Regularization of automatic training of Mahalanobis neurons for small samples of examples of the “Own” image

T.A. Zolotareva^{1,}, A.I. Ivanov¹, O.V. Selishchev¹, and D.M. Skudnev¹*

¹Lipetsk State Pedagogical P.Semenov-Tyan-Shansky University, Lipetsk, 398020, Russia

Abstract. The paper considers the problem of regularization of automatic stable training of networks of artificial Mahalanobis neurons that simultaneously perform recognition of the “Own” image.

1 Introduction

One of the key conditions for the development of artificial intelligence software applications is the implementation of a number of information security rules. The paper deals with the problem of regularization of automatic stable training of networks of artificial Mahalanobis neurons, which simultaneously perform not only recognition of the “Own” image, but also hashing biometric data of “Alien” images. The stability of the training automaton for each of the information security neurons is ensured by the choice of independent biometric data. The problem of selecting independent data is solved by using classical statistical criteria. Unfortunately, single statistical criteria for testing the independence hypothesis do not work well on small samples. To solve this problem, it is necessary to perform a multi-criteria statistical test [1].

2 The ellipsoids of the normal distributions of the data of the image of “Own”

It is known from the practice of processing biometric data, that the biometric parameters of the “Own” image are quite strongly correlated, that is, they are displayed on the plane as an ellipse. Images of “All are Aliens” on the contrary are very poorly correlated and, as a result, are displayed on the plane as a circle. The latter means that the ellipse of the distribution of examples of the “Own” image is located inside the circle of distribution of the “All are Aliens” images, as shown in figure 1.

Figure 1 shows that one quadratic neuron with 4 inputs is equivalent to three or more linear neurons (perceptrons) in terms of the efficiency of distributing data from the “Own” image. The easiest way to verify this is by numerical modeling using real biometric data. A publicly available source of real data is the “BioNeiroAutograph” simulation environment [2], specially created for Russian-speaking universities. The simulation environment

* Corresponding author: zolotarevatatyana2016@yandex.ru

converts the dynamics of handwritten image reproduction into 416 biometric parameters (416 most significant coefficients of the two-dimensional Fourier transform). This tool of obtaining reliable data is designed in such a way that the user can observe all biometric parameters and all weight coefficients of 256 perceptrons [3], automatically trained by the GOST R 52633.5-2011 algorithm [1] on 8 or more examples of the “Own” image. Handwritten images are reproduced using a “mouse” manipulator, a graphic tablet, or a finger on a sensitive computer screen.

Fig. 1. One quadratic neuron is equivalent to 3 or more linear neurons in terms of the efficiency of distributing data from the “Own” image.

It should be noted that optimal for this software product are handwritten words consisting of 5 letters, which the “BioNeiroAutograph” product converts into a 256-bit code. At the same time, each neuron is able to take into account the statistics of 25 input biometric parameters during its training. Such a large number of inputs is redundant. On average, only 16 inputs of linear neurons are involved (they have non-zero weight coefficients). Presumably, transfer to the use of quadratic neurons will reduce the number of neuron inputs from 16 to 4, and the number of neurons themselves can be reduced by half without losing the quality of decisions made by the neural network.

All this is due to the nature of the tasks being solved. Using quadratic neurons, it is much easier to select ellipses of distributions of data of the “Own” image. Linear neurons divide the 416-dimensional “All are Aliens” area by hyperplanes that do not have the right to intersect the 416-dimensional “Own” data hyperellips. The best solutions are hyperplanes that are tangent to the “Own” hyperellips. However, each two-dimensional cross-section of the conditions of a single neuron will always produce images similar to figure 1. The ability to distribute biometric “Own” data by quadratic neurons is always higher than by linear neurons.

