Particle identification using artificial neural networks at BESIII *

QIN Gang(秦纲)[1,2;1)] LÜ Jun-Guang(吕军光)[1;2)] HE Kang-Lin(何康林)[1;3)] BIAN Jian-Ming(边渐鸣)[1,2]
CAO Guo-Fu(曹国富)[1,2] DENG Zi-Yan(邓子艳)[1] HE Miao(何苗)[1,2] HUANG Bin(黄彬)[1,2]
JI Xiao-Bin(季晓斌)[1] LI Gang(李刚)[1,3] LI Hai-Bo(李海波)[1] LI Wei-Dong(李卫东)[1]
LIU Chun-Xiu(刘春秀)[1] LIU Huai-Min(刘怀民)[1] MA Qiu-Mei(马秋梅)[1] MA Xiang(马想)[1,2]
MAO Ya-Jun(冒亚军)[4] MAO Ze-Pu(毛泽普)[1] MO Xiao-Hu(莫晓虎)[1] QIU Jin-Fa(邱进发)[1]
SUN Sheng-Sen(孙胜森)[1] SUN Yong-Zhao(孙永昭)[1,2] WANG Ji-Ke(王纪科)[1,2]
WANG Liang-Liang(王亮亮)[1,2] WEN Shuo-Pin(文硕频)[1] WU Ling-Hui(伍灵慧)[1,2]
XIE Yu-Guang(谢宇广)[1,2] YOU Zheng-Yun(尤郑昀)[4] YANG Ming(杨明)[1,2]
YU Guo-Wei(俞国威)[1] YUAN Chang-Zheng(苑长征)[1] YUAN Ye(袁野)[1]
ZANG Shi-Lei(臧石磊)[1,3] ZHANG Chang-Chun(张长春)[1] ZHANG Jian-Yong(张建勇)[1,3]
ZHANG Ling(张令)[5] ZHANG Xue-Yao(张学尧)[6] ZHANG Yao(张瑶)[6]
ZHU Yong-Sheng(朱永生)[1] ZOU Jia-Heng(邹佳恒)[6]

1 (Institute of High Energy Physics, CAS, Beijing 100049, China)
2 (Graduate University of Chinese Academy of Sciences, Beijing 100049, China)
3 (CCAST(World Laboratory), Beijing 100080, China)
4 (Peking University, Beijing 100871, China)
5 (Hunan University, Changsha 410082, China)
6 (Shandong University, Jinan 250100, China)

Abstract A multilayered perceptrons' neural network technique has been applied in the particle identification at BESIII. The networks are trained in each sub-detector level. The NN output of sub-detectors can be sent to a sequential network or be constructed as PDFs for a likelihood. Good muon-ID, electron-ID and hadron-ID are obtained from the networks by using the simulated Monte Carlo samples.

Key words artificial neural networks, particle identification, PID variables, multilayered perceptrons

PACS 01.50.Pa, 02.50.Sk, 07.05.Mh

1 Introduction

Particle identification (PID) will play an essential role in most of the BESIII physics program[1]. Good μ/π separation is required in precise $f_{\mathrm{D}}/f_{\mathrm{D_s}}$ measurements. Excellent electron ID will help to improve the precision of CKM elements V_{cs} and V_{cd}. The identification of hadronic (π/K/p) particles is the most common tool in BESIII physics analysis, sometimes it is the most crucial tool for the analysis.

Each part of the BESIII detector executes its own functions and provides a vast amount of information which determines the final efficiency of particle identification. The PID ability is quite different for each sub-detector in different momentum range. To improve the PID performance, a powerful technique is required to wisely combine these information together, especially when some of these information

Received 14 February 2007, Revised 7 March 2007

* Supported by CAS Knowledge Innovation Project (U-602(IHEP), U-34(IHEP)), National Natural Science Foundation of China (10491300, 10605030), 100 Talents Program of CAS (U-54, U-25)

1) E-mail: qing@ihep.ac.cn
2) E-mail: lujg@ihep.ac.cn
3) E-mail: hekl@ihep.ac.cn

are highly correlated. In recent years, a couple of PID algorithms have been developed: the likelihood method[2], the Fisher discriminator[3], the H-Matrix estimator[4], the Artificial Neural Network[5], and the Boosted Decision Tree[6], and so on. In this paper, we present an application of the powerful artificial neural network (ANN) method to BESIII PID algorithm.

