PHYSICAL REVIEW D VOLUME 44, NUMBER 7 1 OCTOBER 1991

Quark- and gluon-jet separation using neural networks

I. Csabai, F. Czakó, and Z. Fodor
Eötvös University, Budapest, H-1088 Puskin u. 5-7, Hungary
(Received 23 January 1991)

We present a possible discrimination method by the combination of a neural network (NN) and QCD to separate the quark and gluon jets of e^+e^- annihilation. The network has been trained by jets of the same energy; thus only the internal structure of the jets has been considered. By the combination of the NN and the QCD matrix element, 92% accuracy in identifying jets has been achieved.

The basic particles in quantum chromodynamics (QCD) [1] are the quarks and gluons. Experiments (e.g., electron-positron annihilation) produce hadrons. These fragmentation processes may be described by phenomenological models. Uncertainties in these models present a major obstacle to the detailed testing of QCD. The separation of quark jets from gluon jets is a prerequisite to understand several phenomena. As has been emphasized, a fairly precise identification is needed, e.g., to establish the existence of the three-gluon vertex [2], or to determine the strong coupling constant using azimuthal asymmetries [3].

Gluons carry a stronger color charge than quarks, and one expects this to produce differences in their fragmentation, namely, higher multiplicity, softer hadron spectrum, and broader p_t distribution. There has been intensive theoretical study to calculate variables sensitive to the differences between quark and gluon jets [4], and to provide methods to analyze the experimental data and to see the above-mentioned differences between quark and gluon jets [5,6].

Studies on the fragmentation of quark and antiquark jets in electron-positron annihilation have been quite successful. Less knowledge has been accumulated from experiments on the fragmentation of high-energy gluon jets. Experimental studies encounter the problem that in e^+e^- annihilation quark jets dominantly appear in the two-jet ($q\bar{q}$) topology, but gluon jets turn up only in three or more jet topologies ($e^+e^- \rightarrow q\bar{q}g, \Upsilon \rightarrow ggg$). Because of these difficulties, most studies of gluon-jet fragmentation resulted in inconclusive results. One can study symmetric three-jet events and compare the results with the results of the two-jet events gained at $\frac{2}{3}$ c.m. energies [7]. There are several approaches based on studies of jets with different energies [8].

Despite the considerable amount of experimental effort, unambiguous differences between quark- and gluon-induced jets have not yet been established.

In recent papers a neural network (NN) method for identifying the parton ancestor of a jet has been suggested [9,10]. The method is able to separate gluon and quark jets originating from a Monte Carlo simulation of e^+e^- events with $\approx 85\%$ accuracy. The result is independent of the Monte Carlo model used. The network has access to the energy of the jets and, thus, can exploit the fact that gluon jets usually have a smaller energy.

The primary aim of this Rapid Communication is to present a possible quark/gluon jet discrimination method by combining the NN approach with QCD in electron-positron annihilation. (A possible generalization and details will be published elsewhere [11].) The network has been trained by quark and gluon jets of the same energy; thus, only the internal structure of a jet has been considered. The input data, based on the Monte Carlo event generator used, correspond to a total basis in the sense that all the longitudinal and transversal momenta of the outgoing hadrons with respect to the jet axis have been used in the analysis. On the other hand, the QCD matrix element at a definite jet energy also yields a probability for a jet being a gluon or a quark one.

These studies (e.g., [9,10] and the present one) are remarkably different from real experiments at least in one respect. In a Monte Carlo analysis it is easy to define the parton ancestor of a jet since the event generator is under control. One can store the momenta of the partonic state and then fragment the partons. Comparing the momenta of the partons and the jets, the partons may easily be assigned to the jets. In real experimental situations the parton ancestor is unknown (and in fact the problem is to determine it).

The new interdisciplinary methods of neural networks, in addition to their theoretical significance, have proven successful in various fields and provide an alternative method of data processing.

