E10-2007-17

E. P. Akishina[1], T. P. Akishina[1], V. V. Ivanov[1],
A. I. Maevskaya[2], O. A. Afanas'ev[1,3]

ELECTRON/PION IDENTIFICATION IN THE CBM TRD USING A MULTILAYER PERCEPTRON

[1]Joint Institute for Nuclear Research, Dubna
[2]Institute for Nuclear Research, Troitsk, Russia
[3]Kostroma State University, Kostroma, Russia

Акишина Е. П. и др. E10-2007-17
Идентификация электронов/пионов в CBM TRD с помощью многослойного перцептрона

Рассмотрена задача идентификации электронов/пионов в эксперименте CBM на основе ионизационных потерь энергии и переходного излучения в детекторе TRD. Исследована возможность решения указанной задачи с помощью искусственной нейронной сети. В качестве входной информации для сети использовались выборки, составленные как на основе потерь энергии пионов или электронов в поглотителях детектора TRD, так и на основе «умной» переменной, полученной из исходных данных. Показано, что использование новой переменной позволяет достичь разумного уровня идентификации частиц уже после 10–20 тренировочных эпох, при этом практически отсутствуют колебания относительно тренда и достигается необходимый уровень подавления пионов при условии минимальной потери электронов.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2007

Akishina E. P. et al. E10-2007-17
Electron/Pion Identification in the CBM TRD Using a Multilayer Perceptron

The problem of electron/pion identification in the CBM experiment based on the measurements of energy losses and transition radiation in the TRD detector is discussed. We consider a possibility to solve such a problem by applying an artificial neural network (ANN). As input information for the network we used both the samples of energy losses of pions or electrons in the TRD absorbers and the «clever» variable obtained on the basis of the original data. We show that usage of this new variable permits one to reach a reliable level of particle recognition no longer than after 10–20 training epochs; there are practically no fluctuations against the trend, and the needed level of pions suppression is obtained under the condition of a minimal loss of electrons.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2007

INTRODUCTION

The CBM Collaboration [2, 3] builds a dedicated heavy-ion experiment to investigate the properties of highly compressed baryonic matter as it is produced in nucleus–nucleus collisions at the Facility for Antiproton and Ion Research (FAIR) in Darmstadt, Germany. The scientific goal of the research program of the CBM experiment is to explore the phase diagram of strongly interacting matter in the region of highest baryon densities. This approach is complementary to the activities at RHIC (Brookhaven) and ALICE (CERN-LHC) which concentrate on the region of high temperatures and very low net baryon densities.

The experimental set-up has to satisfy the following requirements: identification of electrons which requires a pion suppression factor of the order of 10^5, identification of hadrons with large acceptance, determination of the primary and secondary vertices (accuracy $\sim$ 30 μm), high granularity of the detectors, fast detector response and read-out, very small detector dead time, high-speed trigger and data acquisition, radiation hard detectors and electronics, tolerance towards delta-electrons.

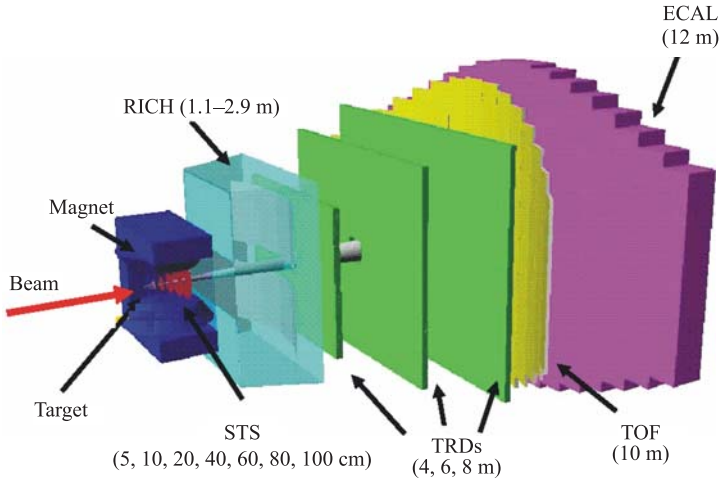

Fig. 1. CBM general layout

Figure 1 depicts the present layout of the CBM experimental set-up. Inside the dipole magnet gap there are a target and a 7-planes Silicon Tracking System

(STS) consisting of pixel and strip detectors. The Ring Imaging Cherenkov detector (RICH) has to detect electrons. The Transition Radiation Detector (TRD) arrays measure electrons with momentum above 1 GeV. The Time of Flight (TOF) detector consists of Resistive Plate Chambers (RPC). The Electromagnetic Calorimeter (ECAL) measures electrons, photons and muons. The CBM set-up is optimized for heavy-ion collisions in the beam energy range from about 8 to 45 AGeV. The typical central Au+Au collision in the CBM experiment will produce up to 700 tracks in the inner tracker (see Fig. 2).