2 The PID system of BESIII

The BESIII detector[1, 7] consists of a beryllium beam pipe, a helium-based small-celled drift chamber, Time-Of-Flight (TOF) counters for particle identification, a CsI(Tl) crystal calorimeter, a superconducting solenoidal magnet with the field of 1 T, and a muon identifier of Resistive Plate Counters (RPC) interleaved with the magnet yoke plates. The preliminary version of the BES Offline Software System (BOSS)[8] has been implemented successfully. The detector simulation[9] is based on Geant4[10]. A tremendous amount of software work has been accomplished but much remains to be done.

2.1 The dE/dx measurements

The Main Drift Chamber (MDC) consists of 43 layers of sensitive wires and works with a 60%/40% He/C_3H_8 gas mixture. The expected momentum resolution σ_p/p is about 0.5% @1 GeV/c. The energy loss in the drift chamber can provide additional information on particle identification. The normalized pulse height, which is proportional to the energy loss of incident particles in the drift chamber, is a function of $\beta\gamma = p/m$, where p and m are the momentum and mass of a charged particle. Fig. 1(a) shows the normalized pulse heights varying with the momentum for different particle species. The electron, muon and pion can not be well separated around 0.2 GeV/c by using dE/dx pulse heights. Similarly, the dE/dx pulse height will fail to separate the electron from the kaon around 0.5—0.6 GeV/c.

There are a lot of factors which can affect the dE/dx measurements[11]: the number of hits, the average pass lengthes in each cell, the space charge and saturation effects, the non-uniformity of electric fields, and so on. After the calibration, the resolution of dE/dx measurements is expected to be 6%—7%. Using dE/dx information, a 3σ K/π separation can be achieved up to 0.6 GeV/c; Good e/π separation can be obtained above 0.4 GeV/c.

2.2 The TOF counter

Outside the MDC is the TOF system, which is crucial for particle identification. It consists of a two-layer barrel array and one layer endcap array. There are two readout PMTs on each barrel scintillator and one on endcap scintillator. The expected time resolution for two layers is from 100 to 110 ps for K and π, giving a 2σ K/π separation up to 0.9 GeV/c.

The physics goal of TOF system is to measure the flight time of charged particle. The velocity (βc) and mass (m) of the charged particle can be calculated by

$$\beta = \frac{L}{c \times t_{\rm mea}}, \quad m^2 = p^2 \times \frac{1-\beta^2}{\beta^2}, \tag{1}$$

where $t_{\rm mea}$ is the measured time-of-flight, L and p are the corresponding flight path and momentum of the charged particle given by the MDC measurements, and c is the velocity of light in vacuum. The typical mass square distributions for the electron, pion, kaon and proton in different momentum range are drawn in Fig. 1(b).

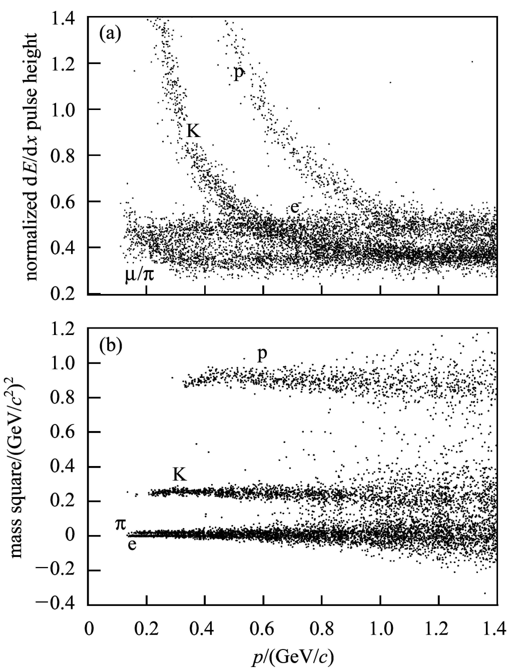

Fig. 1. (a) The normalized pulse heights (dE/dx) vs. momentum of charged particles; (b) The mass square distribution from TOF measurements.

The PID ability relies on the time resolution (σ_t) of the TOF system. The σ_t depends on the pulse height, hit position, and the beam status. Usually the value of σ_t varies in different TOF counters due to the different performances of the scintillator, PMT, and electronics. Since the distinctive TOF measurements correlate due to the common event start time, the weighted time-of-flight for two layers is obtained by a correlation analysis depicted in Ref. [12].

