http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (15): 3374-3377, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3374.3377

Significance Test Algorithm of Crowd Flow in Public Fitness Areas Based On Regression Analysis

Guanghong Liu
Institute of Physical Education, Shenyang Sport University, Shenyang 110102, China

Abstract: The increase of crowd come in and go out of fitness places can reflects the increase of the number of fitness people from the side. Fitness places often have installed camera, their daily video recording can be used as raw data of health situation in the area. In order to better statistical the number of people in fitness places with high density population, research means on the video should develop toward a more accurate goal that is easier to achieve. In the places with higher population density, target occlusion problem among each other is more prominent, which makes it difficult to detect and trace independent entity in a crowded area and the difficulty to precisely acquire the body's movement trajectory is strengthened. On the basis of studying the characteristics of the video study object (crowd flow), this study establishes a linear regression model to estimate the population flows. The study first introduces the principle of video motion segmentation and the extraction method of eight categories of image features and then discusses the principles of regression estimation and significance test approach, finally verifies the reasonableness of theoretical models in the text by the data, which provides a theoretical basis for video analysis and provides a better technical foundation for the regional public fitness study.

Key words: Crowd flow, regression estimation, significance test

INTRODUCTION

With economic development and social progress, the people have gradually pay attention to health problems in the material life. As we all know, the public health as a sport objectively needs some places, equipment and other economic inputs. These objective conditions become a necessary condition for the sport (Liu, 2006). And the displayed data from these objective conditions can reflects the overall situation of the public health from the side (Zhang and Yue, 2013). The study selects the video produced in public fitness places as the study object, in order to get the crowd flow data by analyzing information in the video record and reflect the health status of population fitness in this area (Yan, 2008). Video analysis technology and artificial intelligence technology are in constant development; intelligent video monitoring technology has become a very broad study field (Hu, 2007). Currently, intelligent monitoring studies are focused on a small number of individuals and the analysis of population movement are few (Xiang, 2007); but with social development and the demands of public safety, the analysis of population movement have gradually got the attention of scholars and experts (Wang, 2005); crowd flow estimation studies for VCR video in public fitness places are less (Xiao, 2010).

In this study, on the basis of previous studies, it uses the feature extraction theory of video object, does a schematic analysis for the regression model of data estimation, reaches corresponding conclusions through data validation and regression analysis, in order to provide a more convenient way for the regional health statistics analysis through the discussion in this article.

Regression prediction estimation model: Series of different numbers of people and the corresponding feature numbers of different numbers of people are known, we need to establish an appropriate model to estimate the corresponding unknown number of people to the known number of features and regression estimation is a classic learning method.

Regression estimation mathematical model: Regression estimation is mathematical models based on a function model of additive noise (Hou, 2005) and supposes one unknown function has the parametric form as shown in the Eq. 1:

$$f_0(x) = f(x, a_0) \tag{1}$$

In Eq. 1 $a_0 \in \Lambda$ is the unknown parameter vector; in addition we assume that at any point of x_i, the noise values of this function can be measured, as shown in Eq. 2:

$$y_i = f(x_i, a_0) + \varepsilon_i \tag{2}$$

In Eq. 2 noise ε_i and x_i are mutually independent and its distribution obey density function p(ε); through the observational data and a series of two-dimensional points, the parameter α_0 can be estimated using certain parameter estimation method (Feng, 2007); the maximum functional produced by maximum likelihood estimation is shown in Eq. 3:

$$L(a)=\sum_{i=1}^{1}\ln p\left(y_i-f(x_i,a)\right) \quad (3)$$

In Eq. 3 p(ε) is a determined function; if we use a normal distribution with zero mean and a fixed variance as the noise model, its expression is in Eq. 4:

$$p(\varepsilon)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) \quad (4)$$

Substitute the Eq. 4 into the Eq. 3, then the form of Eq. 5 can be obtained:

$$L^*(a)=-\frac{1}{2\sigma^2}\sum_{i=1}^{1}\left(y_i-f(x_i,a)\right)^2-1\ln\left(\sigma\sqrt{2\pi}\right) \quad (5)$$

The parameters in Eq. 5 can be obtained by the least square method.