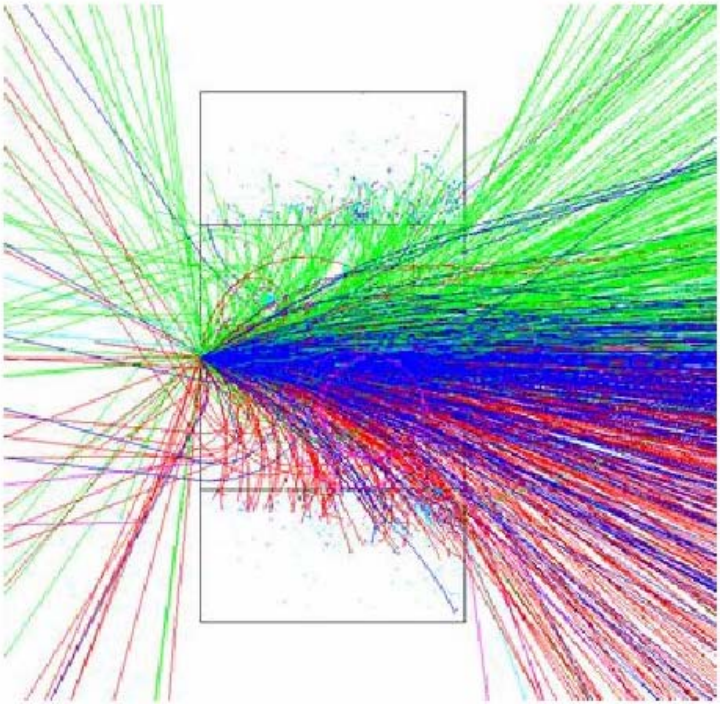

Fig. 2. Visualization of a typical CBM event

Large track multiplicities together with the presence of a non-homogeneous magnetic field make the reconstruction of events extremely complicated. It comprises local track finding and fitting in the STS and TRD, ring finding in RICH, cluster reconstruction in ECAL, global matching between STS, RICH, TRD, TOF and ECAL, and the reconstruction of primary and secondary vertices. Therefore, the collaboration performs the extensive analysis of different event recognition and reconstruction methods, in order to understand better the geometry of detectors and to investigate specific features of useful events [3].

The measurement of charmonium is one of the key goals of the CBM experiment. The main difficulty lies in the extremely low multiplicity expected in Au+Au 25 AGeV collisions near J/ψ production threshold. For detecting J/ψ meson in its dielectron decay channel the main task is the separation of electrons and pions. One of the most effective detectors for electron/pion separation is a multiwire proportional chamber TRD.

The TRD must provide electron identification and tracking of all charged particles. It has to provide, in conjuction with the RICH and the electromagnetic calorimeter, sufficient electron identification capability for the measurements of charmonium and low-mass vector mesons. The required pion suppression is a factor of about 100 and the required position resolution is of the order of 200–300 μm. In order to fulfill these tasks, in the context of the high rates and high particles multiplicities in CBM, a careful optimization of the detector is required.

In the technical proposal of the CBM experiment there were presented preliminary results on the estimation of the electron identification and pions suppression applying a maximum likelihood ratio test (see details in [3]). A standalone Monte Carlo C++ based simulation code was developed to perfom the simulations. The following processes were realized in the simulations: i) energy loss of electrons and pions in the gas detector due to the procedure described in [1]; ii) for electrons, production and absorption of the transition radiation (TR) in the radiator, absorption of TR in the mylar foil and absorption of TR in the active gas volume. The results of these simulations have demonstrated that the TRD will have 9 to 12 layers that can fulfill the required electron/pion indentification in the CBM experiment.

It must be noted that the application of a maximum likelihood ratio test requires a very accurate determination of distribution functions of energy losses by pions and electrons (see details on page 87 in [3]), which is not so simple to fulfill in practice.

Recently, the use of artificial neural networks in multi-dimensional data analysis has become widespread [4–7]. One of such problems consists in classifying individual events represented by empirical samples of finite volumes pertaining to one of the different distributions composing the distribution to be analyzed. A layered feed-forward network — a multilayer perceptron (MLP) — is a convenient tool for constructing multivariate classifiers, although its learning speed and power of recognition critically depend on the choice of input data.

In this work, we investigate a possibility to apply the MLP for identification of electrons and pions using the measurements of ionization energy losses and TR in the TRD detector.

1. MAIN PRINCIPLES OF PATTERN RECOGNITION BASED ON A FEED-FORWARD ANN

The MLP consists of a few layers of neurons: a layer of input neurons, one or several hidden layers and a layer of output neurons. The connections in such a network are set up between layers of neighbouring layers and the information movement is going on only in one direction: from input layer to output layer. The choice of the MLP architecture includes the determination: 1) number of layers, 2) number of neurons in each layer, 3) format of input data.

The analyzed data are received by the layer of input neurons; their number corresponds to the dimension of the input data (input pattern). A number of hidden layers and neurons are determined by a problem under the solution. The results of input patterns analysis are obtained from output neurons.