2.3 The CsI(Tl) calorimeter

The CsI(Tl) crystal electromagnetic calorimeter (EMC) contains 6240 crystals, and is used to measure the energy of photons precisely. The expected energy and spatial resolutions are 2.5% and 0.6 cm @1 GeV, respectively. The character of electromagnetic shower is distinctive for electron, muon and hadron, the energy deposit and the shape of shower in calorimeter can be used as discrimination variables to do PID.

In the CsI(Tl) crystal, the energy deposit by ionization is about 0.165 GeV for the charged particles passing at normal incidence through the EMC. Electron and positron lose all of their energies in the calorimeter by producing the electromagnetic showers, the ratio of deposit energy to the track momentum (E/p) will be approximately unity. Sometimes the energy deposit of hadrons will have an E/p ratio higher than that of the expected by ionization due to the nuclear interaction with materials. Fig. 2(a) shows the energy deposit vs momentum for e, μ and π in EMC.

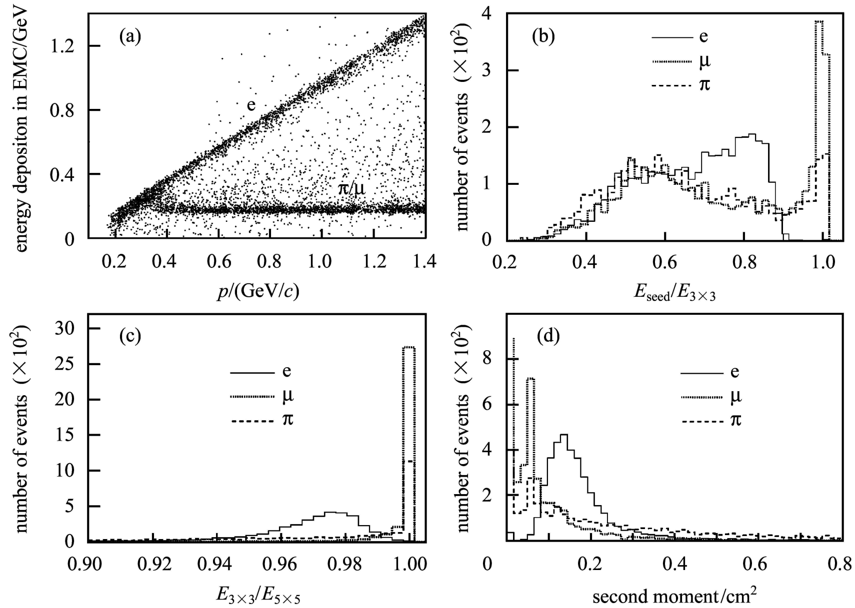

Fig. 2. (a) Energy deposit in EMC vs. the momentum for e, μ and π; (b) Ratio of $E_{\mathrm{seed}}/E_{3\times3}$ for e, μ and π; (c) Ratio of $E_{3\times3}/E_{5\times5}$ for e, μ and π; (d) Second-moment S distribution for e, μ and π.

The "shape" of shower can be characterized by the three energies: E_{seed}, the energy deposited in the central crystal; $E_{3\times3}$, the energy deposited in the central 3×3 crystal array; and $E_{5\times5}$, the energy deposited in the central 5×5 crystal array. The ratios of $E_{\mathrm{seed}}/E_{3\times3}$ and $E_{3\times3}/E_{5\times5}$ for e, μ and π at 1 GeV/c are plotted in Fig. 2(b) and 2(c), respectively.

The second-moment S is defined as

$$S=\frac{\sum_i E_i\cdot d_i^2}{\sum_i E_i}\,, \tag{2}$$

where E_i is the energy deposit in the i-th crystal, and d_i is the distance between the i-th crystal and the center position of reconstructed shower. The original idea of S was developed by the Crystal Ball experiment[13] to distinguish the cluster generated by π^0 and γ. The S distributions for e, μ and π at 1 GeV/c are shown in Fig. 2(d).