Linear regression models: Usually the Eq. 6 is called Gaussian-Markov linear model, denoted as $(Y, X\beta, \sigma^2 I_n)$:

$$\begin{cases} Y=X\beta+\varepsilon \\ E(\varepsilon)=0, COV(\varepsilon,\varepsilon)=\sigma^2 I_n \end{cases} \quad (6)$$

The meaning of the variables in Eq. 6 is shown in Eq. 7:

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \beta=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (7)$$

When k = 1 in the Eq., Equation 6 is a linear regression equation; when k>1, Eq. 6 is the regression plane equation; use the least square method to strike the estimated value of each element in vector β, the residual sum of squares is shown as Eq. 8 below:

$$Q=\sum_{i=1}^{n}\left(y_i-\beta_0-\beta_1 x_{i1}-\cdots-\beta_k x_{ik}\right) \quad (8)$$

When the Q in Eq. 8 takes the minimum value, substitute the corresponding $\hat{\beta}_i$ into the regression plane equation as shown in the Eq. 9:

$$y=\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_k x_k \quad (9)$$

Equation 9 is the required regression equation.

Significance test: As for the significance test of the obtained regression equation:

$$y=\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots+\hat{\beta}_k x_k$$

We take k = 1 for an example, assume:

$$H_0:\beta_1=0 \qquad H_1:\beta_1\neq 0$$

And conduct tests; the text uses test value F to conduct test; when H_0 sets up, test value F is shown in Eq. 10:

$$F=\frac{U(n-2)}{Q_e}\sim F(1,n-2) \quad (10)$$

The U, Q_e in Eq. 10 is shown in the Eq. 11:

$$\begin{cases} U=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \\ Q_e=Q\left(\hat{\beta}_0,\hat{\beta}_1\right)=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \end{cases} \quad (11)$$

The corresponding probability distribution to the F value is the F distribution and the probability is p; so when $F>F_{1-\alpha}(1,n-2)$, we have p>a, then reject H_0, otherwise accept it. This study takes the 95% confidence coefficient, so we take α = 0.05; the 95% confidence interval of regression coefficients β_0, β_1 is expressed as the tdistribution shown in Eq. 12:

$$\begin{cases} \left[\hat{\beta}_0-t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}_e\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{L_{xx}}}, \hat{\beta}_0+t_{1-\frac{\alpha}{2}}(n-2)\hat{\sigma}_e\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{L_{xx}}}\right] \\ \left[\hat{\beta}_1-t_{1-\frac{1}{2}}(n-2)\frac{\hat{\sigma}_e}{\sqrt{L_{xx}}}, \hat{\beta}_1-t_{1-\frac{\alpha}{2}}(n-2)\frac{\hat{\sigma}_e}{\sqrt{L_{xx}}}\right] \end{cases} \quad (12)$$

In Eq. 12, $\hat{\sigma}_e$ means residual standard deviation; the variable description and calculation methods are shown in Eq. 13:

$$\hat{\sigma}_e^2=\frac{Q_e}{n-2}, L_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad (13)$$

Stepwise regression analysis method can produce optimal regression equation and the basic idea has the following four points:

- Start from an independent variable, depending on the effect significant degree of the independent variables X, individually introduces the regression equation in descending order
- When an introduced variable becomes insignificant since the behind variable is introduced, you need to get rid of the former variable
- The introduction of a independent variable or rejection of an independent variable from the regression equation, is called one step of stepwise regression
- For each step the F-value test must be conducted to ensure that it contains only variable that has significant effect on Y before introducing a new significant variable

DATA RELATIONSHIPS AND RESEARCH RESULTS

Relationship between side length of the square and the number of boxes: For certain image, use a series of squares to cover it, which can draw a list of N values and the quantitative relationship between R and N is shown in Table 1.

The relationship trend between the data R and N in Table 1 is shown in Fig. 1.

The corresponding statistics of box numbers for the population images of different densities is shown in Table 2.

Research results of regression estimation: For the independent regression prediction of eight types of features, axis x represents the value of each feature (Liu, 2008); axis y indicates the number of people, the obtained regression equation, confidence interval, F-value, the error value and the MSE of each feature, as shown in Table 3.