The self-organizing adaptive NN model used in this paper was introduced by Kohonen and has been described in detail in the author's book [12]. The modified version of the original model, the learning vector quantization (LVQ), has already proven to be more advantageous in certain statistical pattern recognition problems [13] compared to the widespread feed-forward model [14].

In general, the application of NN consists of three distinct periods: learning, testing, and generalization. In discrimination problems the learning phase corresponds to the construction of the discrimination function

With the present model, learning and testing are realized in a supervised way: jet data vectors along with the information on their established quark or gluon origin are given to the NN. In the testing phase one tests the performance of the discrimination function produced by the network with data never seen by the NN. The generalization, the active phase of the network, yields predictions for unknown samples. To perform a generalization on Monte Carlo data is of no significance; therefore, we encourage

 © 1991 The American Physical Society

the experimentalists to perform this phase and classify the available experimental data.

In the r category, d-dimensional LVQ model for any category c we have a fixed number (n_c) of d-dimensional reference vectors $\mathbf{k}_i^{(c)}(i=1,\ldots,n_c;\ c=1,\ldots,r)$. The learning period starts with the initialization of the reference vectors (e.g., the first n_c input data for each category are equal to $\mathbf{k}_i^{(c)}$). During the adaptive learning process the reference vectors are moved at time t according to the incoming data $\mathbf{x}^{(c)}(t)$. [Here the incoming data $\mathbf{x}(t)$ received at t belong to the category c.] The "movements" are organized by the learning rule in order to obtain an optimal discrimination with the *nearest-neighbor* method. The learning algorithm is the following.

Let $\mathbf{k}_w^{(m)}(t)$ be the reference vector closest to a given input $\mathbf{x}^{(c)}(t)$ in some (e.g., Euclidean) metric at time t:

$$\rho[\mathbf{k}_w^{(m)}(t)-\mathbf{x}^{(c)}(t)]=\min_{i,s}\rho[\mathbf{k}_i^{(s)}(t)-\mathbf{x}^{(c)}(t)]\,, \quad (1)$$

where ρ is the distance and the minimum is with respect to all reference vectors and for all categories. If $\mathbf{x}^{(c)}(t)$ belongs to the same category as the closest reference vector ($c=m$) then this reference vector is updated in the following way:

$$\mathbf{k}_w^{(m)}(t+1)=\mathbf{k}_w^{(m)}(t)+\alpha(t)[\mathbf{x}^{(c)}(t)-\mathbf{k}_w^{(m)}(t)]\,, \quad (2)$$

if the categories are not identical ($m\neq c$),

$$\mathbf{k}_w^{(m)}(t+1)=\mathbf{k}_w^{(m)}(t)-\alpha(t)[\mathbf{x}^{(c)}(t)-\mathbf{k}_w^{(m)}(t)]\,, \quad (3)$$

for all the other reference vectors,

$$\mathbf{k}_i^{(m)}(t+1)=\mathbf{k}_i^{(m)}(t)\,, \quad (4)$$

where $0<\alpha(t)<1$ and $\alpha(t)$ is decreasing monotonously with discrete time.

In the test phase the NN returns the category represented by the reference vector closest to the given input. This result may be compared to the correct answer to measure the classification performance of the NN.

We have illustrated these ideas in Fig. 1. The presented discrimination problem is taken from the real quark/gluon jet data. We have plotted jets having 30-GeV energy in two dimensions (transversal momentum fractions of the first versus the third particles). In the naive approach the discriminating curve is a straight line [Fig. 1(a)]. Following the method outlined above (e.g., having two-two reference vectors) the discriminating curve is not a straight line; it has two bends [Fig. 1(b)]. Consequently the discrimination performance improves. By increasing the number of the reference vectors the discriminating curve becomes more and more sophisticated further improving the success rate.

