Use Information Standards methods to select the best linear regression model of digital image

Fathala A.Chachan
Administration and Economics, an nl-mustansiriya University

aiiiii_f@yahoo.com

Abstract. In this paper , several models of digital linear regression were proposed to test the effect of digital treatments on the image, and then choose the relatively better model among the appropriate models and test their quality by adopting special criteria (Akaike Information criteria, Bayesian information criteria BIC, Hannan – Quinn criteria HQC) In this research it was found that the image was processed and underwent its work (rotate, change its background, minimize it, and improve its contrast ) it has better quality compared to other pictures for having lowest value (AIC,BIC,HQC)

Keywords: simple linear regression, Akaike Information criteria, Bayesian information criteria, Hannan – Quinn criteria.

1. INTRODUCTION

the plan of computerized imaging frameworks and the advancement of programs since the sixties of the final century, where computerized pictures are characterized by a more noteworthy run of grayscale as well as contain data within the frame of numbers, which encouraged the investigation of models and scientific strategies, as well as the ease of managing with computerized information from Mail, which empowers analysts who need to work on the same information to access it in a simple and simple as simple to store, make strides, analyze, show, categorize, classify and recognize them physically and electronically and get as much data as conceivable. Of a number of little squares stacked nearby each other issue with a framework comprising of columns and columns and each square of these boxes speaks to what is known as an component or sharpness of the picture (pixels)

The inquire about comprises of a few segments:: Inserting and reading the image using Matlab, than study for (10) cases Optimize image contrast and cut from background , Optimize image contrast and minimize size, Minimize the image and subtract from its background, Improved image contrast and rotation, Rotate the image and cut from its background, Rotate the image and reduce its size, Optimize the image contrast, trim from its background, and minimize it, Adjust the image contrast, cut its background, and rotate it, Improved image rotation Rotated and minimized, Optimize the image contrast, truncate its background, then minimize it, rotate it, and then Estimation of regression models for (10) cases and extraction of the coefficient of determination for each case, and finally Testing the quality of the models based on special criteria, including the Akaike Information criteria (AIC) and the Bayesian information criteria (BIC) and the Hannan-Quinn criteria (HQC) standard.
2. RESEARCH GOAL
The aim of the research is to use several linear regression models for digital images to test the effect of digital treatments on the image and to test the quality of the models based on special criteria, such as the Akaike Information Criteria (AIC) and the Bayesian information criteria (BIC) Hannan - Quinn criteria HQC.

3. Digital Image
The digital picture could be a lattice of focuses organized in columns, lines and each point of the advanced picture called Pixel, which is the littlest portion of the advanced image. This point can be recognized inside the advanced picture by the coordinate framework Z, K, and each point of the picture encompasses a numerical esteem speaking to the color of this point and the color extend between zero which speaks to the dark color and a super esteem speaking to the white speaks to the grayscale that ranges from zero for dark color and 255 for white color.

4. IMAGE ANALYSIS STATISTICAL
It is the extraction of information and data from the picture, where one of the foremost critical errands of picture examination is to distinguish the fundamental data extricated for the reason of measurable examination of information, and incorporates picture examination classification of the picture into two sorts, to be specific the wave classification and based on factual characteristics and non-wave classification is based to a great extent on the degree of closeness between designs.

5. IMAGE PROCESSING
It may be a handle whose input could be a picture and its output is another image by applying a set of operations on the first picture to urge the foremost exact data that can be included within the examined picture. These forms are:

1. Optimize image contrast and cut from background
2. Optimize image contrast and minimize size
3. Minimize the image and subtract from its background,
4. Improved image contrast and rotation.
5. Rotate the image and cut from its background.
6. Rotate the image and reduce its size.
7. Optimize the image contrast, trim from its background, and minimize it.
8. Adjust the image contrast, cut its background, and rotate it.
9. Improved image rotation Rotated and minimized.
10. Optimize the image contrast, truncate its background, then minimize it, rotate it

