

Classifying LEP Data with Support Vector Algorithms

P. Vannerem^a K.-R. Müller^b B. Schölkopf^b A. Smola^b
S. Söldner-Rembold^a

^a*Fakultät für Physik, Hermann-Herder-Str. 3, D-79104 Freiburg, Germany*

^b*GMD-FIRST, Rudower Chaussee 5, D-12489 Berlin, Germany*

Abstract

We have studied the application of different classification algorithms in the analysis of simulated high energy physics data. Whereas Neural Network algorithms have become a standard tool for data analysis, the performance of other classifiers such as Support Vector Machines has not yet been tested in this environment.

We chose two different problems to compare the performance of a Support Vector Machine and a Neural Net trained with back-propagation: tagging events of the type $e^+e^- \rightarrow c\bar{c}$ and the identification of muons produced in multihadronic e^+e^- annihilation events.

1 Classification algorithms

Artificial Neural Networks (ANN) are a useful tool to solve multi-dimensional classification problems in high energy physics, in cases where one-dimensional cut techniques are not sufficient. They are used both as hard-coded chip for very fast low-level pattern recognition in on-line triggering [1,2] and as a statistical tool for particle and event classifications in offline data analysis. In offline data analysis, a Monte Carlo simulation of the physics process and the detector response is necessary to train an ANN by supervised learning. ANN algorithms have been applied successfully in classification problems such as gluon-jet tagging [3] and b-quark tagging [4]. The ANN classifiers [5] constructed in this paper have sigmoid nodes. Design and tuning issues were solved by applying practical experience rules [6].

* To be submitted to the Proceedings of AIHENP99, Crete, April 1999

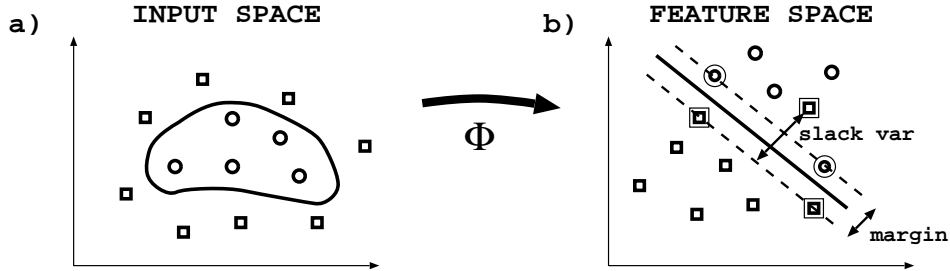


Fig. 1. A Support Vector Machine maps the input vectors of 2 different classes (a) into a higher dimensional space (b) and determines the separating hyperplane with maximum *margin* in that space. For non-separable problems additional *slack variables* are introduced.

This paper presents the application of a recently proposed machine-learning method, called Support Vector Machines (SVM) [7,8], to high energy physics data. The underlying idea is to map the patterns, i.e., the n -dimensional vectors \mathbf{x} of n input variables, from the input space to a higher dimensional feature space with a non-linear transformation (Fig. 1). Gaussian radial basis functions are used as a kernel for the mapping. After this mapping the problem becomes linearly separable by hyperplanes. The hyperplane which maximises the margin is defined by the support vectors which are the patterns lying closest to the hyperplane. This hyperplane which is determined with a training set is expected to ensure an optimal separation of the different classes in the data.

In many problems a complete linear separation of the patterns is not possible and additional slack variables for patterns not lying on the correct side of the hyperplane are therefore introduced. The training of a SVM [9,10] is a convex optimisation problem. This guarantees that the global minimum can be found, which is not the case when minimising the error function for an ANN with back-propagation. The CPU time needed to find the hyperplane scales approximately with the cube of the number of patterns.

2 Data sets and the OPAL experiment

Two distinctly different problems were chosen for comparing ANN and SVM classifiers, charm tagging and muon identification.

The first problem is to classify (tag) $e^+e^- \rightarrow q\bar{q}$ events according to the flavour of the produced quarks, separating c-quark events from light quark (uds) and b-quark events. Flavour tagging is necessary for precision measurements of electroweak parameters of the Standard Model. The events are divided into two hemispheres by a plane perpendicular to the thrust axis. The flavour tag is applied separately to both hemispheres, which contain the jets from the

