UDC 621.39

COMPARISON OF METRIC FUNCTIONS AND CORRELATION WHEN FINDING A MEASURE OF SIMILARITY BY THE COEFFICIENTS OF POLYNOMIAL

ALEXANDR ZALESKI, VALERY CHERTKOV
Polotsk State University, Belarus

The main functions of the metric and correlation are disclosed in the article. The analysis of Hausdorff metric function is made, as well as the metric based on the root-mean-square error, and also the normalized averaged correlation function is depicted. The results of the experiment are compared. The conclusions are drawn.

Image processing using digital transformations is increasingly used to solve a variety of applied problems in communication, radar, measurement technology, medicine and other fields of science and technology. Various algorithms for finding similar images are widely used in the media, search engines, security systems. By searching for similar images, image classification tasks are effectively handled [1].

An effective approach to comparison of reference images is the use of image representation in the form of unordered sets of some pre-selected elements.

In order to find similarity of identification images similarity of measures are used. Measures of similarity are subdivided into four types: correlation coefficients; measures of distance; associative coefficients and probability of coefficients similarity. Although all four types of similarity measures were widely used at previous time, only correlation and distance coefficients were widely used [2].

1. The correlation coefficient - a measure of mutual influence nature of two random variable changes. Defined by the expression (1):

$$R^{COR} = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(o_{ij}-\bar{o})(b_{ij}^{*}-\bar{b}^{*})}{\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(o_{ij}-\bar{o})^{2}} * \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(b_{ij}^{*}-\bar{b}^{*})^{2}}} \quad (1)$$

where $\bar{o}$ и $\overline{b^{*}}$ - average images o and b^{*}

The value of the correlation coefficient varies from -1 to +1, with zero indicating that there is no connection between the objects.

The main drawback of the correlation coefficient as the similarity measure is that it is sensitive to the form by reducing sensitivity to the magnitude of differences between variables. Furthermore, the correlation calculated in this manner has no statistical sense [3].

2. Measure of distance (metric) is widely popular. Two objects are identical when description of their variables takes the same value. In this case, the distance between them is zero. Distance measures dependent on the choice of scale measurements and usually not bounded above. One of the most well-known distances is a function based on the Hausdorff metric, defined as (2):

$$R^{H} = 1 - \frac{1}{L} max_{ij} \left| a_{ij} - b_{ij}^{*} \right|, \quad (2)$$

where $i \in 0 \ldots N-1,\ j \in 0 \ldots N-1$

The Hausdorff metric is a natural metric defined on the set of all non-empty compact subsets of a metric space. Thus, the Hausdorff metric transforms the set of all non-empty compact subsets of a metric space into a metric space. Used as a measure of similarity, it is very sensitive to noise. Consequently, metricity in itself does not guarantee the strictness of the evaluation of similarity itself, but at the same time this measure satisfies the aggregate of other requirements, but not a metric, can in most cases give a better estimate of similarity than another measure that is a metric [4].

In this experiment, the metric determined the measure of similarity from the data of polynomials. Since the metric can only work with data.

Function based on root-mean-square error, defined as (3):

$$R^{s} = 1 - \frac{1}{LN^2}\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(a_{ij} - b_{ij}^{*}\right)^2} \quad (3)$$

The root-mean-square deviation characterizes the measure of data dispersion, but now (unlike dispersion), it can be compared with the original data, since the units of measurement are the same, but this indicator in its pure form is not very informative, since it contains too many intermediate calculations [5].

As the measure of uncertainty, it also participates in many statistical calculations. With its help, the degree of accuracy of various estimates and forecasts is established. If the variation is very large, then the standard deviation will also turn out to be large, hence, the forecast will be inaccurate, which will be expressed, for example, in very wide confidence intervals.

The disadvantage of these distance measures is that the evaluation of similarity strongly depends on the differences in the data shifts. Moreover, the metric distances change under the influence of scale measurement transformations of variables [6] .

3. Computational experiment data.

In the course of the experiment, graphs were found for investigating and comparing with the reference polynomial the results of modeling and searching for each of the functions. The results of modeling of the Hausdorff function are presented in the graph 1.