The recognition problem consists in determination of belongness of the input pattern (represented by a vector of features) to one or several a'priori determined classes [8]. The analyzed patterns are put to the MLP input layer, and the information obtained from the output layer permits one to get the answer to which class belongs a particular pattern.

Main principals of work of the MLP network and its application to the recognition problem are convenient to consider in an example taken from the paper by B. Denby [5]. Let one need to construct a classifier of events that belong to the class «a» or to the class «b» (see Fig. 3). The discriminator function

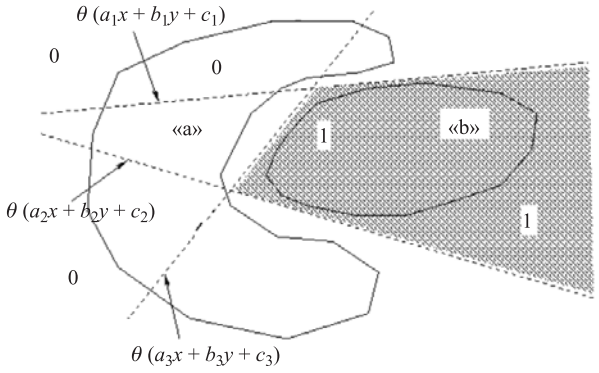

Fig. 3. Areas of events corresponding to two different classes: «a» and«b»

corresponding to the solution of this problem has the following form:

$$\begin{aligned} D &= \theta[\theta(a_1x + b_1y + c_1) + \theta(a_2x + b_2y + c_2) + \\ &+ \theta(a_3x + b_3y + c_3) - 2], \end{aligned} \tag{1}$$

where the threshold function $\theta(x - x^{'})$ equals 0 when $x < x^{'}$ and 1 for $x \geqslant x^{'}$. The parameters a_i, b_i and $c_i, i = 1, ..., 3$, are chosen in such a way that the fuction

(1) must take the value 1 in the area which includes the boundaries of the class «b» and the value 0 — in all other areas.

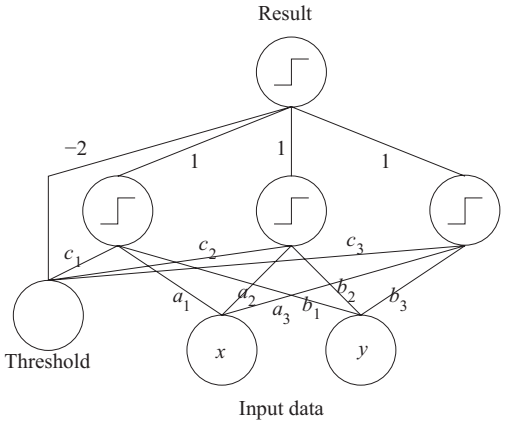

Fig. 4. Scheme of the discriminating function

Figure 4 shows the scheme of a discriminator that realizes function (1). On input x and y are given random variables that correspond to the current pattern. These values are multiplied by the coefficients a_i and $b_i, i = 1, ..., 3$, which determine the weigths of lines connecting the inputs of the scheme with the first layer of the threshold discriminators (TD). To the input of TDs the tresholds are also inputted which are responsible for parallel shift corresponding to the separating line on a plain. The signals from outputs of the hidden TDs are multiplied by new coefficients (which, in our case, equal to 1) and together with a threshold equal to –2 are put to the output TD.

The considered scheme of the discriminator function is a simplified model of a three-layered neural network of the feed-forward type in which the threshold discriminators play the role of neurons. If one replaces the threshold discriminators to a smooth function, for example, of the sigmoid type, then a neural network acquires very important property — **the ability to learn**. It is possible to classify complex patterns of a high dimension using such a network, which is practically impossible by traditional methods.

The network tuning (determination of coupling weights between neurons and thresholds) on a concrete problem is realized through its training using as usual the algorithm of backward error propagation [9].

During the training process the network is tuned onto the analyzed distributions that are realized through the weights w_{ij} correction. This problem is solved by the minimization of an error functional (the functional of the network energy). The error functional E presents a sum of differences between the output signals

and the target values:

$$E = \frac{1}{2}\sum_p \sum_i (\mathbf{y}_i^{(p)} - \mathbf{t}_i^{(p)})^2,$$

where p is the index of the input pattern, i. e., the event from a data set to be used for the network training, i is the index of the output neuron, $\mathbf{y}_i^{(p)}$ the output result obtained from the ith neuron in the pth pattern and $\mathbf{t}_i^{(p)}$ is the target value.

In the MLP each jth neuron is realizing the transformation $y_j = g(\sum_i w(j,i)y_i + \theta_j)$, where y_j is the output signal of the neuron under consideration, w_{ij} is the weight of connection between ith and jth neurons, g is the transition function (the transition function of a sigmoid type in the MLP is usually used, for example, $g(x) = \frac{1}{2}[1 + \tanh(x)]$). The architecture of the MLP network is presented in Fig. 5.