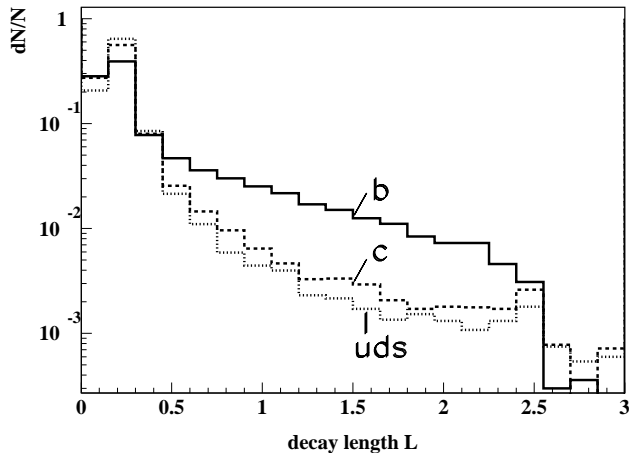


Fig. 2. The distribution of the reconstructed decay length L is an input distribution to the flavour tag.

two produced quarks. For a signal (s) with background (bg) the efficiency ε is defined as

$$\varepsilon = N_{\text{tag}}^s / N_{\text{tot}}^s,$$

where N_{tag}^s are the number of correctly tagged hemispheres and N_{tot}^s are all signal hemispheres in the sample. The purity π is given by

$$\pi = N_{\text{tag}}^s / (N_{\text{tag}}^s + N_{\text{tag}}^{\text{bg}})$$

with $N_{\text{tag}}^{\text{bg}}$ being the number of tagged background hemispheres.

Due to the high mass ($\simeq 5$ GeV) and long lifetime ($\simeq 1.5$ ps) of b hadrons, hemispheres containing b-quarks can be tagged using an ANN with typical efficiencies of 25% and purities of about 92% [11].

However, for the lighter charm quark, the measured fragmentation properties and secondary vertex quantities are very similar in charm events and uds events (Fig. 2). High purity charm tags with low efficiency are possible using $D^{*\pm}$ mesons or leptons with high transverse momentum from semi-leptonic decays. Applying an ANN or SVM charm tag to kinematic variables defined for all charm events is expected to increase the charm tagging efficiency at the cost of lower purities.

The second problem is the identification of muons which are produced in the fragmentation of $e^+e^- \rightarrow q\bar{q}$ events. Muons are usually not absorbed in the calorimeters. They are measured in muon chambers which build the outer layer of a typical collider detector. A signal in these chambers which is matched to a track in the central tracking chamber is already a good muon discriminator.

The OPAL detector at LEP has been extensively described elsewhere [12,13].

The event generator JETSET 7.4 [14] is used to simulate e^+e^- annihilation events ($e^+e^- \rightarrow q\bar{q}$), including the fragmentation of quarks into hadrons measured in the detector. The fragmentation model has to be tuned empirically to match kinematic distributions as measured with the detector [15]. The response of the detector is also simulated. A data set of simulated e^+e^- collisions at a centre-of-mass energy $\sqrt{s} = m_{Z^0}$ is used for the charm identification problem. A second Monte Carlo data set at $\sqrt{s} = 189$ GeV has been used for the muon identification problem.

3 Problem 1: Charm quark tagging

The Monte Carlo events had to fulfil preselection cuts which ensure that an event is well reconstructed in the detector. These cuts result in a first overall reduction of the reconstruction efficiency. The input variables for the machine-learning algorithms were chosen from a larger dataset of 27 variables, containing various jet shape variables, e.g. Fox-Wolfram moments and eigenvalues of the sphericity tensor, plus several secondary vertex variables and lepton information. A jet finding algorithm [16] clusters the tracks and the calorimeter clusters into jets. Only the highest energy jet per hemisphere is used. The variables containing information about high transverse momentum leptons were removed in order to avoid a high correlation of this charm tag with the charm tag using leptons. The 14 variables with the largest influence on the performance of an ANN (27-27-1 architecture), trained to classify charm versus not-charm, were picked from the larger set. The variable selection method used is equivalent to a method which selects the variables with the largest connecting weights between the input and the first layer. This method has been shown to perform a fairly good variable selection [17]. It would be interesting to try Hessian based selection methods in comparison [6]. A ANN classifier with a 14-14-1 architecture was found to have the best performance. More complex architectures did not improve the classification.

At generator level five different quark types are distinguished. Due to their very similar decay and jet properties, Monte Carlo events coming from u,d and s quarks are put into one single class (uds). The data set consists of 10^5 hemispheres per uds, c and b class. However, the efficiencies and purities are calculated assuming a mixture of quark flavours according to the Standard Model prediction. This set is divided into training, validation and test sets of equal size. The learning machines were trained on equal number of events from the three classes. The supervision during the learning phase consisted of a charm versus not-charm label (udsb), thus distinguishing only two classes.