In general, ellipsoids of normal distributions of “Own” data image are described by quadratic forms, historically called the Mahalanobis metric (an Indian scientist who was engaged in biometrics (anthropometry of skulls) in the 30s of the XX century) or the Mahalanobis neuron:

$$\begin{cases} e_i^2 = [E(\bar{x}) - \bar{x}_i]^T \cdot [r]^{-1} \cdot [E(\bar{x}) - \bar{x}_i] \\ z(e_i^2) \rightarrow "0" & \text{if } e_i^2 \leq k \\ z(e_i^2) \rightarrow "1" & \text{if } e_i^2 > k \end{cases} \tag{1}$$

where $\bar{x}_i$ - is the vector of pre-normalized biometric parameters of the i -th example of the “Own” image, all standard deviations of the parameters of this vector are single; $E(\bar{x})$ - is the vector of mathematical expectations, analyzed biometric parameters; $[r]$ - is the correlation matrix of bio-parameters; k – is the value of the limit of the Mahalanobis neuron.

Usually, the value of the quantizer limit of the Mahalanobis quadratic neuron is chosen so that all examples of the training sample give the condition “0”:

$$k = \max(e_i^2) \cdot 1.01 \quad (2)$$

In this case, the Mahalanobis quadratic neuron provides the probability of errors of the first kind inversely proportional to the size of the training sample. For a typical training sample of 20 examples, we get:

$$P_1 \approx \left(1 - \frac{1}{n}\right) = \left(1 - \frac{1}{20}\right) = 0.95 \quad (3)$$

3 The conditionals for correlated data

In the classical literature of the twentieth century on images recognition [4, 5, 6], the Mahalanobis neuron (1) is considered as a convenient theoretical construction, far from practical application. The fact that square-shaped neurons were not used in practice in the 80s of the twentieth century is most likely due to the lack of small-sized, low-consuming processors at that time. Today, the situation has changed, there are enough microcontrollers, but the ideological inertia that has been formed for decades makes it difficult to look at these mathematical constructions in a new way.

Unfortunately, we still live in the ideological paradigms of the early middle ages and the postulate of linear algebra about the “curse of dimension”. All this makes it difficult to apply computer technology everywhere when solving any problems. In our case, the problem of automatic training of the Mahalanobis neuron comes down to the problem of calculating the inverse correlation matrix in the system (1).

With the advent of the first computing machines in the middle of XX century, engineers and researchers faced ill-defined challenges. This class of problems includes solving systems of linear equations and inversions of linear algebra matrices [7, 8, 9]. Each matrix has a so-called condition number:

$$\text{cond}[r] \approx \frac{\max(\lambda_i)}{\min(\lambda_i)} \quad (4)$$

where λ_i - are the eigenvalues of the invertible matrix.

The condition number $\text{cond}[r]$ can be considered in engineering terms as the gain of the noise of the initial data and the noise of the computer. The noise of a digital computer can always be reduced to any given value by increasing the length of the used bit grid. A completely different situation occurs with the noise of calculating mathematical expectations included in the system (1). In order to calculate the expected value accurately enough, we will need a training sample of 200 examples or even 2000 examples, while real training should be performed on a training sample of 20 examples of the “Own” image.

It is the low accuracy of calculating the mathematical expectation and standard deviation that prevents the widespread practical use of square-shaped neurons. The second obstacle is the ideology of searching for the so – called most informative parameters or “Occam’s razor” (William of Occam is a 12th-century English monk who is credited with formulating the principle: “Plurality must never be posited without necessity”). The

pernicious consequences of the total use of Occam’s razor can be easily shown by a simple example of the circulation of symmetric correlation matrices with very high values of correlation coefficients. The results of calculating the condition numbers for matrices of 3, 5, and 7 orders are shown in figure 2.

Fig. 2. Indicators of the condition number of strongly correlated data.

It seems that the proponents of searching for the most informative parameters are right. Strongly correlated data must be thrown out. However, this is only true for Mahalanobis neurons with elliptic quantizers (1). If we switch to using Bayes neurons with hyperbolic quantizers, the situation changes dramatically [10, 11, 12, 13]. Bayes neurons are the inverse of Mahalanobis neurons. The higher the correlations controlled by Bayes neurons, the more effective they are.

The symmetrization of correlations is convenient because it gives easily interpreted results in theoretical studies. In particular, it allows seeing how quickly the values of the condition numbers fall when the correlation relationships weaken. Figure 3 shows data on the condition numbers for weakly correlated data.