6. CRITITERIA INFORMATION [3]
It could be a degree of the relative quality of measurable modeling connected to a information set, as well as an assess of the amount of blocked data and of these criteria:

Akaike Information Standard AIC: This basis is utilized to choose the leading one show from a set of models based on the data criteria in which the analyst (Akaike 1973) depended on two vital parts: the primary agent of the change of the mistake and the moment agent of the show course, as within the taking after equation

\[
AIC = n \ln\left(\frac{sse}{n}\right) + 2q
\]

sse: - Total error boxes
q: - The number of parameters of the model
n: - Number of views

Where the model that corresponds to its lowest value is chosen.
Bayesian information standard BIC: The BIS standard for the previous standard was developed by Akaike to be more precise and flexible. The general formula is

$$BIC = n \ln(\frac{sse}{n}) - (n - q) \ln(\frac{q}{n}) + q \ln(n) + q \ln\left(\frac{\sigma_{\hat{x}}^2}{\sigma_a^2}\right)$$.

BIC: - represents the standard BIS information
sse: - Total error boxes
By neglecting some limits we get the approximate short form, which is

$$BIC = n \ln(\frac{sse}{n}) + q \ln(n)$$

q: - The number of parameters of the model
n: - Number of views
Where the model that corresponds to its lowest value is chosen the Hannan-Quinn criterion: In 1979, an alternative standard was proposed for BIC by replacing b

As follows:

$$BIC = n \ln\left(\frac{sse}{n}\right) + q \ln(n)$$

Where the model that corresponds to its lowest value is chosen

7. SIMPLE LINEAR REGRESSION [2]

A technique used to study and analyze the effect of a quantitative variable called an independent variable on another quantitative variable called the dependent variable. The regression model can be represented by the following equation: $y = a + bx$

Where it represents
a: the constant gradient or the fraction of the y axis
b: Straight line inclination
The two values are calculated from the following two ratios:

$$b = \frac{\sum xy - \bar{x} \bar{y}}{\sum x^2}, \quad a = \bar{y} - b\bar{x}$$

8. PRACTICAL

This investigate has been chosen a particular picture (unique picture) and conducted on them 10 picture processors where these computerized pictures were treated on the premise that they are pictures in dark and white which the number of focuses composed of the picture from white to dark is256A picture could be a set of Points based on the past depiction of computerized images. And after that ponder the degree of closeness between the first picture and the conventional picture through the taking after steps

1. Inserting and reading the image using Matlab
2. Optimize image contrast and reduce background size, reduce image contrast and reduce background size, improve image contrast and rotate, rotate image and trim from background, rotate and reduce image size, improve image contrast and cut off Improve image contrast, crop background and rotate, improve image rotation and rotate, reduce image contrast, crop background, minimize, rotate
3. Estimation of regression models for (10) cases and extraction of the coefficient of determination for each case
4. Testing the quality of the models based on special criteria, including the Akaike Information criteria (AIC) and the Bayesian information criteria (BIC) and the Hannan-Quinn criteria (HQC) standard.
Table (1)
Results of the effect of the process (10 treatments) on the
Original image using simple regression analysis and the coefficient of selection

| treatments | Regression Models                                      | $R^2$ |
|------------|--------------------------------------------------------|-------|
| 1          | Optimiz image contrast and cut from background         | 0.904 |
| 2          | Optimize image contrast and minimize size              | 0.908 |
| 3          | Minimize the image and subtract from its background    | 0.786 |
| 4          | Improved image contrast and rotation                   | 0.877 |
| 5          | Rotate the image and subtract from its background      | 0.835 |
| 6          | Rotate the image and reduce its size                   | 0.594 |
| 7          | Optimize the image contrast, cut from its background, and minimize it | 0.915 |
| 8          | Adjust the image contrast, cut its background, and rotate it | 0.907 |
| 9          | Improved image rotation Rotated and minimized          | 0.914 |
| 10         | Optimize the image contrast, truncate its background, then minimize it, rotate it | 0.917 |