2.4 The muon system

The magnet return iron has nine layers of Resistive Plate Chambers (RPC) in the barrel and eight layers in the endcap to form a muon counter. All the

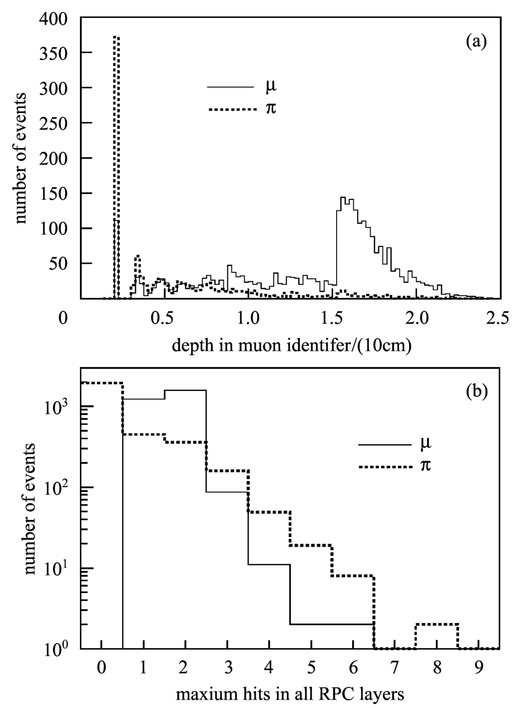

Fig. 3. (a) The travel depth of μ and π in muon counter; (b) The maximum number of hits for μ and π in all RPC layers.

RPCs have been manufactured, tested and installed. The spatial resolution obtained is about 16.6 mm.

The energy of electron which is exhausted in the calorimeter, cannot reach the muon counter. Most of the hadrons passed the material of calorimeter, magnet, and would be absorbed in the return irons. Muons have a quite strong punching ability. Usually they will produce one hit in each layer, hadrons may produce many hits in a certain layer if an interaction occurs. The distance of muon hit to the extrapolated position of inner track will be helpful to reduce the hadron contamination to a lower level, since the hits generated by the secondary muon from the decay of π/K cannot match the inner track very well. Fig. 3 shows the distributions of travel depth and the maximum number of hits in all RPC layers for μ and π with the momentum distributing in 0.8—1.5 GeV/c.

3 Brief description of the artificial neural network

An artificial neural network[5] is a computational structure inspired by the study of biological neural processing. Feed-forward neural networks, also known as multilayered perceptrons, are most popular and widely used. The output of a feed-forward neural network trained by minimizing, for example, the mean square error function, directly approximates the Bayesian posterior probability without the need to estimate the class-conditional probabilities separately. A feed-forward neural network (NN) is shown schematically in Fig. 4. Such networks provide a general framework for estimating non-linear functional mapping between a set of input variables $\boldsymbol{x}(x_1,x_2,\cdots,x_N)$ and an output variable $O(\boldsymbol{x})$ (or a set of output variables) without requiring a prior mathematical description of how the output formally depends on the inputs.

The network is made of neurons characterized by a bias and weighted links in between, those of links are called synapses. A layer of neurons makes independent computations on the data, so that it receives and passes the results to another layer. The next layer may in turn make its independent computations and pass on the results to yet another layer. Finally, the processed results of the network can be determined from the output neurons. As indicated in the sketch, all neuron inputs to a layer are linear combinations of the neuron output of the previous layer. For a given neuron j in layer k, we have the following equation

$$x_j^k = A\left(w_{0j}^k + \sum_{i=1}^{M_{k-1}} w_{ij}^k \cdot x_i^{k-1}\right), \qquad (3)$$

where $x_i^{k-1}(i=1,2,\cdots,M_{k-1})$ represents the input signal from the previous layer $k-1$, M_{k-1} is the total number of neurons in layer $k-1$, w_{ij}^k's represent the synaptic weights of neuron j, the bias term w_{0j}^k (not shown in Fig. 4) is acquired by adding a new synapse to neuron j whose input is $x_{0j}^{k-1}=1$. The transfer from input to output within a neuron is performed by means of an "activation function" $A(x)$. In general, the activation function of a neuron can be zero (deactivated), one (linear), or non-linear. For a hidden layer, a typical activation function used in Eq. (3) is a sigmoid

$$A(x) = \frac{1}{1+\mathrm{e}^{-x}}\,. \qquad (4)$$

The transfer function of the output layer is usually linear. As a consequence: an neural network without hidden layer should give identical discrimination power as a linear discriminant analysis like the Fisher discriminator. In case of one hidden layer, the neural network computes a linear combination of sigmoid.