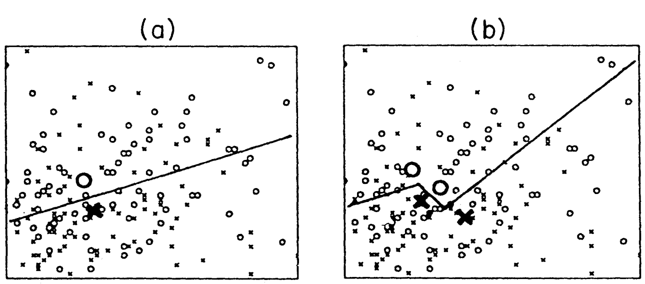

FIG. 1. Discrimination of 30 GeV quark and gluon jets (see text), (a) naive approach; (b) LVQ result with 2-2 reference vectors.

To study the properties of quark and gluon jets we have used Monte Carlo data generated by JETSET 6.3 [15]. The partonic configuration is calculated from perturbative QCD in the form of second-order matrix element for finite jet resolution parameter y_{min}. We have changed the QCD part of JETSET 6.3 to get rid of the approximation of Gutbrod *et al.* [16], by the method of the Mark J collaboration (for more details see [17]). In the Monte Carlo analysis fragmentation parameters fitted to data from the CERN e^+e^- collider LEP given by the OPAL Collaboration [18] have been used.

We have used 1 000 000 hadronic events at the Z^0 resonance. In the hadronic final states not only the charged, but neutral particles have been considered as well. Soft particles gaining energy less than 5% of the total energy of the jet have been neglected.

Having generated an event we applied the LUCLUS cluster algorithm [12] to determine the number, the energies, and the directions of the jets. Jets obtaining less than 5% of the total center-of-mass energy ($E_{\text{c.m.}}$) have always been excluded. Our Monte Carlo studies show that $\approx 99\%$ of the three-jet events is originated from three-parton events, there is only a very small 0.6% and 0.4% background from four- and two-parton events, respectively.

We have analyzed the following data samples.

(a) General three-jet events. A quark (antiquark or gluon) jet is defined as the jet closest to the quark (antiquark or gluon) parton, with the closeness defined as the relative angle between them. The minimal angle between two jets was set to 35°.

(b) In order to compare the results, particularly the results obtained by imposing a cut on the energy of the jets, with those of [10,11], we have also studied the JETSET 7.2 parton shower option with the default values of the generator. In this analysis the same cluster algorithm has been used as above. The d_{join} cut has been set to 2.5 GeV, with an additional constraint requiring exactly or at least three jets. A quark jet is defined as the jet closest to a quark with the closeness defined in the same way as in the clustering. Jets not assigned to a quark have been defined as gluon jets.

The NN input information we used are the following.

(a) The jet label (whether a particular jet is of quark or gluon origin).

(b) The energy (E_{jet}) and the invariant-mass square (Q_{jet}^2) of the jet.

(c) The longitudinal p_l/p_{jet} and transversal p_t/p_{jet} momenta fractions of all the outcoming particles. Previously mentioned soft particles have been rejected.

Fifty reference vectors have been used in the network analysis. In each run reference vectors have been associated in equal number with quark and gluon categories. Initial reference values for the vectors were picked out from the input data; namely, the first 25 quark and 25 gluon data were chosen.

Test results show decreasing success rate fluctuations when the number of the reference vectors have been increased. The accuracy of the prediction monotonically increases with the number of learning steps and seems to saturate after about 8000 steps (Fig. 2). 8000 patterns have been used in the test period to determine the network's performance.

The performance of a given method is presented in the form 84% (76/88). This notation indicates that the method gives 76% and 88% identification rates for gluon and quark jets, respectively, while the weighted average (for all the quark and gluon jets used in the analysis) identification rate is 84%.

Quark and gluon jet separation. As gluons are produced in bremsstrahlunglike processes, they usually have smaller energies than the primary quark jets. Imposing a cut for our three-jet sample at 23 GeV (jets with energy larger than 23 GeV are called quark jets, while others gluon jets) results in a success rate of 88% (77/93).

