Bayes-Minkowski measure and building on its basis immune machine learning algorithms for biometric facial identification

A E Sulavko
Omsk State Technical University, 11 Mira ave., Omsk, 644050, Russia

Abstract. In this paper we propose a new Bayes-Minkowski proximity measure that can be used to process correlated biometric, biomedical, and other type of data (with the normal distribution law or close to it). The Bayes-Minkowski measure is an antagonist criterion with respect to the Minkowski measure, since it shows opposite properties. It is possible to build a hybrid network of classifiers and apply immune learning algorithms to the network based on these proximity measures. It was demonstrated in the work on the example of tasks of identification and verification of a person’s personality by facial image. The achieved errors probabilities of person’s identification and verification by face features were: 0.0029 and 0.0017, respectively.

1. Introduction
At the present day one of the main global trend is associated with the development of technology of artificial intelligence. The most popular approaches among other to solving the problems of classification, clustering and regression are the device of multilayer artificial neural networks (ANN) and deep learning methods. Ideas of the use of "deep" neural network architectures are actively popularized by large corporations including Google, NVidia, Intel, Microsoft. In many cases, these ideas have been brought to effective practical solutions. However, the mentioned apparatus has disadvantages [1]:
1. Iterative learning algorithms become unstable on small samples (they are tend to retraining and to decrease in accuracy with little changes in the parameters of the ANN). It requires a huge database of examples (tens, hundreds of thousands) to train the multi-layer networks.
2. Modern "deep" ANNs contain more than a hundred layers (e.g. ResNet-150) which are redundant considering Hornik and Helt-Nielson theorems (1989 and 1987). Mentioned theorems based on earlier Kolmogorov’s work (any continuous function can be approximated with any accuracy by ANN with one hidden layer, the real limitations of shallow networks are being actively studied [2, 3]). The increase the number of layers is often unjustified and leads to a significant increase of the training sample and training time.
3. Theorems about the representability of functions in the form of ANNs and the convergence of training procedures give a little information about the composing of the optimal ANN configuration for a given task or learning sample. Therefore, deep neural network training hasn’t been fully automated and it is always being under human control (a research engineer need to select too many parameters that affect the neural network structure and training algorithm that creates a high labor costs).
4. Existing training algorithms are more or less exposed to the following problems: retraining, attenuation of gradients, getting into the local minimum of the error surface, “network paralysis”.
The above shortcomings does not allow the use of "deep" neural networks in a number of tasks that are characterized by low volume of training sample. For example, they include AI tasks from the field of medicine. The sample creation for studies is associated with the need to verify the patient’s disease
which is often associated with invasive studies. Typically, a sufficient amount of samples are collected over the years. Most clearly problems of a small volume and a low representativeness of training samples are appeared at the creation of automatic machines for biometric identification and authentication. The specificity of these tasks is that the biometric system must be tuned quickly (you cannot require the user to repeat entering biometric data many times, otherwise the system will not be in demand in practice). This example is characterized by the fact that the problem of sampling shortages will not disappear in the future. It is no matter how many biometric data has been accumulated by researchers around the world. In real practice, the system will still be trained on a small number of examples (10-30). Therefore, it is necessary to develop high-precision models of classifiers and algorithms for their rapid stability training in small samples.

In this paper a new Bayes-Minkowski proximity measure that can be used to process correlated biometric, biomedical, and other type of data (with the normal distribution law or close to it) was proposed. The Bayes-Minkowski measure is an antagonist criterion with respect to the Minkowski measure, since it shows opposite properties. It is possible to build a network (committee) of classifiers and apply immune learning algorithms [4] to the network with and without a teacher based on these proximity measures. It was demonstrated in the work on the example of tasks of identification and verification of a person’s personality by facial image.