The number of parameters (the synaptic weights w_{ij}^k's in Eq. (3)) need to grow only as the complexity of the problem grows. The parameters are determined by minimizing an error function, usually the mean square error between the actual output O^p and the desired (target) output t^p,

$$E = \frac{1}{2N_p}\sum_{p=1}^{N}(O^p - t^p)^2\,, \qquad (5)$$

with respect to the parameters. Here p denotes a feature vector or pattern. The stochastic optimization algorithm used in learning enables the model to be improved a little bit for each data point in the training sample. Neural networks provide a very practical tool because of the relatively small computational times required in their training. The fast convergence as well as the robustness in supervised learning of multilayered perceptrons are due to efficient and powerful algorithms developed in recent years.

Good generalization, that is good predictions for new inputs, is controlled by model complexity, The traditional approaches used to control model complexity are structure stabilization (optimizing the size of the network) and regularization. In the former one starts with large networks and prunes connections or starts with small networks and adds neurons as necessary. In the latter, one penalizes complexity by adding a penalty term to the error function.

There are many new and sophisticated approaches to achieve good generalization. Here, it is important to note that the generalization error (g.e.) of an NN can be decomposed into the sum of the bias-squared (b^2) plus the variance (σ^2), i.e., the generalization error, g.e.$=\sqrt{b^2+\sigma^2}$. The goal is to minimize the g.e., that is, finding the best compromise between bias and variance. Ensembles of networks, such as committees or stacks, can be used to control bias and variance[5].

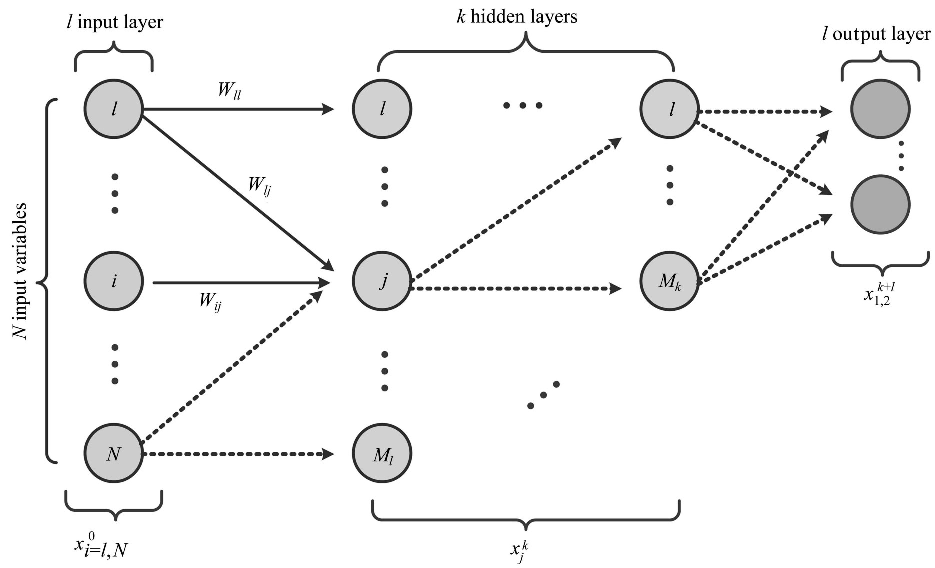

Fig. 4. The schematic structure of a multilayered perceptrons' neural network: the input layer contains N neurons as input variables ($x^0_{i=1,2,\cdots,N}$); the output layer contains (two) neurons for signal and background event classes; in between the input and output layers are a number of k hidden layers with arbitrary number of neurons ($x^k_{j=1,2,\cdots,M_k}$).

4 Applying the ANN technique in PID algorithm at BESIII

In this paper, a class of Multilayer Perceptrons (MLP)[14] network will be applied to BESIII PID algorithm, which is implemented in ROOT[15]. The PID variables described in Section 2.3 and Section 2.4 are correlated each other. With no loss of information, a cell analysis may be not sufficient for the likelihood method to get an optimal result. As we know, the correlations of PID variables among sub-detectors are reasonably small and could be ignored. For example, the $\mathrm{d}E/\mathrm{d}x$ measurement and the energy deposit in EMC have almost no influence on the time of flight measurements. This allows us to configure the networks sequentially.