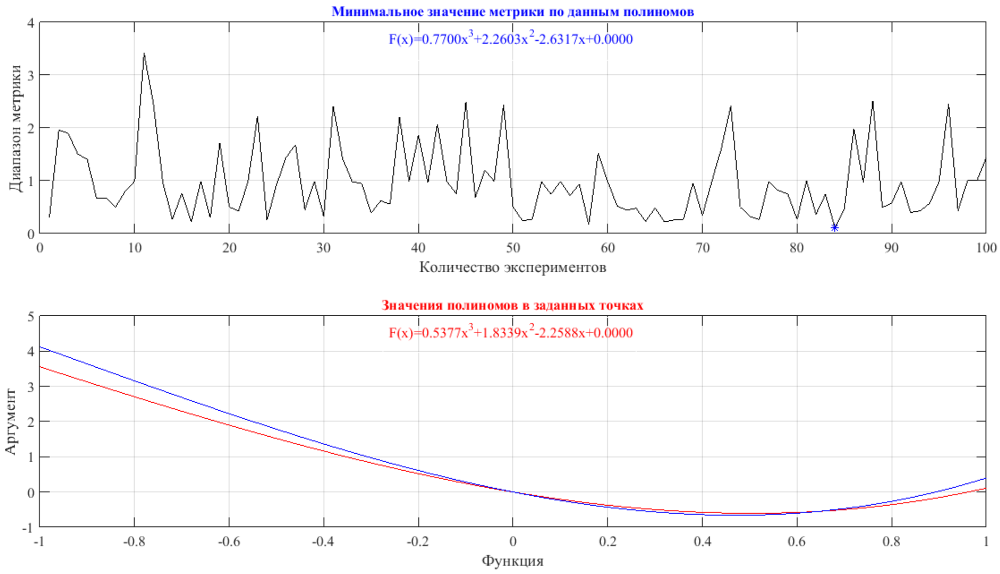

Fig.1. The graph of the Hausdorff function simulation

The result of the experiment on the coefficients of polynomials:

Reference polynomial:
$F(x) = 0.5377x^3 + 1.8339x^2 - 2.2588x + 0.000$

The polynomial found by the Hausdorff function:
$F(x) = 0.7700x^3 + 2.2603x^2 - 2.6317x + 0.000$
The results of modeling the root-mean-square error function are presented in the graph 2.

The result of the experiment on the coefficients of polynomials:

Reference polynomial:
$F(x) = 0.5377x^3 + 1.8339x^2 - 2.2588x + 0.000$

The polynomial found by the function of the root-mean-square error:
$F(x) = 0.2346x^3 + 1.8569x^2 - 2.2076x + 0.000$

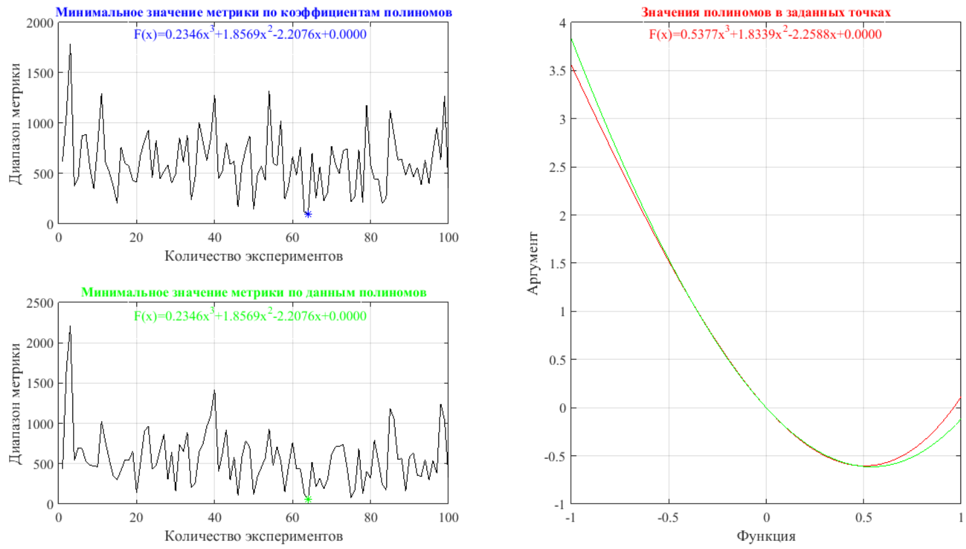

Fig.2. The graph of the function simulation based on the root-mean-square error

The results of modeling the normalized averaged correlation function are presented in graph 3.