2. Minkowski measure and “curvature” of features space

Quadratic proximity measures and networks based on them are used in biometric applications of the theory of pattern recognition [5]. The indicated proximity measures are generalized in the form of a Minkowski measure (1):

\[ y = \sum_{j=1}^{n} \left( \frac{m_j - a_j}{\sigma_j} \right)^p \]

where \( a_j \) is a value of \( j \) feature from vector \( \tilde{a} \), representing a recognizable image (biometric or other) \( m_j \) and \( \sigma_j \) – mean and standard deviation values of \( j \) feature for those class of images, with which the \( \tilde{a} \) image is compared, \( p \) – power coefficient determining the level of “curvature” of space. While \( p=1 \) we get a “city block measure”, while \( p=2 \) – Pearson’s measure. The curvature of the feature space arises due to the presence of correlation relationships between dimensions (Fig. 1). The feature space is Euclidean nor equally curved in biometric identification and authentication problems. Rather, the level of curvature of the feature space changes relative to the observer.

![Figure 1. The compression direction of the space of two features: a. if there is a positive correlation between them, b. in the presence of a negative correlation between them c. with independence of features](image)

We can reduce the number of classification errors by changing the parameter \( p \). To demonstrate this computational experiment on pattern recognition in a space of 100 abstract (imitated) features was conducted. All features had a normal distribution of values (the most common case for biometrics). 500 classes of images were generated that differed among themselves by the distribution parameters of the values of the features (the class was described by two vectors of parameters - mathematical
expectations and standard deviations). Then, 100 feature vectors are generated for the corresponding parameters of each class using the Monte Carlo method. Two data sets were generated. Each of them included 500 classes with 125 sample images per class. In the second set the values of each feature within the class were sorted in ascending order before the formation of images. So in the first set signs remained independent random variables, and in the second one, correlation dependence appeared between them.

With each set of data was held separate session on pattern recognition in the verification regime by cross-comparison. Each classifier was trained on 25 randomly generated examples. The remaining 100 examples were used for testing. At the end of the session, the relative frequencies of occurrence of 1st and 2nd kind errors were calculated. In biometric systems, these indicators are called the probabilities (or percentage) of errors of "false rejection" (FRR) and "false acceptance" (FAR). Comparison of biometric systems is often performed by the coefficient of equal error rate (EER = FRR = FAR). The generalized experimental results are shown in Fig. 2.

Thus, if the features are dependent, then an increase in \( p \) can lead to a decrease in the number of recognition errors by several times. This important aspect must be taken into account when constructing any classifiers (neural network, immunological, ensemble), where any measure of distance in the feature space is used in one form or another.

### 3. Bayes-Minkowski measure and its properties

Bayes-Minkowski measure (2) proposed. These metric take on smaller values, the higher the correlation coefficient between the \( t \) and \( j \) features. A similar experiment on pattern recognition was conducted with the Bayes-Minkowski measure (Fig. 3).

\[
y_t = \sum_{j=1}^{n} \left( \frac{(m_t-a_t)}{\sigma_t} \right) - \left( \frac{(m_j-a_j)}{\sigma_j} \right).
\]

From the presented data it can be seen that the \( p \) optimum change in each considered case. If the features are independent then the minimum in EER is achieved at \( p>1 \), if they are correlated – at \( p<1 \). The dynamics of changes in EER \( (p) \) for the Bayes-Minkowski measure has the opposite tendency as compared to the EER \( (p) \) for the Minkowski measure. If the features are correlated, the Bayes-Minkowski measure gives a higher result than if the features are independent.
4. The Minkowski and Bayes-Minkowski classifiers ensemble

We formed a similar space of features but with the additional condition: 50 features were independent, 50 features – correlated. We generated 500 classes of 125 images (25 training and 100 test examples). For each class of images we constructed a committee of 200 different classifiers – 100 classifiers are based on the Minkowski measure (50 classifiers with $1 < p < 3$, 50 classifiers with $9 < p < 100$), 100 based on the Bayes-Minkowski measure (50 classifiers with $0.5 < p < 1$, 50 classifiers with $5 < p < 9$) with randomly specified parameters $50 \geq n \geq 10$. For each classifier we defined the subspace of correlated or independent features, randomly choosing $n$ features from 50 possible depending on its type and value of the parameter $p$. We apply the bagging method (bootstrap aggregation) when teaching classifier committees - we trained each basic (weak) classifier using only 16 random examples out of 25.