4.1 The configuration of PID networks

The PID variables selected from each sub-detector together with the incident momentum and the transverse momentum have been grouped and trained separately, each sub-detector (the barrel part and the endcap part) has one output. In this step, the network for each sub-detector is quite simple. Almost all sub-networks are configured with one hidden layer containing $2N$ hidden neurons, where N is the number of the input neurons. 50 000 single track events for each particle species with momentum ranging from 0.1 GeV/c to 1.6 GeV/c and $-0.83<\cos\theta<0.83$ are subjected to networks, where θ is the incident polar angle. The output values are constrained to be 1, 2, 3, 4 and 5 for electron, muon, pion, kaon and proton, respectively. The training results for each sub-detector are shown in Figs. 5(a)—5(d). Since the offline software system for some of the endcap detectors are not well established now, here and below we just show the training results for the barrel part of detectors.

The bands of muon and pion are merged into one single peak (around 2.5) in the outputs of $\mathrm{d}E/\mathrm{d}x$, TOF and EMC. The information in EMC and MUC is hard to be applied to kaon and proton samples. The output of EMC can help to separate muon and pion slightly. But the output of MUC could separate muon from hadrons clearly.

The outputs of neural network from sub-detectors can be combined in several ways to get the near-optimal discrimination variables. For example, the probability density functions (PDF) for the resulting variables can be used as the basis for a likelihood analysis, or can be used as the input variables for a sequential network. At present, a conventional likelihood analysis based on the neural output variables and a sequential network analysis are applied in parallel to the BESIII PID algorithm. Here, we are only concerned with the study of the sequential networks. Basically, the network consists of two input momentum variables and four input PID variables. The momentum variables are the incident momentum and the transverse momentum. The PID variables include the neural outputs from dE/dx ($O_{\mathrm{d}E/\mathrm{d}x}$), TOF ($O_{\mathrm{TOF}}$), EMC ($O_{\mathrm{EMC}}$) and MUC ($O_{\mathrm{MUC}}$) system (the barrel part and the endcap part separately). The network is configured with one hidden layer of ten hidden neurons.

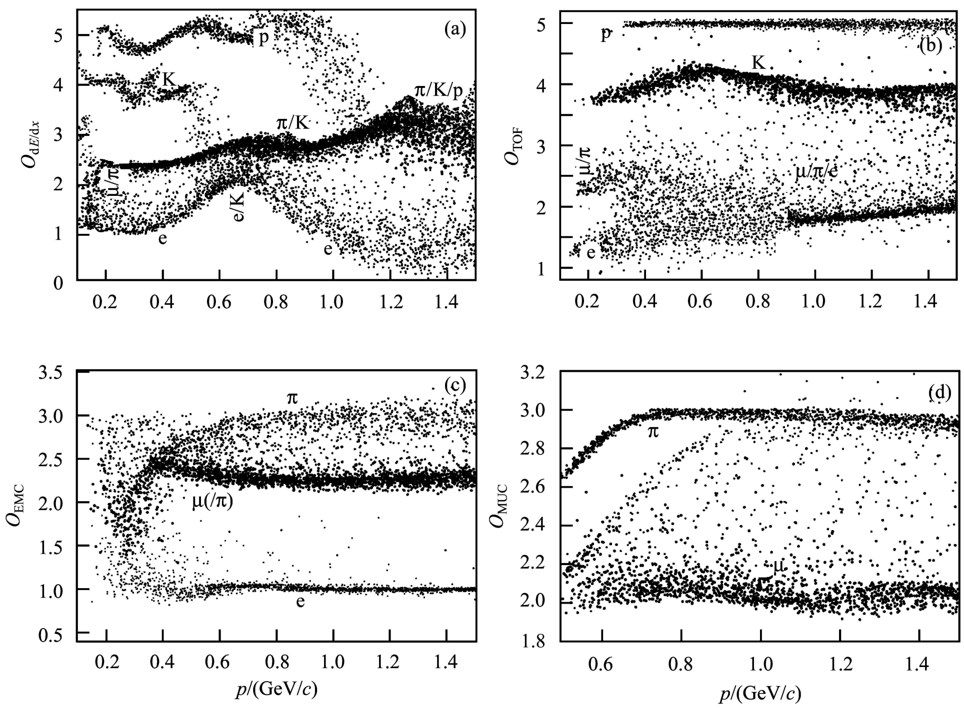

Fig. 5. NN outputs of sub-detectors. (a) dE/dx output; (b) TOF output; (c) EMC output; (d) MUC output.

4.2 Results

Electron, muon, and hadron separations are studied with several simulated Monte Carlo samples through different configurations of networks. Cuts are put on the output of final discrimination variables (the output of sequential network O_{seq}). The most PID interests in BESIII physics program are the e/π, μ/π and π/K separations.