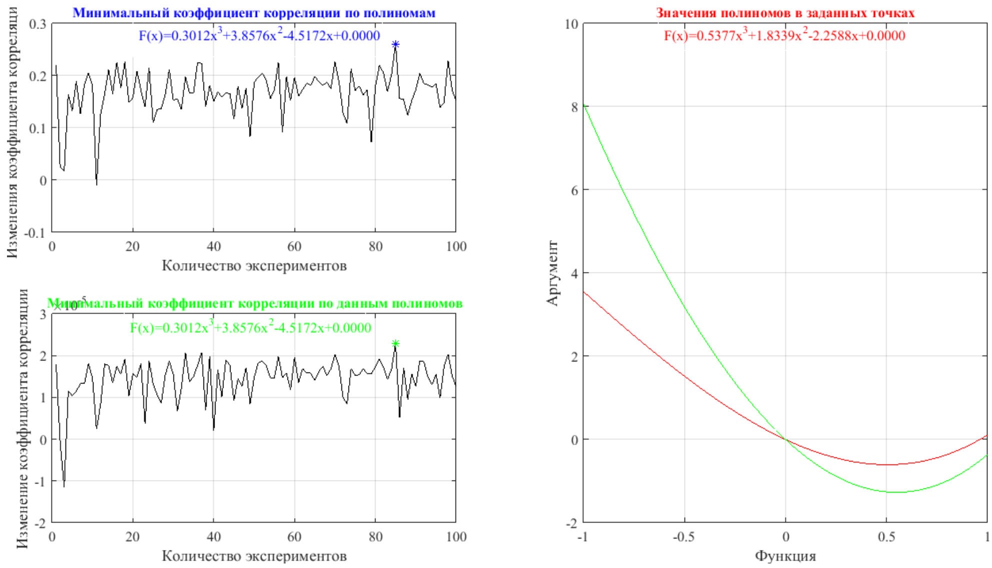

Fig.3. The graph of the simulation of the normalized averaged correlation function

The result of the experiment on the coefficients of polynomials:

Reference polynomial:
$F(x) = 0.5377x^3 + 1.8339x^2 - 2.2588x + 0.000$

The polynomial found by the normalized averaged correlation function:
$F(x) = 0.3012x^3 + 3.8576x^2 - 4.5172x + 0.000$

CONCLUSION. In this scientific article, some metrics and the correlation coefficient were examined in detail, in the context of their comparison in the search for a measure of similarity with respect to the coefficients of the polynomial. The computational experiment and the analysis of the obtained graphs both show that the best indicators for finding the measure of similarity by the coefficients of polynomials, showed metrics based on the root-mean-square error and the metric based on the Hausdorff function. The correlation function in the experiment for investigating a measure of similarity by coefficients proved to be less suitable, in view of the large error in determining the similarity. It is established that the most promising method is determination of a measure of similarity with respect to the coefficients of a polynomial, since, in comparison with the data, it has an order of magnitude higher accuracy.

REFERENCES

1. Воронцов, К.В. Алгоритмы кластеризации и многомерного шкалирования : курс лекций / К.В. Воронцов. – МГУ, 2007.
2. Аль-Фурайджи, О.Дж. Прогрессивный поиск схожести двух изображений на основе сопоставления угловых контурных точек / О.Дж. Аль-Фурайджи, В.Ю. Цветков // Доклады бел. гос. университета информатики и радиоэлектроники. – 2010. – № 5 (51). – С. 66–71.
3. Чубукова, И.А. Data Mining : курс лекций [Электронный ресурс] / И.А. Чубукова // Интернет-университет информационных технологий. – Режим доступа: www.intuit.ru/department/database/datamining.
4. Васильев, Н. Метрические пространства / Н. Васильев // Квант. — 1990. — № 1.
5. Харин, Ю.С. Теория вероятностей, математическая и прикладная статистика : учебник / Ю.С. Харин, Н.М. Зуев, Е.Е. Жук. – Минск : БГУ, 2011. – 464 с.
6. Метрические пространства [Электронный ресурс]. – Режим доступа: http://dfgm.math.msu.su/files/ivanov-tuzhilin/2014-2015/METRGEOM2014-1.pdf.