Taking as a basis the model of the artificial immune network from work we also applied the immune algorithms for training classifier committees[4]. The presented model was modified. As well as the mechanism for the generation and selection of computing elements – detectors (now the minimum power of the detectors is increased, and the correlation of their decisions is not controlled), its architecture was changed - the Minkowskogko and Bayes-Minkowski metrics were the basis for the new detectors architecture. In this work only the mechanism of formation of innate immunity [4] (AIS training with a teacher) was involved and the “acquired immunity” mechanism [4] (subsequent retraining of AIS without a teacher) was not involved. The effectiveness of the modified AIS model depends on the following parameters: $N$ is number of detectors (weak classifiers) comparing the input biometric image with the user's etalon; $I$ is number of iterations of AIS training (the longer the system learns the higher the reliability of solutions).

During the iterative process of setting up the committee the immune algorithm changes the distribution of the types of basic classifiers and selects the optimal parameters ($n$ and $p$) for them so that the number of system errors decreases. We repeated the pattern recognition experiment which results can be seen in Fig. 4 and 5.

**Figure 3.** The effect of $p$ and $n$ by EER: a. The features are independent; b. The features are correlated.

**Figure 4.** The effect of $p$ by EER (the level of features correlation is different).

**Figure 5.** Pattern recognition results based on a classifier committee (the level of features correlation is different).
It can be seen from the results that the classifier committee gives a much lower percentage of errors, and when using the committee’s immune training algorithms, the probability of error decreases even more significantly. Interestingly, with the mixed nature of the correlation of attributes, the optimum for both proximity measures is achieved at $p = 5$ (Fig. 4).

5. **Face identification**

In order not to be limited by the description of the theory, we apply the results of simulation to solve practical problems – identification (one-to-many comparison) and verification (one-to-one comparison) of a person using a face image. Database of images of the faces of 90 subjects from [6] was used to conduct a computational experiment. Before the beginning of experiment each image was processed to extract features using a technique based on the Viola-Jones method and the Hough transform (also from the work [6]).

The following face parameters were used as features:
- Distances between eyes, a center of a face, a tip of a nose (in pixels, values are normalized by a diagonal of a face in a frame);
- Area of eyes, a nose, a mouth (values are normalized by a face area);
- Coefficients of correlation of intensity and color parts of pixels (according to RGB) between all pairs of the following area of a face: eyes, a nose, a mouth. These features describe facial asymmetry;
- Parameters describing a color of eyes and skin.

Committees consisting of 200 classifiers are formed for each subject (the type of each classifier and the values of the $p$ and $n$ parameters were set taking into account the correlation of the features of the corresponding subject). 25 random examples of the subject’s face image were also used to train the committee. The observed results were presented in Fig. 5 and 6. In Table 1 we presented their comparison with those achieved earlier.

![Figure 6](image.png)

**Figure 6.** Characteristic curves illustrating the dependence of FRR and FAR on the immune system threshold when verifying facial images: a. $N=100$ (simple committee), b. $N=200$ (simple committee), c. $N=100$ (training with immune algorithm, $I=100$), d. $N=200$ (immune algorithm, $I=100$).
Figure 7. Characteristic curves illustrating the dependence of EEE on the immune system threshold when identifying facial images.

Table 1. The achieved level of errors in the recognition of personality by face features.

| Method                                           | Face image database, number of subjects | Probability of error |
|--------------------------------------------------|----------------------------------------|----------------------|
| neural networks [7]                              | ORL, 40                                | 0.0779               |
| immune network and genetic algorithm [8]         | ORL, 40                                | 0.003                |
| immune network [9]                               | Yale Face Database                     | EER=0.018            |
| negative selection based immune system [10]      | proprietary base, 65                   | 0.99                 |
| DocFace+ (deep convolutional neural networks [11])| dataset of Chinese Identity Cards (53591 different ID-selfie pairs) | FAR=0.001, FRR=0.0249 |
| hybrid neural networks (face + keyboard handwriting + signature) [6] | proprietary base, 90                 | EER=0.009            |
| present work (single committee, identification)  | proprietary base, 90                   | 0.0063               |
| present work (immune training, identification)   | proprietary base, 90                   | 0.0029               |
| present work (single committee, verification)    | proprietary base, 90                   | EER=0.004            |
| present work (immune training, verification)     | proprietary base, 90                   | EER=0.0017           |

The obtained results are significantly superior results achieved by other researchers, as well as the results obtained in the framework of the previous work of our team [6].