4.2.1 Muon identification

The muon candidate is required to have response in μ-counter. The sequential network is trained with two PID variables: the O_{MUC} and the O_{EMC}. The μ-ID abilities are studied in different momentum partitions by comparing the discrimination results from O_{MUC} and O_{seq}. Fig. 6(a) and 6(b) show the variations of the muon identification efficiency and pion contamination rate as functions of incident track momentum, where the track momentum is required to be greater than the cut-off threshold (~500 MeV/c). Above 0.8 GeV/c, the muon identification efficiency is around 90%, and the pion contamination rate is about 5%. The additional information from EMC may help to improve the μ-ID ability.

As experienced in BaBar experiment[16], the additional variables, e.g., the goodness of muon track fit and the goodness of the muon track matching to the extrapolation position from inner track system, may help to reduce the background contamination rates. These information will be studied at BESIII in the future.

4.2.2 Electron identification

Figure 6(c) shows the variations of electron identification efficiency and pion misidentification rate at different momentum range by setting cuts on O_{EMC}. Above 0.6 GeV/c, one can see that the electron-ID efficiency is greater than 95% while the pion contamination rate can go down to the 10^{-3} level. The e/π separation is quite bad for low momentum tracks (less than 0.4 GeV/c).

Both $O_{\mathrm{d}E/\mathrm{d}x}$ and O_{EMC} are good discrimination variables to separate electron from pion above 0.4 GeV/c. The O_{TOF} can separate e/π very effectively below 0.3 GeV/c. The network is trained with $O_{\mathrm{d}E/\mathrm{d}x}$, O_{TOF} and O_{EMC}. As shown in Fig. 6(d), the network offers a nearly uniform acceptance and background contamination between 0.25 GeV/c and 1.6 GeV/c. It is an interesting result that the acceptance hole between 0.2 GeV/c and 0.4 GeV/c has almost vanished by adding an appropriate cut, where no detector has clear discrimination power for electron. The system obviously makes the inference that the particle has to be an electron if it is not one of the others. This is the combined contribution from the sub-detectors.

4.2.3 π/K separation

As a general view, the proton identification is extremely good with TOF and dE/dx information at BESIII. Here and below, only the K/π separation is focused on. As discussed in Section 2, the dE/dx can identify K/π very effectively below 0.6 GeV/c; the two-layer TOF can separate K/π up to 0.9 GeV/c.

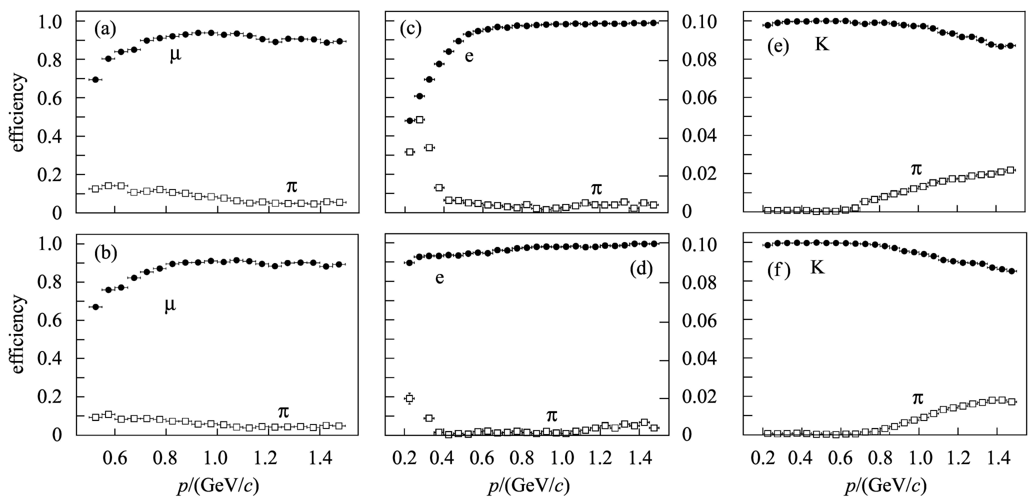

Fig. 6. PID efficiency and contamination rate in different momentum partitions. (a) μ/π separation with O_{MUC}; (b) μ/π separation with O_{MUC} and O_{EMC}; (c) e/π separation with O_{EMC}; (d) e/π separation with O_{EMC}, O_{TOF} and $O_{\mathrm{d}E/\mathrm{d}x}$; (e) K/π separation with likelihood method; (f) K/π separation with neural networks. In (c) and (d), pion contamination rates are enlarged by a factor of 10.