The table above shows the extent to which the original image responds to the changes that occurred when using 10 image processors by using the linear regression model and the selection coefficient by which to know which cases are close and identical to the real picture. When the image has been modified to improve its contrast and subtract from its background and then reduce the rotation after the coefficient of determination ($R^2$) of the image (0.917), which means that the image after treatment explains the (91.7%) of the variation in the original image. The case in which image contrast and truncation were improved From the background and then reduced it as the coefficient of identification ($R^2$) of the image (0.915), while the lowest value of the coefficient of identification ($R^2$) when the image was rotated and then reduced size as the value ($R^2$) of the image(0.594)

9. Image Quality Test
After the formation of 10 regression models, the quality of the models has now been tested by using three criteria for information (Akaike Information Criteria and Bayesian information criteria BIC) and the Hannan - Quinn criterion HQC. The image, which has been improved to contrast and cut from the background and then minimized then recycled with the highest quality to have the lowest standard of information (Akaike standard 1897.04, standard Bis 1912.64, Hannan-Quinn standard 1903.238).
| Pictures before and after treatment | AIC       | BIC       | HQC     |
|------------------------------------|-----------|-----------|---------|
| Optimiz image contrast and cut from background | 1946.269  | 1954.07  | 1949.37 |
| Optimize image contrast and minimize size | 1928.998  | 1936.798 | 1932.098 |
| Minimize the image and subtract from its background | 2238.78   | 2246.58  | 2241.88 |
| Improved image contrast and rotation | 2045.81   | 2043.61  | 2038.91 |
| Rotate the image and cut from its background | 2143.24   | 2151.042 | 2146.34 |
| Rotate the image and reduce its size | 2472.92   | 2480.721 | 2476.021 |
| Optimize the image contrast, trim from its background, and minimize it | 1905.65   | 1917.394 | 1910.297 |
| Adjust the image contrast, cut its background, and rotate it | 1936.43   | 1948.13  | 1941.08 |
| Improved image rotation Rotated and minimized | 1909.11   | 1920.81  | 1913.76 |
| Optimize the image contrast, truncate its background, then minimize it, rotate it | 1897.04   | 1912.64  | 1903.238 |
10. Conclusions
1. The results of the research show that the use of simple regression analysis method is more accurate in distinguishing between images
2. Using the coefficient of selection to compare regression models shows that R10 > R52 > R32 > R62 > R72 > R92 > R22 > R82 > R12 > R42. 
3. To compare the images and study the extent of to compare the quality of images and to study the extent of image similarity before and after treatment

AIC10 < AIC7 < AIC9 < AIC2 < AIC8 < AIC1 < AIC4 < AIC5 < AIC3 < AIC6
BIC10 < BIC7 < BIC9 < BIC2 < BIC8 < BIC1
< BIC4 < BIC5 < BIC3 < BIC6
HQC10 < HQC7 < HQC9 < HQC2 < HQC8 < HQC1
< HQC4 < HQC5 < HQC3 < AIC6

Images that have been modified to improve their contrast and cut from the background and then reduced after the rotation of the highest quality to have the lowest standard of information.

11. RECOMMENDATIONS
1. We suggest advance thinks about comparing the strategy of relapse investigation to determine the degree to which the initial picture of the changes that influence it
2. Other tests can be used to evaluate alternative models thus enhancing our confidence in the alternative model such as jackknife, cross validation, relative maximum absolute error
3. The bootstrap can be used to select the best regression model

Regression models can be formed at other distributions such as uniform, lognormal distribution.

References
[1] J. Schurmann (1996), Pattern Classification : a Unified View of Statistical and Neural Approaches, Jon Wiley and Sone, New York.
[2] Scott, H., (2010), “The Effectiveness of Information Criteria in Determining Unit Root and Trend Status” CESIS, NO.213
[3] De-Graft, H. (2010), “Comparison of Akaikie information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship” Journal of Development and Agricultura Economics Vol. 2(1) pp. 001-006.