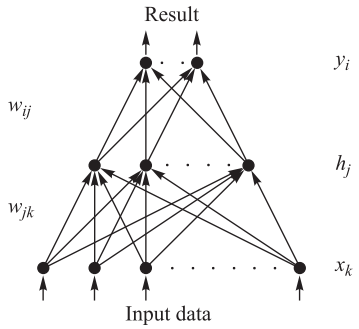

Fig. 5. Scheme of the multilayer perceptron with one hidden layer

Here, x_k, h_j and y_i denote correspondingly the input, hidden and output neurons; w_{jk} are the weights of input neurons with hidden ones, w_{ij} are the weights of hidden neurons with output neurons. The signals $a_j = \sum_k w_{jk}x_k$ and $a_i = \sum_j w_{ij}h_j$ are put to the inputs of hidden and output neurons, correspondingly. The values of output signals from these neurons are determined from relations

$$h_j = g(a_j/T) + \theta_j, \qquad y_i = g(a_i/T) + \theta_i,$$

where $g(a,T)$ is the transition function (T is the «temperature» that gives the value of its slope, and θ is the threshold of the corresponding node — neuron).

The output signals from the hidden and output layers are the functions of corresponding weights. To calculate the changes of weights w_{ij} and w_{jk}* on each iteration step, we have:

$$\Delta\omega_{ij} = -\eta\frac{\partial E}{\partial \omega_{ij}}, \tag{2}$$

and

$$\Delta\omega_{jk} = -\eta\frac{\partial E}{\partial \omega_{jk}}. \tag{3}$$

*Usually the gradient descent method is used for these calculations.

If we rewrite the expressions (2) and (3) in details, then we get:

$$\frac{\partial E}{\partial \omega_{ij}} = \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial a_i}\frac{\partial a_i}{\partial \omega_{ij}} = \delta_i g'(a_i) h_j, \tag{4}$$

and

$$\frac{\partial E}{\partial \omega_{jk}} = \sum_i \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial a_i}\frac{\partial a_i}{\partial h_j}\frac{\partial h_j}{\partial a_j}\frac{\partial a_j}{\partial \omega_{jk}} = \sum_i \delta_i g'(a_i)\omega_{ij} g'(a_j) x_k, \tag{5}$$

where the variable δ_i is determined from the following expression:

$$\delta_i = y_i - f(x_i).$$

Thus, for connections between hidden and output layers we have:

$$\Delta\omega_{ij} = -\eta\delta_i g'(a_i) h_j + \alpha\Delta\omega_{ij}^{\text{old}}. \tag{6}$$

Similarly, for connections between the input and hidden layers of neurons we get the expression:

$$\Delta\omega_{jk} = -\eta\sum_i \omega_{ij}\delta_i g'(a_i) g'(a_j) x_k + \alpha\Delta\omega_{jk}^{\text{old}}. \tag{7}$$

In expressions (6) and (7) η is the parameter that controls the speed of the network training [9], $\alpha\Delta\omega_{ij}^{\text{old}}$ and $\alpha\Delta\omega_{jk}^{\text{old}}$ are the moments which suppress the oscillations at the network output. The procedure of the network training is repeated untill the value of the output signal is close to the target value.

It must be noted that in practice the weights are usually corrected not for each training pattern, but on the basis of a small set of patterns. This permits one to accelerate the training process. Usually it is necessary to do a few passages on a whole training set to minimize the error functional and to get a reliable set of weights. At the end of the training process the weigths are fixed and the quality of the network training is estimated on the basis of testing data set.

2. ELECTRON/PION IDENTIFICATION APPLYING MLP

The J/ψ phase space distribution and decay kinematics were calculated with the PLUTO event generator [10] for beam energy 25 AGeV. The background was calculated with the UrQMD event generator [11] for the same beam energy. For calculations only events with one J/ψ electron daughter mixed with one minimum bias UrQMD event were used.

Simulations of signal and background events were performed with the CBM software framework CBMROOT (based on ROOT package [12]), using GEANT3