6. Conclusion

The Minkowski measure is suitable for pattern recognition in the space of weakly dependent features. The reliability of pattern recognition can be significantly increased by adjusting the power-law coefficient $p$. Since it has almost opposite properties the Bayes-Minkowski measure is an antagonist measure with respect to the Minkowski measure. The Bayes-Minkowski metric is suitable for pattern recognition in a space of highly dependent features (by changing $p$, you can also increase the reliability of pattern recognition). Both measures work differently for different $p$ values (in fact we have different proximity measures for different $p$, the solutions of which are less correlated, the stronger this parameter differs).
It is possible to form a committee of classifiers capable of more efficient decision-making from these proximity measures. When forming a committee, it is also worth using the bagging method (training different classifiers on different intersecting subsets of the training sample) and dividing the feature space into subspaces, but not randomly (for example, as in the “random forest” method), but depending on the correlation of the features. The selection of basic classifiers can be performed by an iterative immune algorithm (in particular, the algorithm proposed in [4] can be taken as the basis of the algorithm). As shown by an experimental verification the developed method of forming a decision network from the Minkowski and Bayes-Minkowski classifiers is highly effective. The achieved errors probabilities of person’s identification and verification by face features were: 0.0029 и 0.0017, respectively.

7. References

[1] Ivanov A I, Lozhnikov P S and Sulavko A E 2017 Evaluation of signature verification reliability based on artificial neural networks, Bayesian multivariate functional and quadratic forms (Samara: Computer Optics) 5 pp 765-774
[2] Kůrková V and Sanguineti M 2017 Probabilistic lower bounds for approximation by shallow perceptron networks Neural Networks vol 91 pp 34-41
[3] Kůrková V and Sanguineti M 2016 Model complexities of shallow networks representing highly varying functions Neurocomputing vol 171 pp 598-604
[4] Sulavko A E and Shalina E V 2019 Biometric authentication of information systems users using keyboard handwriting based on immune network algorithms Applied informatics 3(81) pp 39-53
[5] Ivanov A I, Kachajkin E I and Lozhnikov P S 2016 A Complete Statistical Model of a Handwritten Signature as an Object of Biometric Identification Control and Communications (SIBCON): Int. Siberian conference (Moscow: Electronic Materials) pp 1–5
[6] Lozhnikov P S and Sulavko A E 2018 Generation of a biometrically activated digital signature based on hybrid neural network algorithms Journal of Physics: Conf. Series vol 1050 pp 012047-1–012047-12
[7] Alexandrino J L, Cavalcanti G D C and Carvalho Filho E C B 2009 Hybrid intelligent system clonart applied to face recognition Int. Joint Conference on Neural Networks (Atlanta, Ga: IEEE) pp 102–107
[8] Luh G 2014 Face recognition based on artificial immune networks and principal component analysis with single training image per person Immune Computation 2(1) pp 21–34
[9] Sadeghi M, Maghooli K and Moin M S 2014 Using Artificial Immunity Network for Face Verification The International Arab Journal of Information Technology vol 11 4 pp 354-361
[10] Silva J C D 2015 et al. Identity Recognition Using an Artificial Intelligence Based on Artificial Immune System Asia-Pacific Conf. on Computer Aided System Engineering pp 158–162
[11] Shi Y and Jain A K 2019 DocFace+: ID Document to Selfie Matching IEEE Transactions on Biometrics, Behavior, and Identity Science vol 1 no 1 pp 56-67

Acknowledgments
The reported study was funded by RFBR according to the research project № 18-37-00399