A Method to Improve the Rate of Case Investigation Based on Multiple Linear Regression Model

Zhongfeng Guo[1], Jianhui Lv[2]

[1,2]College of Information Science and Engineering of Northeastern University, Shenyang 110819, China.

E-mail: [1]guo.zhf@neusoft.com; [2]lvjianhui2012@163.com

Abstract:

In order to analyze the influence of people's behavior attribute to crime, this paper uses multiple linear regression model, and data fit of the population and the number of criminal cases, the population and the years. Construct iteration model, and establish the relationship between population increase and key attributes. Experiment show when the population changes, key behavior property which affects the crime rates needs to readjust. This kind of mathematical model based on multiple linear regressions can be well used in case investigation, and achieves good results.

Keywords: linear regression, case investigation, behavior attribute, data fitting

1. Introduction

In recent years, with the improvement of material civilization and spiritual civilization. Our country is becoming more and more stable and harmonious. But there are still a lot of cases should be solves in every day. In order to detect more cases, scientific methods should be used to help the police. In our country, criminal offence involves criminal law and criminology. Because of the different discipline nature and the different research purpose, academic circles studies of crime have different perspectives and working focuses. They don't make overall analysis about all criminal behaviors, and most of their researches are only in the experimental stage. It does not have any practical significance for the police to solve the case. According to the statistics of Professor Bai Jianjun, the results show that the causes of crime theory have been more than 130. So it is difficult for us to conduct a comprehensive analysis on the causes of crime. And it is also difficult for us to narrow down the suspects' crowd to raise the work efficiency.

Through analyzing all the behavioral attributes of the population and abstracting the main behavior attribute data model to determine the weight of key attributes, it is an important direction of our study. This paper gives a detail scheme to raise the efficiency of handling cases.

2. Population Increase Based on Multiple Linear Regression

Human behavior is a key factor in investigating case[1-4]. Although the reference value of some behavioral attributes is very little, they also affect the case detection rate to a certain extent. In order to determine which behavior is the key attribute and the importance degree of the key attribute. this paper involves 8 kinds of behavior attribute, namely: natural characteristics,family background, education level, regional limits, age, income, criminal record, bad habits, which are marked as B1,B2,B3,B4,B5,B6,B7 and B8.

In these behavioral attributes, most properties and population have hidden relationships. For example: as the population increase, the pressure of competition increase, people's income level will change; at the same time, the regional limits immigration and emigration also affect the population increase. 8 kinds of factor have different effects to crime rates, and the difference lies in the different weights. In order to explore the effects of population growth on 8 kinds of behavioral properties, constructing the mathematical model of linear regression [5], as is shown in equation

$$\begin{cases} f(x_1,\ldots,x_8;\beta_1,\ldots,\beta_8;u) \\ = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_8 x_8 + u_t \\ E(u) = 0 \\ \sigma^2 = u_t \end{cases} \quad (1)$$

Among it U represents a set of all the investigation subjects. β_i represents influence coefficient of behavior attribute to crime rate, x_i represents the index of behavior attribute and $f(x_1,\ldots,x_8;\beta_1,\ldots,\beta_8;u)$ represents crime rate. The relationship of behavioral attributes is not easy to observe, but it has close relation with the increase and decrease of population. Real-time population plays the bridge role between two of them. We can get the abstract formula relying on the relationship between population and crime rate.

$$\begin{aligned} & f(x_1,\ldots,x_8;\beta_1,\ldots,\beta_8;u) \\ & = N'(x_1,\ldots,x_8;\beta_1,\ldots,\beta_8;u) \\ & = N_{\text{abstract}} \end{aligned} \quad (2)$$

Among it, $N_{abstract}$ is a function of N. This paper use Liaoning province as an example for modeling, setting up 22 years' population increase of crime rate model. Because 22 years is a short time , Malthusian population model can be applied to forecast. Setting r as the change rate of population, it has relationships with the local population N and the time t, which is shown as formula (3) :

$$\begin{cases} \dfrac{dN}{dt} = rN(t) \\ N(0) = N_0 \end{cases} \quad (3)$$

N_0 is the number of population in Liaoning province in 1990.The index corresponding to 8 kinds of behavioral attributes has a logical mathematical relationship with time t .In order to illustrate the problem, and set the hypothetical life as 250 years life; this paper mainly lies on the Logistic model [6]:

$$\begin{cases} x = c_1 / (1 + \exp(c_2 + c_3 t)) \\ x_1 = x_1(t) \\ x_2 = x_2(t) \\ \quad \vdots \\ x_8 = x_8(t) \end{cases} \quad (4)$$