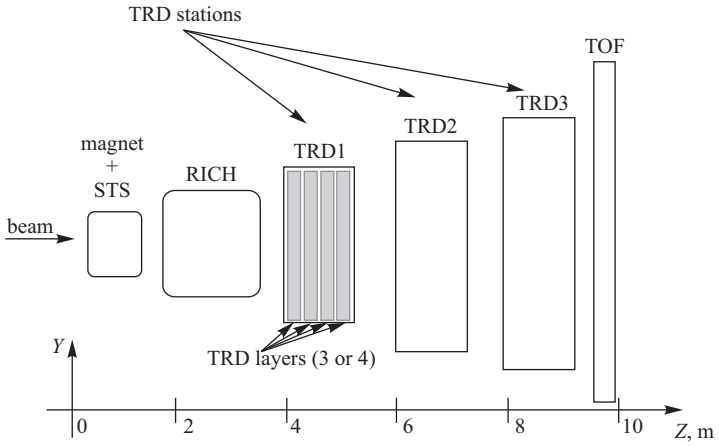

Fig. 6. Schematic view of the TRD

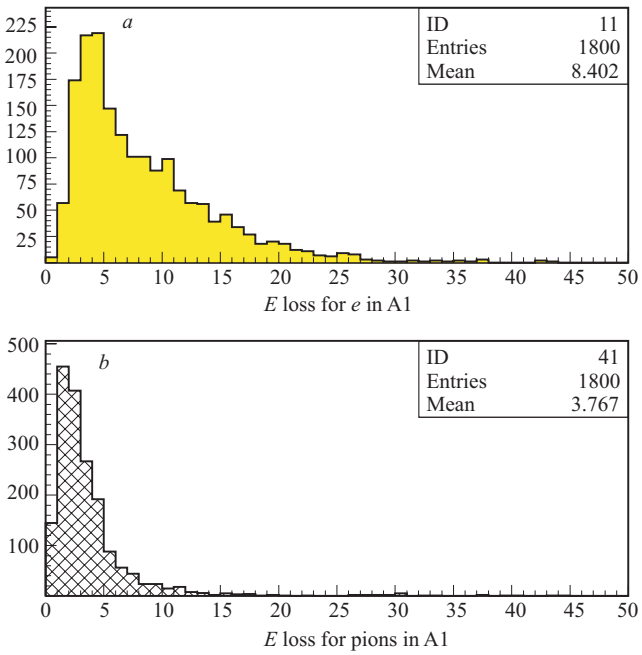

Fig. 7. Distributions of energy losses (including transition radiation) by electrons (a) and energy losses by pions (b) in the first absorber of the TRD

[13] transport through standard set-up with a gold target 250 mkm thickness, the beam pipe, STS and TRD. Set-up with 4 TRD arranged in 3 stations was used for simulation (Fig. 6). Only electrons (signal and background) and pions

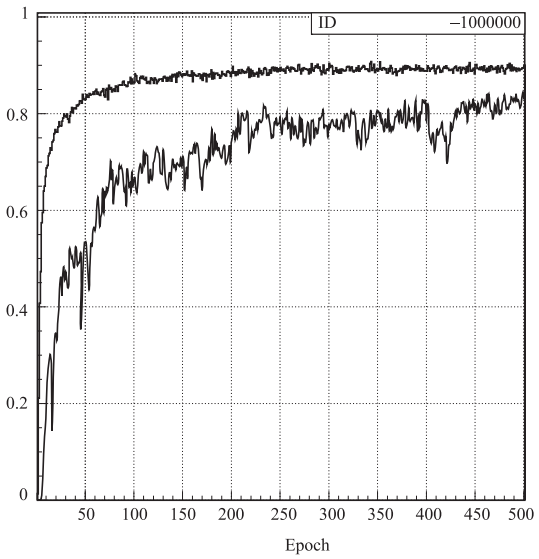

Fig. 8. The efficiency of pion/electron identification by the MLP for original (bottom curve) and transformed (top curve) samples

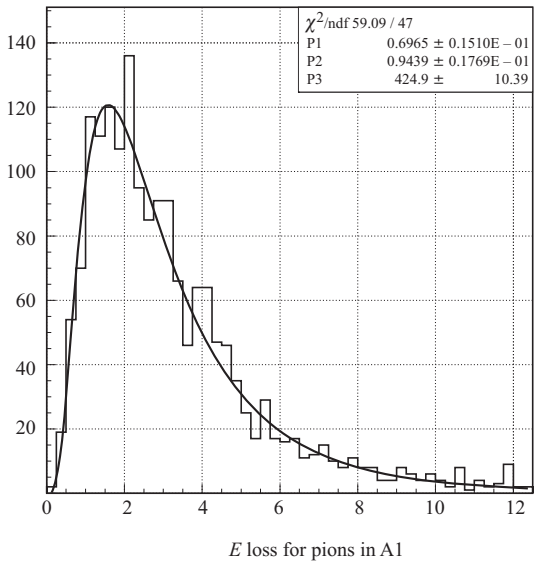

Fig. 9. Approximation of energy losses by pions in the first absorber of the TRD by the density function of the log-normal distribution (9)

(background) which made hits in 6 STS stations and 12 TRD layers participated in J/ψ reconstruction. Electron transition radiation energy loss in TRD gas was added to GEANT dE/dx.