The regression equation in Table 3 is shown below.

$$y = 3.6152 + 0.0049x_1 \quad (22)$$

$$y = -1.5516 + 0.0137x_2 \quad (23)$$

$$y = -2.4755 + 0.0324x_3 \quad (24)$$

$$y = 1.5346 + 1.8048x_4 \quad (25)$$

$$y = -2.2386 + 0.0283x_5 \quad (26)$$

$$y = 13.7853 + 1.0982x_6 \quad (27)$$

$$y = 4.4233 + 0.0210x_7 \quad (28)$$

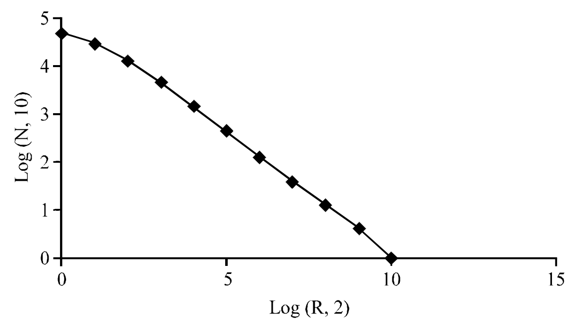

Fig. 1: Data change trend of $\log_{10}$ (N) with $\log_2$ (R)

Table 1: Relationship between side length and the number of boxes

R	N
2^0	44000
2^1	27466
2^2	11786
2^3	4265
2^4	1386
2^5	421
2^6	121
2^7	37
2^8	12
2^9	4
2^{10}	1

Table 2: Corresponding statistics of box numbers for the population images of different densities

No. (people)	Box dimension	Variance
12	1.1861	0.49380
40	1.2750	0.25260
45	1.4662	0.18024

Table 3: Independent regression prediction of eight types of features

No.	Regression equation	b_0 confidence interval	b_1 b_0 confidence interval	F-value	Error	MSE
x_1	Eq. 22	(3.1062,4.2041)	(0.0048,0.005)	6064.9	4.0188	5.0308
x_2	Eq. 23	(-2.235,-0.8682)	(0.0134,0.014)	6638.4	3.5077	4.3773
x_3	Eq. 24	(-4.0462,-2.066)	(0.0316,0.0338)	3504.4	3.8513	4.8934
x_4	Eq. 25	(-0.5419,2.8928)	(1.7022,1.9524)	822.2	5.7009	6.8458
x_5	Eq. 26	(-3.0494,-1.728)	(0.0277,0.029)	6520.7	4.1350	4.9890
x_6	Eq. 27	(13.108,14.9678)	(0.9958,1.1781)	772.0	4.5761	5.9618
x_7	Eq. 28	(4.0141,4.8954)	(0.0206,0.0214)	10354.6	3.2658	4.1490
x_8	Eq. 29	(-56.69,-44.919)	(53.8154,62.804)	648.6	4.7515	6.0593

$$y = -50.5410 + 58.1041x_8 \quad (29)$$

CONCLUSION

This study describes feature extraction method and linear regression estimation methods of the research object during the video analysis process and uses data to verify the accuracy of the model's algorithm. Based on the characteristics of study object in the video, it establishes a linear regression model to estimate the population flows, discusses the principles of regression estimation and significance test method, finally verifies the reasonableness of the theoretical models by the data, which provides a theoretical basis for video analysis and provides a better technical foundation for the study of regional public fitness condition.

REFERENCES

Feng, L., 2007. Retrieval of human motion data based on energy model. J. Comput. Aided Des. Comput. Graph., 19: 1015-1021.

Hou, Z.Q., 2005. A background reconstruction algorithm based on pixel intensity classification. J. Software, 16: 1568-1576.

Hu, H.M., 2007. An interval estimator of error variance in nonparametric regression model. J. China Univ. Petrol., 31: 162-167.

Liu, M.S., 2006. Video object extraction based on motion connectivity. Comput. Eng., 32: 226-228.

Liu, X.R., J.L. Zhou and X.H. Li, 2008. Grey level dependent matrix based crowd density estimation in fourier spectrum. Contr. Automation, 24: 310-314.

Wang, E.D., 2005. Estimate technology study of crowd movement and density. Safety, 3: 21-24.

Xiang, J., 2007. Human motion retrieval with time-sequence features based on data-driven decision tree analysis. J. Image and Graph., 12: 1761-1765.

Xiao, B.X., 2010. Survey on human mobcap data feature extraction and retrieval. Appli. Res. Comput., 27: 10-17.

Yan, X.J., 2008. Object tracking method based on block motion estimation. J. Image Graph., 13: 1869-1876.

Zhang, B. and H. Yue, 2013. Bio-mechanical mathematical model analysis for race walking technique. Int. J. Applied Math. Stat., 44: 469-476.