$x_i(t)$ is the relationship between the index of behavioral attributes and time. The formula (1) -- (4) are substituted into the transformation formula, and we can obtained multiple linear regression model:

$$\begin{cases} (\beta_1,\beta_2,\ldots,\beta_8)\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_8(t) \end{pmatrix} = [N_{\text{abstract}}]^2 - \sigma^2(t) \\ \\ \beta_1 + \beta_2 + \ldots + \beta_8 = 1 \end{cases} \quad (5)$$

Do not consider the change of population, that is to say, at the moment t ,put $(\beta_1,\beta_2,\ldots,\beta_8) = (W(1),W(2),\ldots,W(8))$ into the model. We can calculate the value of σ^2 ,$W(j)$ shows the total weight of behavioral attributes. Put it into model (5), $x_1(t), x_2(t),\ldots,x_8(t)$ can be obtained by data fitting.Its only task is to get the value of $\beta_1,\beta_2,\ldots,\beta_8$.

3. Behavioral Attributes Fitting -- Taking Liaoning Province as an Example

This part mainly aims at solving the linear regression model according to 8 kinds of behavioral attributes. Taking Liaoning Province as an example. The data is from the network resources and the Chinese criminal database, and the reference time is from 1990 to 2011.

3.1. Population Quantity (Million) and Time (Year) Fitting

Deformation formulas (3):

$$\begin{cases} N(t) = N_0 e^{r(t-t_0)} \\ N_0 = 4194 \end{cases} \quad (6)$$

The data fit chart as in Fig1.

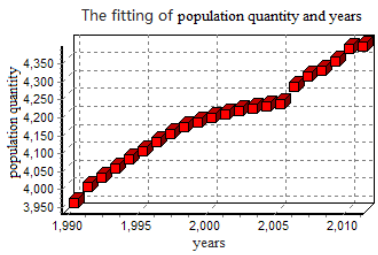

Fig1.population quantity and years fitting

Through the analysis of Fig1, it is easy to know that $r = 2‰ > 0$. And the population of Liaoning Province shows increasing trend from 1990 to 2011.

The concrete model as follows: equation (7)

$$N(t) = N_0 e^{2‰t} \qquad (7)$$

3. 2. Fitting abstract model

According to the relative data of population quantity and the number of case. The data fit chart as in Fig2

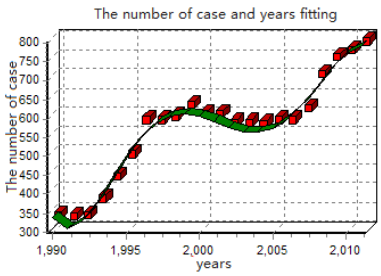

Fig2.the number of case and years fitting

Set the highest order as five, and get the polynomial model as follows (8).

$$\begin{aligned} & f(x_1,\ldots,x_8;\beta_1,\ldots,\beta_8;u) \\ & = 38151.215 - 655.9N \\ & +2.707N^2 - 0.0042N^3 \\ & +2.2679e^{-6}N^4 + 1.1271e^{-11}N^5 \end{aligned} \qquad (8)$$

3. 3. The Final Data Model

The Logistic model, in which only c_1 has a different value, as shown in table

Table1.Logistic model's Parameter of behavior attribute

attribute	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8
parameter c_1	1.1	4.4	6.6	1.1	4.4	5.5	1.1	8.8
$c_2 = 0.1, c_3 = 0.008$								

With those parameters we get the function of Logistic model as follows:

$$x(t,i) = \frac{c_{1,i}}{1+e^{0.1+0.008t}} \qquad (9)$$

With all the equations on the above we get the final data model as follows:

$$\begin{cases} (\beta_1,\beta_2,\ldots,\beta_8)\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_8(t) \end{pmatrix} = [N_{\text{abstract}}]^2 - \sigma^2(t) & (1) \\ x(t,i) = \dfrac{c_{1,i}}{1+e^{0.1+0.008t}} & (2) \\ N_{\text{abstract}} = k_0 + k_1 N + k_2 N^2 +,\ldots,k_5 N^5 & (3) \\ N(t) = N_0 e^{r(t-t_0)} & (4) \\ \beta_1 + \beta_2 + \ldots + \beta_8 = 1 & (5) \end{cases} \qquad (10)$$