Fig. 3. Low level of condition numbers for weakly correlated biometric data.

Figure 3 shows that the condition numbers for weakly correlated data are almost independent of the dimension of the inverted matrices. The latter means that we can always achieve acceptable values of the condition number of the inverted matrix by selecting weakly dependent biometric parameters [14]. As shown in figure 4, the distribution of the values of the correlation coefficients of real biometric data is close to the normal law with a flattened vertex. In other words, there are a lot of poorly correlated data, but their correct selection is hindered by a small sample size. For 20 examples of the “Own” image, the accuracy of calculating correlation coefficients is low, and it needs to be improved by special data processing methods.

Fig. 4. Example of a histogram showing the distribution of the values of the pair correlation coefficients of biometric parameters of the 21st example of the handwritten word “Penza” according to the data of the “BioNeuroAutograph” simulation environment [2].

Figure 4 shows that the central section with a flat top of the data distribution is well described by the uniform law of distribution of values with a probability density amplitude of 1.12 and a correlation value range from -0.33 to +0.33. That is, the probability that pairs of probability coefficients fall into this interval is $P=1.12 \cdot 0.66=0.739$. The total number of coefficients of the correlation matrix $(416 \cdot 416 - 416)/2=86320$, among them, we will have to check $86320 \cdot 0.739 \approx 63790$ pairs of biometric data samples for the validity of the independence hypothesis. In the considered procedure of regularization of calculations, we are dealing with a problem of quadratic computational complexity.

4 Conclusion

Thus, the advantage of quadratic neurons over perceptrons in neural network processing of biometric data is obvious. Regularization of automatic training of Mahalanobis neurons for small samples of examples of the “Own” image is convenient because it gives easily interpreted results in theoretical studies.

As a result, it becomes possible to significantly enhance the protective properties of neural network converters of biometrics into code due to the realization of several topological advantages of quadratic neurons in comparison with already well-studied networks of linear neurons.

References

1. Ivanov A I, Zolotareva T A 2020 *Artificial intelligence in protected execution: synthesis of statistical neural network automata of multi-criteria verification of the hypothesis of independence of small samples of biometric data: preprint* (Penza: “Penza State University” publishing house) p 102
2. Ivanov A I, Zakharov O S 2009 *JSC “Penza Scientific and Research Electrotechnical Institute”* <http://пниэи.рф/activity/science/noc/bioneuroautograph.zip>
3. Ivanov A I 2013 *Automatic training of large artificial neural networks in biometric applications. Textbook* p 30 http://пниэи.рф/activity/science/noc/tm_IvanovAI.pdf

4. Duda R, Hart P 1976 Image recognition and scene analysis (M.: Publishing house "Mir") p 511
5. Hunt E 1978 *Artificial intelligence* (M.: Mir) p 558
6. Tou J T, Gonzalez R C 1980 *Pattern Recognition Principles* (M.: Sov. Radio) p 408
7. Lawson C, Hanson R 1966 *Numerical solution of problems by the least squares method* (Moscow: Nauka) p 230
8. Forsythe J, Moler K 1969 *Numerical solution of systems of linear algebraic equations* (M.: Mir)
9. Rice J 1984 *Matrix calculations and mathematical support* (Moscow: Mir) p 412
10. Lozhnikov P S, Ivanov A I, Kachaykin E I, Sulavko A E 2015 *Information security issues. Federal state unitary enterprise* **3** 48
11. Ivanov A I, Lozhnikov P S, Sulavko A E, Serikova Y I *Infocommunication technologies* **15(2)** 186
12. Ivanov A I, Lozhnikov P S, Sulavko A E 2017 *Computer optics* **41(5)** 765
13. Lozhnikov P S 2017 *Biometric protection of hybrid document flow: monograph* (Novosibirsk: SB RAS publishing house) p 130
14. Volchikhin V I, Ivanov A I, Malygina E A, Serikova Y I 2018 *Measurement. Monitoring. Management. Control* **1(23)** 66