We have checked this simple energy cut method for the reproducible multijet sample of [10] (JETSET 7.2, minimal angle between jets 40°, $d_{\text{join}}=2.5$ GeV, requiring at least three jets, neglecting jets with less than 5% of the total $E_{\text{c.m.}}$, 5400 hadronic events). They report 84% success rate with a feedforward NN with one hidden layer. Surprisingly, the simple energy cut leads to a slightly higher result (85%). In our opinion their network has just almost learned the fact that jets with small energies are usually gluons, while jets with larger energies are the quark ones, ignoring the rest of the input information.

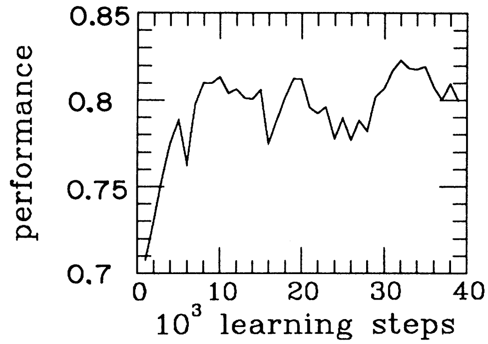

FIG. 2. The success rate of the LVQ method (30 GeV jets) vs the number of learning steps.

The underlying perturbative QCD matrix element gives the $p_g(x_a|x_b,x_c)$ probability that a jet (in a three-jet event) with a given $x_a=2E_a/E_{\text{c.m.}}$ energy fraction is a gluon jet (the two other jets have x_b and x_c energy fractions). For instance, in first order of α_s this probability is clearly

$$p_g^{(\text{QCD})}(x_a|x_b,x_c)=\frac{x_b^2+x_c^2}{(1-x_b)(1-x_c)}\left[\frac{x_a^2+x_b^2}{(1-x_a)(1-x_b)}+\frac{x_b^2+x_c^2}{(1-x_b)(1-x_c)}+\frac{x_c^2+x_a^2}{(1-x_c)(1-x_a)}\right]^{-1}. \tag{5}$$

One can use this probability (5) to identify jets, e.g., if $p_g^{(\text{QCD})}>\frac{1}{2}$ then a jet is said to be a gluon; otherwise it is said to be quark. This identification method gives an 87% (77/92) performance for hadronic three-jet events.

Our first NN approach was to teach our network the data set including energy, invariant-mass squared, and all the transversal and longitudinal momenta of the six most energetic particles. The performance has been quite independent of the number of learning steps and of the number of particles in the input data.

The obtained success rate of 86% does not exceed the result where no NN has been used, only the energy as a cut has been considered. It seems that the NN has picked out energy as the most important selection factor rather than the hidden information contained in other details. Therefore the failure of this type of brute-force approach is confirmed. Thus a different approach, based on a separate study of kinematic properties and the internal structure of jets, is needed.

To exclude energy from the NN input data the jets have been sorted into ten energy classes (from 5 to 50 GeV by 5 GeV) of 8000 jets each (4000 quark and 4000 gluon ones). Learning and testing were performed separately for each energy class.

Thus the NN has been able to utilize only the information contained in the Q_{jet}^2 and in the momenta fractions of the particles in the jets. The reference vectors were expected to represent and find the fine details hidden in the data.

The above procedure resulted in a success rate of 70% as a weighted average over all energy categories. This result seemed to be rather low but the information yielded by the NN could be considered to be independent of the energy value based QCD categorization. Therefore to improve the performance one would try to combine the information yielded by the NN and the underlying QCD matrix element. The NN approach is a nearest-neighbor method with discrete output, thus the first step is to extend it to continuous values in order to obtain the probability for a given jet being of quark or gluon origin.