4. Model Solution

The concrete steps of data model solving as follows:

a)Put $(\beta_1,\beta_2,\ldots,\beta_8) = (W(1),W(2),\ldots,W(8))$, $N_0 = 4194$, $r = 2‰$, $t_0 = 0$ into formula (10.1). From 1990 to 2011, and get 22 corresponding values of $\sigma^2(t)$.

b).put results of formula (9) and a) into formula (10.1). After that, 22 new equations about $\beta_1,\beta_2,\ldots,\beta_8$ will be created, with $\beta_1+\beta_2+,\ldots,+\beta_8=1$ .There are 23 equations in total, and only one equation is an identity, which is $\beta_1+\beta_2+,\ldots,+\beta_8=1$. The others are recursive equations. In order to ensure the data of $\beta_1,\beta_2,\ldots,\beta_8$ has higher adaptability; Many times equation solution should be done.

c).Dividing the 22 recursive equations into three groups. Every group includes seven or eight equations. Each group combines with $\beta_1+\beta_2+,\ldots,+\beta_8=1$ .Structure three new equations, and get the value of $\beta_1,\beta_2,\ldots,\beta_8$ from three equations.

d).Get the average value of the three group of $\beta_1,\beta_2,\ldots,\beta_8$ .At last, every behavior attribute's weight will be calculated.

e).Reordering the weight of 8 behavior attribute and get the final result.

Implement the data model, get three groups of values of $\beta_1,\beta_2,\ldots,\beta_8$. As Table2 shows:

Table2. Values of $\beta_1,\beta_2,\ldots,\beta_8$

item	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
first group	0.0676	0.0901	0.0691	0.2146	0.0767	0.1504	0.1989	0.1327
second group	0.0774	0.1004	0.0749	0.2211	0.0801	0.1420	0.1836	0.1208
third group	0.0773	0.0991	0.0741	0.2105	0.0799	0.1590	0.1901	0.1102
average	0.0741	0.0965	0.0727	0.2154	0.0789	0.1505	0.1908	0.1212

The total weight of each behavior attribute in descending order is as follows:

$B_4 \to B_7 \to B_6 \to B_8 \to B_2 \to B_5 \to B_1 \to B_3$

As the result shows, it is easy to know that B_4, B_7, B_6, B_8 are the key attributes among 8 kinds of behavior attribute .That is to say regional limits, criminal record, income and bad habits are very important attributes when the police detect cases.

5. Conclusions

By the solution of this model, it is clear to see that when the population change is considered, the linear regression model could co-ordinate 8 kinds of behavioral attributes into the model which is concerned about the number of crime cases. It is much better to show the effect of various factors on the crime rate and has strong reliability to determine the key attribute.This model improves the efficiency of handling cases. In order to get a better model, the next step of the study is to accelerate convergences' speed of the algorithm by use of rough set.

6. References

[1] Guixin Wang, Yiyun Liu. Analysis of the Characters and Reasons of the Crime Committed by Floating Population in Shanghai [J].Population Journal, 2006, (3):44-48.

[2] Gang Chen, Shu Li, Yili Chen.Study on the Effect on the Crime Rate by the Floating Population[J].Chinese Journal of Population Science. 2009, (4): 52-60.

[3] Xiaohu Zhang. Social Stratification Analysis of the Crime Rate Increasing during the period of Social Transformation [J].2002, (1):91-106.

[4] Xi Sun. On the Behavioral Attributes of Possessory Offences[J]. Legal System and Society, 2007, (3):91-104.

[5] Jinjing Xie. The Comprehensive Model of Exchange Rate Determination with Multi-factors--An Example of Empirical Analysis Based on USD and CAD[D]. Shanxi: Shanxi University of Finance &Economics, 2010.

[6] Xiaozhu Jian. The Fit of C and γ Parameters within Logistic Model (in Chinese). Master's dissertation. Jiangxi Normal University, 2006