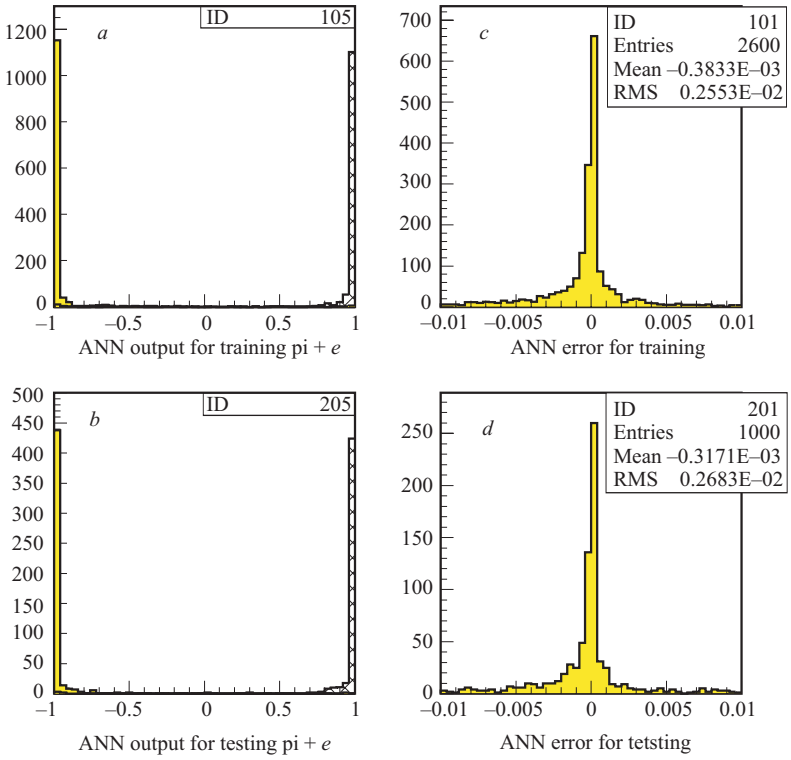

Fig. 10. Distributions of the MLP output signals obtained at the training (a) and testing (b) stages; the right plots show the distributions of errors between the target value and the MLP output signal at the training (c) and testing (d) stages

The files of two types were formed on the basis of data generated with the help of the GEANT3 code. The former included the information on ionization energy losses by pions in $n = 12$ modules of the TRD, and the latter involved information on energy losses by electrons, including the transition radiation. Each file included 1800 events*.

The three-layered perceptron from the package JETNET3 [14] has been used for particle identification. The network included $n = 12$ input neurons (according to the number of absorbers in the TRD), 35 neurons in the hidden layer and one output neuron. It was assumed that for pion events the output signal must equal –1, and for electron events — +1. To estimate the quality of the MLP training, we assumed that the network correctly identified the event given to the input, if

*By event we mean a sample of the volume n composed from energy losses of pion or electron detected by the TRD.

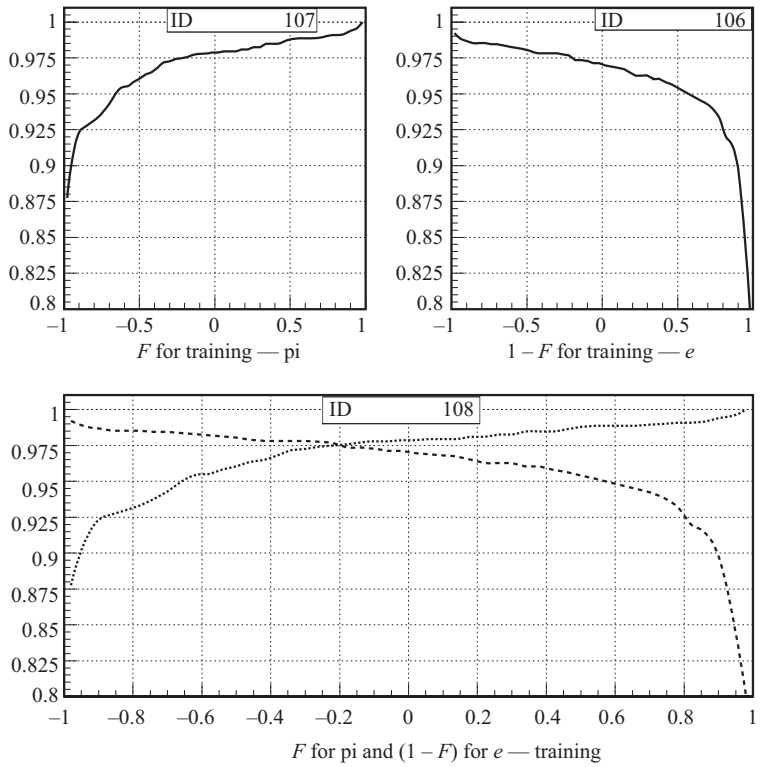

Fig. 11. The cumulative probability (for the MLP training stage) $F(y_t) = P_r(y < y_t)$ for pion and the dependence $1 - F(y_t)$ for electron events; the bottom plot shows the summary dependence for pions and electrons

the absolute error between the output signal and the target value did not exceed 0.05. An algorithm of the backward error propagation has been used for the error functional minimization at the stage of ANN training [9].

Initially, the events were formed using the set of energy losses ΔE_i, $i = 1, ..., n$ corresponding to the passage through the TRD pions or electrons. Figure 7 shows distributions of energy losses (including transition radiation) by electrons (a) and energy losses by pions (b) in the first absorber of the TRD detector. The distributions of energy losses in other TRD absorbers are of similar character.