It is straightforward to define the variable R for any jet given by $\mathbf{x}$:

$$R=\frac{\rho(\mathbf{x}-\mathbf{k}_w^{(q)})-\rho(\mathbf{x}-\mathbf{k}_w^{(g)})}{\rho(\mathbf{x}-\mathbf{k}_w^{(q)})+\rho(\mathbf{x}-\mathbf{k}_w^{(g)})}, \tag{6}$$

where $k_w^{(q/g)}$ is the nearest quark or gluon reference vector to the input $\mathbf{x}$ [as in Eq. (1)]. R falls within $[-1,1]$.

We calculated R values for the gluon- and quark-jet test samples separately. $f_q(R)$ and $f_g(R)$ denote the obtained density functions shown in Fig. 3(a). These functions may be used to calculate a probabilistic value for a given input with a certain R value. The input jet is expected to be a quark jet with a probability of $p_q^{(\text{NN})}$ and a gluon jet with probability $p_g^{(\text{NN})}$:

$$p_q^{(\text{NN})}(R)=\frac{f_q(R)}{f_q(R)+f_g(R)},$$
$$p_g^{(NN)}(R)=\frac{f_g(R)}{f_q(R)+f_g(R)}, \tag{7}$$

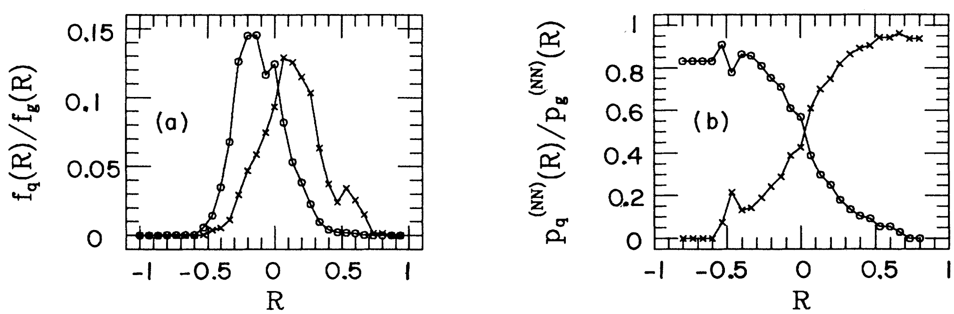

FIG. 3. (a) Quark and gluon density functions in R; (b) probability of being a quark or gluon jet calculated from Eq. (7).

respectively. These probabilities are plotted in Fig. 3(b) as a function of R.

Provided the probabilities $p_q^{(QCD)}$ by (3) and $p_q^{(NN)}$ by (8) are independent one easily gets the combined $\tilde{p}_q$ probability for an input to be a quark jet:

$$\tilde{p}_q = \frac{p_q^{(QCD)} p_q^{(NN)}}{1 - p_q^{(QCD)} - p_q^{(NN)} + 2 p_q^{(QCD)} p_q^{(NN)}} . \tag{8}$$

The above combination of the probabilities with a $\tilde{p}_q = \frac{1}{2}$ cut has led to a 92% success rate. This indicates that the optimal way is to combine the information of the fragmentation properties and the underlying QCD matrix element.

The increase from 87% success rate (only QCD matrix element approach) to the present 92% accuracy does not seem to be a very significant change; however, this result means that the misidentification rate is 62% higher either in the QCD method or in the brute-force NN approach (e.g., our first approach or [10]) compared to the combined one.

We have applied conventional, NN, and combined QCD-NN methods to separate quark and gluon jets.

The conclusions may be summarized as follows.

(i) The kinematic separation based on an energy cut or $p_{q/g}^{(QCD)}$ results in $\approx 87\%$–88% identification success rate.

(ii) The NN alone can reach $\approx 86\%$. Thus the performance of the method based on kinematic variables and the naive NN are about the same. This is the reason why the success rate in [9,10] with NN is not higher than that with the kinematic cut for the same event sample. Clearly what the NN does is nothing more than realize this simple kinematic information.