Traditionally, a likelihood method which combines the TOF and dE/dx information will be applied to the hadron identification. To construct the PDFs, data are divided into several bins in momentum and $\cos\theta$ partitions to obtain the corresponding resolutions and offsets. Fig. 6(e) shows the variations of kaon identification efficiency and pion contamination rate as functions of momentum. In the real world, there are often tails on distributions due to track confusion, nonlinearities in detector response, and many other experimental sources which are imperfectly described in PDFs. Herein, it is helpful to apply NN technique in hadron separation.

For hadron separation, the network is trained with two PID variables: the O_{TOF} and the $O_{\mathrm{d}E/\mathrm{d}x}$. The PID ability is studied with the cuts on the output of sequential network O_{seq}. The results are shown in Fig. 6(f). Both the likelihood method and the neural network method give the similar results. Below 1 GeV/c, one can see that the kaon-ID efficiency is greater than 95% while the pion contamination rate is less than 10%. The K/π separation is extremely good for low momentum tracks (less than 0.8 GeV/c).

5 Summary

The shapes of PID variables in EMC and MUC systems are complicated and there maybe exist nonlinear correlations between PID variables. It is difficult for the likelihood method to construct the PDFs analytically and handle the correlations properly. From our studies, good electron-ID and muon-ID can be easily achieved from the neural network at BESIII with the full detector information. In a simple application, e.g., for hadron separation, we get the similar results from the neural network and the likelihood method where the PID variables in TOF and dE/dx systems are quasi-Gaussian, and the correlation between two-layer TOF measurements is approximately linear. The flexible configuration of PID networks can be employed in the simple and complicated application.

There are still a lot of factors which have to be taken into account while applying the artificial neural network technique in the particle identification at BESIII. For example, one or several input variables have to be removed due to the imperfect consistency between data and Monte Carlo simulation. The impurity of training sample may introduce the additional systematical uncertainties. More detailed studies are needed in the future. Now the likelihood and the artificial neural network PID algorithms are studied in parallel at BESIII. The final PID algorithm will definitely be a combination of all the methods which have been created and applied in high energy physics experiment, one way might be by using the likelihood method to combine the neural network outputs from sub-detectors. Of course, to make the final decision of PID algorithm, there is still a long way to go.

References

1 BESIII Design Report. Interior Document in Institute of High Energy Physics, 2004

2 Carli T, Koblitz B. Nucl. Instrum. Methods, 2003, **A501**: 576—588; Holmström L, Sain R, Miettinen H E. Comput. Phys. Commun., 1995, **88**: 195

3 Fisher R A. The Use of Multiple Measurements in Taxonomic Problems, Annals of Eugenics 7. 1936, 179—188

4 Bhat P C (DØ Collaboration). 1995, arXiv:hep-ex/9507007

5 Bishop C M. Neural Networks for Pattern Recognition, Clarendon, Oxford, 1998; Beale R, Jackson T. Neural Computing: An Introduction. Adam Hilger, New York, 1991

6 Roe B P, YANG H J et al. Nucl. Instrum. Methods, 2005, **A543**: 577—584

7 Harris F A (BES Collab.). 2006, arXiv:physics/0606059

8 LI Wei-Dong, LIU Huai-Min et al. The Offline Software for the BESIII Experiment, Proceeding of CHEP06. Mumbai, India, 2006

9 DENG Zi-Yan et al. HEP & NP, 2006, **30**(5): 371—377 (in Chinese)

10 Agostinelli S et al. (Geant4 Collab.). Nucl. Instrum. Methods, 2003, **A506**: 250

11 WANG Da-Yong, HE Kang-Lin, LI Wei-Guo et al. Software Developments and Studies of BESIII dE/dx Reconstruction. In: The Publications for the 5th Seminar of Nuclear Instrument and Its Applications. China, 2005 (in Chinese)

12 HU Ji-Feng et al. HEP & NP, 2007, **31**(10): 893 (in Chinese)

13 Edwards C et al. Phys. Rev., 1982, **D25**: 3065

14 Brun R, Rademakers F et al. ROOT Users Guide 5.12. 2006, 73—76

15 Brun R, Rademakers F et al. An Object-Oriented Data Analysis Framework, http://root.cern.ch

16 Narsky I, StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data. 2005, arXiv:physics/0507143