In spite of the fact that the distribution of energy losses by electrons, significantly differs from the character of the distribution of energy losses by pions, for such a choice of input data the training process was going on very slow (see bottom curve in Fig. 8), there were large fluctuations (against the trend) of the efficiency of events identication by the network. Moreover, in spite of a

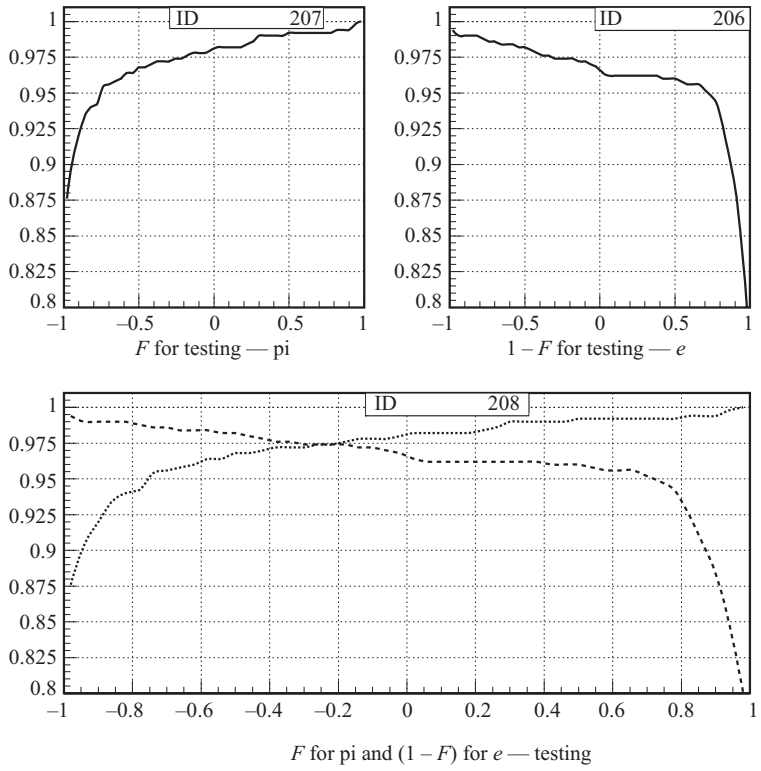

Fig. 12. The cumulative probability (for the MLP testing stage) $F(y_t) = P_r(y < y_t)$ for pion and the dependence $1 - F(y_t)$ for electron events; the bottom plot shows the summary dependence for pions and electrons

large number of training epochs, one can not reach the needed level of pions suppression.

In this connection, the sets of a new variable λ were formed on the basis of the original samples (see details in [15]):

$$\lambda_i = \frac{\Delta E_i - \Delta E^i_{mp}}{\xi_i} - 0.225, \qquad i = 1, 2, ..., n, \tag{8}$$

where ΔE_i is the value of energy loss in the ith TRD absorber, ΔE^i_{mp} is the most probable value of energy loss, $\xi_i = \frac{1}{4.02}$ FWHM of distribution of energy loss for pion in the ith absorber [16].

In order to determine the value of most probable energy loss ΔE^i_{mp} and the value FWHM of distribution of energy losses by pions in the ith absorber, the indicated distributions were approximated by the density function of a log-normal

distribution (see Fig. 9)

$$f(x) = \frac{A}{\sqrt{2\pi}\sigma x} \exp^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}, \tag{9}$$

where σ is the dispersion, μ is the mean value, and A is a normalizing factor [16].

As one can see from Fig. 9, the distribution of energy losses of pions quite well follows the distribution (9).

The sample of obtained values $\lambda_i,\ i = 1, ..., n$ was ordered due to values $(\lambda_j,\ j = 1, ..., n)$ and for each of them the values of Landau distribution function $\phi(\lambda)$ with the help of the DSTLAN function (from the CERNLIB library [17]), which was used to form the input pattern for the network, were calculated.

The described procedure of the initial data transformation permits one to pass from the problem of classification of samples which belong to one of two different overlapping distributions to the problem of classification of empirical distributions corresponding to pion and electron events. In this case, the reliable level of pion/electron identification by the network is reached after 10–20 training epoches in conditions of practical absence of fluctuations against the trend, and very quickly the needed level of pions suppression under the condition of a minimal loss of electrons is reached (see the behaviour of the top curve in Fig. 8).

Figure 10 presents the distributions of values of the MLP output signals obtained at the training (a) and testing (b) stages; the right plots show the distributions of errors between the target value and the MLP output signal at the training (c) and testing (d) stages.

Figure 11 shows the dependence (for the training stage) of the cummulative probability $F(y_t) = P_r(y < y_t)$ for pion and the dependence $1 - F(y_t)$ for electron events; the bottom plot shows the summary dependence for pions and electrons.