(iii) We have sorted the data into ten equidistant energy categories (from 5 to 50 GeV) and analyzed them independently with the NN. Combining these results with the underlying QCD probabilities based on the kinematical configuration of the jet system we have reached 92% success rate, a prediction accuracy superior to previous work.

The whole NN algorithm combined with the QCD method is very easy to implement in parallel hardware; thus, it offers a very promising method for on-line triggering in high-energy experiments.

The authors would like to thank F. Csikor and G. Pocsik for constructive discussions on particle physics. Many thanks go to T. Geszti for his inspiring ideas on neural networks.

[1] H. Fritzsch *et al.*, Phys. Lett. **47B**, 365 (1973); D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973); H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).

[2] M. Bengtsson and P. Zerwas, Phys. Lett. B **208**, 306 (1988); M. Bengtsson, Z. Phys. C **42**, 75 (1989); Z. Fodor, Phys. Rev. D **40**, 267 (1989); **40**, 3590 (1989); S. Bethke *et al.*, Z. Phys. C **49**, 59 (1991).

[3] A. Djouadi *et al.*, Phys. Lett. B **241**, 260 (1990).

[4] A. H. Mueller, Nucl. Phys. **B241**, 141 (1984); J. B. Gaffney and A. H. Mueller, *ibid.* **B250**, 109 (1985); G. S. H. Dzhaparidze, Z. Phys. C **32**, 59 (1986).

[5] H. P. Nilles and K. H. Streng, Phys. Rev. D **23**, 1944 (1981); O. Nachtmann, Z. Phys. C **16**, 257 (1983); L. M. Jones, Phys. Rev. D **39**, 2550 (1989); **42**, 811 (1990).

[6] Z. Fodor, Phys. Rev. D **41**, 1726 (1990).

[7] HRS Collaboration, M. Derrick *et al.*, Phys. Lett. **165B**, 449 (1985); Mark II Collaboration, A. Petersen *et al.*, Phys. Rev. Lett. **55**, 1954 (1985); TASSO Collaboration, W. Braunschweig *et al.*, DESY Report No. 89-032, 1989 (unpublished).

[8] JADE Collaboration, W. Bartel *et al.*, Phys. Lett. **123B**, 460 (1983); CELLO Collaboration, in *Lepton and Photon Interactions,* Proceedings of the International Symposium on Lepton and Photon Interactions at High Energies, Hamburg, West Germany, 1987, edited by R. Rückl and W. Bartel [Nucl. Phys. B (Proc. Suppl.) **3** (1987)]; AMY Collaboration, Y. K. Kim *et al.*, in *Proceedings of the XXV International Conference of High Energy Physics,* Singapore, 1990, edited by K. K. Phua and Y. Yamaguchi (World Scientific, Singapore, 1991).

[9] L. Lönnblad *et al.*, Phys. Rev. Lett. **65**, 1321 (1990).

[10] L. Lönnblad *et al.*, Nucl. Phys. **B349**, 675 (1991).

[11] I. Csabai *et al.* (work in progress); T. Sjöstrand and M. Bengtsson, Comput. Phys. Commun. **43**, 367 (1987).

[12] T. Kohonen, *Self-Organization and Associative Memory,* 2nd ed. (Springer, Berlin, 1988).

[13] G. Barna and K. Kaski, Phys. Scr. **T33**, 110 (1990).

[14] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing* (MIT, London, 1988).

[15] T. Sjöstrand, Report No. LU TP 86-22, 1986 (unpublished).

[16] F. Gutbrod *et al.*, Z. Phys. C **21**, 235 (1984).

[17] F. Csikor *et al.*, Mod. Phys. Lett. A **3**, 1177 (1988).

[18] OPAL Collaboration, M. Z. Akrawy *et al.*, Z. Phys. C **47**, 505 (1990).