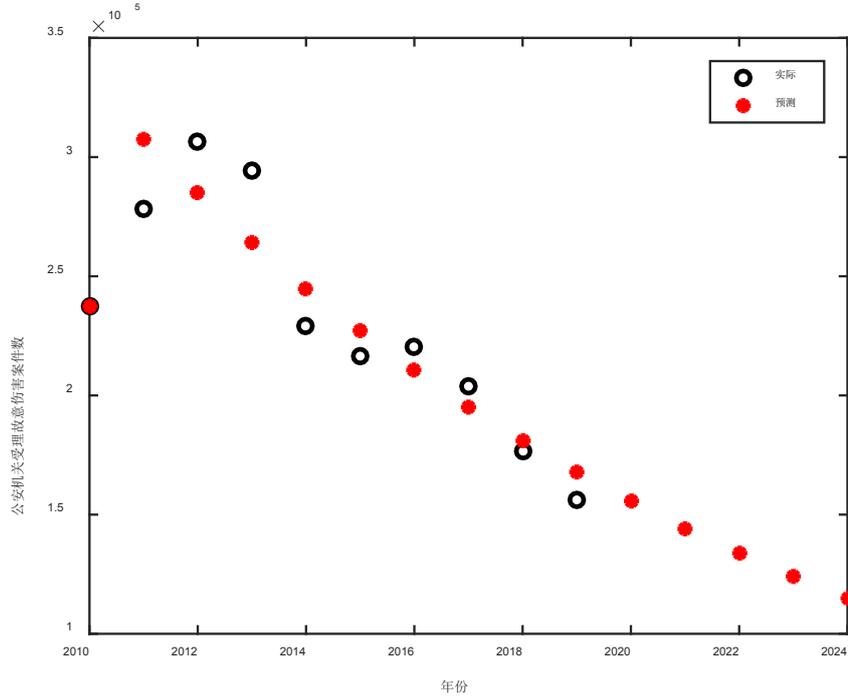


图 1 2010—2029 年实际值与预测值之间的差距

MATLAB 程序如下:

```
clear;clc;
x_0=[237471,278482,306328,294449,229277,
216716,220193,204054,176387,156090];
x_1=cumsum(x_0);
n=length(x_0); for i=1:(n-1)
C(i)=(x_1(i)+x_1(i+1))/2;
end
Y_n = x_0(2:end);
B=[-C;ones(1,n-1)]';
GreyParameter=inv(B'*B)*B'*Y_n;
a=GreyParameter(1);u=GreyParameter(2);
F=[];F(1)=x_0(1);
for i=2:(n+5)
F(i)=(x_0(1)-u/a)/exp(a*(i-1))+u/a ;
end
G=[];G(1)=x_0(1);
for i=2:(n+5)
G(i)=F(i)-F(i-1);
end
t1=2010:2019;
t2=2010:2024;
G
```

```
plot(t1,x_0,'ko', 'LineWidth',2)
hold on
plot(t2,G,'r*', 'LineWidth',2)
xlabel(' 年份 ', 'fontsize',12)
ylabel(' 公安机关受理故意伤害案件数 '), 'fontsize',12)
set(gca, 'LineWidth',1);
legend(' 实际 ', '预测 ');
e = x_0-G(1:n);
```

5 结语

上述两种预测方式中, 移动平均法具有易于理解的特点; 灰色预测对数据量的要求不高, 并且实操性强。我们希望学生可以扎实掌握这部分内容, 在步入警务实践工作后, 在做好数据保密工作的基础上, 能够将这两种方法学以致用。通过这两种数据处理方法, 能够更加深入地挖掘出数据背后代表的意义, 为警务工作提供新思路。

参考文献

- [1] 陈笛. 盗窃类犯罪的时间序列分析[D]. 北京: 中国人民公安大学, 2017.
- [2] 刘小娟, 高连生. 灰色系统理论在犯罪动态预测中的应用[J]. 中国人民公安大学学报(社会科学版), 2005(1): 44-48.
- [3] 马翠玲, 司守奎, 李彪, 等. 融入数学史, 借助MATLAB实现中心极限定理形象化教学[J]. 大学数学, 2014, 30(S1): 115-118.