At the stage of the MLP testing the event type was determined by the value of the output signal y: when it does not exceed the preassigned threshold y_t, then the event was assumed to belong to pion, in the opposite case — to electron. Figure 12 shows the acummulative probability $F(y_t) = P_r(y < y_t)$ for pion and the dependence $1 - F(y_t)$ for electron events; the bottom plot shows the summary dependence for pions and electrons.

For threshold $y_t = 0.84$ the error of the first order α — part of electron events interpreted as pion events — was 9.4%, and the error of the second order β — part of pion events interpreted as electrons — was 0.6%. Thus, the suppresion of pion events is equal to 167. In the case, when we do not apply the above-described transformation of the original data, $\alpha = 11\%$, $\beta = 2.5\%$, and the suppression of pion events will consist of approximately 40.

CONCLUSION

We reported the electron/pion identification using energy losses in 12 layers of the CBM TRD applying a feed-forward ANN. The J/ψ phase space distribution and decay kinematics were calculated with the PLUTO event generator for beam energy 25 AGeV. The background was calculated with UrQMD event generator for the same beam energy. For analysis only events with one J/ψ electron daughter mixed with one minimum bias UrQMD event were used. As the input data for the MLP we used both the samples of energy losses of electrons/pions in the TRD absorbers and the new variables obtained on the basis of initial samples. We show that when one uses the transformed samples, then there are practically no fluctuations in the efficiency of particle recognition and after 10–20 training epoches the MLP reaches a reliable level of analyzed patterns recognition. If for the former case with approximately 11% of electrons loss one may reach the pion suppression around 40, then in the latter case for lower level of electrons loss we get the pions suppression at the level 170.

The application of the ANN method for simulated full physics events in the CBM TRD is under investigation.

REFERENCES

1. *Andronic A. at al.* // Nucl. Instr. Meth. A. 2004. V. 519. P. 508.
2. Letter of Intent for the Compressed Baryonic Matter Experiment; http://www.gsi.de/documents/ DOC-2004-Jan-116-2.pdf
3. Compressed Baryonic Matter Experiment. Technical Status Report, GSI, Darmstadt, 2005 (http://www.gsi.de/onTEAM/dokumente/public/DOC-2005-Feb-447 e.html).
4. *Nauman Th.* Multidimensional Data Analysis // «Formulae and Methods in Experimantal Data Evaluation», vol. 3 (European Physical Society, Geneve, 1983).
5. *Denby B.* Tutorial on Neural Networks Applications in High Energy Physics: 1982 Perspective // Proc. of the Second Intern. Workshop on Software Engineering, Artifical Intelligence and Expert System in High Energy Physics. France, 1992. P. 287.
6. *Kisel I. V., Neskoromnyi V.Ṅ., Osokov G. A.* Application of Neutral Networks in Experimental Physics // Part. & Nucl. 1993. V. 24. P. 1551.
7. *Ivanov V. V. et al.* JINR Preprint P10-94-300. Dubna, 1994.
8. *Fogelman S. F.* Neural Networks for Patterns Recognition: Introduction and Comparison to Other Techniques // Proc. of the Second Intern. Workshop on Software Engineering, Artifical Intelligence and Expert System in High Energy Physics. France, 1992. P. 277.

9. *Rumelhart D. E., Hinton G. E., Williams R. J.* Learning Internal Representations by Error Propagation in D. E. Rumelhart, J. L. McClelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol.1: Foundations. MIT Press, 1986.

10. http://www-hades.gsi.de/computing/pluto/html/PlutoIndex.html

11. *Bleicher M., Zabrodin E., Spieles C. et al.* Relativistic Hadron–Hadron Collisions in the Ultra-Relativistic Quantum Molecular Dynamics Model (UrQMD) // J. Phys. G. 1999. V. 25. P. 1859.

12. ROOT — An Object-Oriented Data Analysis Framework, User's Guide V. 5.08, December 2005.

13. GEANT — Detector Description and Simulation Tool, CERN Program Library, Long Write-up, W5013 (1995).

14. *Peterson C., Rögnvaldsson Th., Lönnblad L.* JETNET 3.0 — A Versatile Artificial Neural Network Package // Comput. Phys. Commun. 1994. V. 81. P. 185.

15. *Zrelov P. V., Ivanov V. V.* The Relativistic Charged Particles Identification Method Based on the Goodness-of-Fit ω_n^3-Criterion // Nucl. Instr. and Meth. in Phys. Res. A. 1991. V. 310. P. 623–630.

16. *Eadie W. T., Dryard D., James F. E., Roos M., Sadoulet B.* Statistical Methods in Experimental Physics, North-Holland Pub.Comp., Amsterdam-London, 1971.

17. *Koelberg K. S.* CERN Computer Centre Program Library, G110.

Received on February 2, 2007.

Редактор *В. В. Рудниченко*

Подписано в печать 10.05.2007.
Формат 60 × 90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 1,13. Уч.-изд. л. 1,59. Тираж 290 экз. Заказ № 55771.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.
E-mail: publish@jinr.ru
www.jinr.ru/publish/