The outputs of both learning machines are shown in Fig. 3. The two classes c and udsb are separated by requiring a certain value for the output. This defines

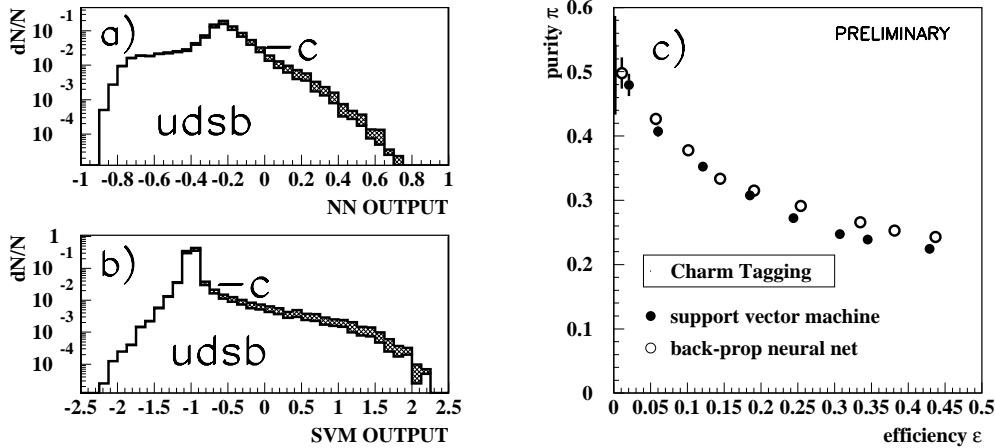


Fig. 3. The output distributions of a) the ANN and b) the SVM. c) Purity versus efficiency for both charm tags. Statistical errors are shown.

the efficiency and purity of the tagged sample. The purity π as a function of the efficiency ε for the two charm tags are shown in Figure 3 with the statistical errors. The performance of the SVM is comparable to the performance of the ANN with a slightly higher purity π for the ANN at larger efficiencies ε .

4 Problem 2: Muon identification

The muon candidates are preselected by requiring a minimum track momentum of 2 GeV and by choosing the best match in the event between muon chamber track segment and the extrapolated central detector track. Ten discriminating variables containing muon matching, hadron calorimeter and central detector information on the specific energy loss, dE/dx , of charged particles were chosen from a larger set of variables.

The Monte Carlo data set consists of $3 \cdot 10^4$ muons and $3 \cdot 10^4$ fake muon candidates. This set was divided into training, validation and test sets of equal size. After training, the tag is defined by requiring a certain value for the output of the learning machines. The resulting purity as a function of efficiency for muon identification is shown in Fig. 4. For high efficiency the performance of the SVM is very similar to the ANN.

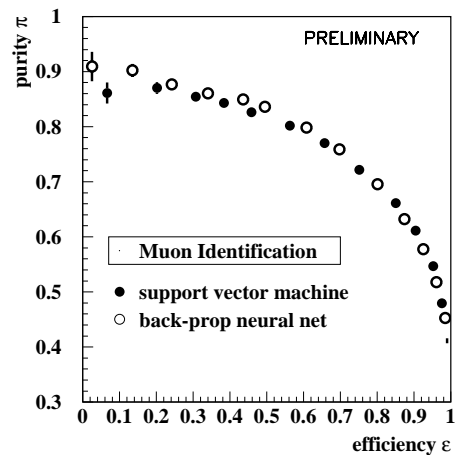


Fig. 4. The separation performance on identifying muons shown for a Neural Net trained with back-propagation and a Support Vector Machine. Statistical errors are shown

5 Conclusion

We have compared the performance of Support Vector Machines and Artificial Neural Networks in the classification of two distinctly different problems in high energy collider physics: charm-tagging and muon identification. The constructed SVM and ANN classifiers give consistent results for efficiencies and purities of tagged samples.

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft with the grants SO 404/1-1, JA 379/5-2, JA 379/7-1. We also would like to thank Frank Fiedler and Andreas Sittler of DESY for their help with the preparation of the OPAL data. We thank Gunnar Rättsch and Sebastian Mika of GMD-FIRST for helping to analyse the data.

References

- [1] For a review see: B. Denby, Proceedings of AIHENP 1993, ed. D. Perret-Gallix, World Scientific (1993).
- [2] C. Kiesling et al. (Principles and Hardware), S. Udluft et al. (Training and Control) and L. Janauschek et al. (Physics Results), Proceedings of AIHENP

- 1999, eds. G. Athanasiu, D. Perret-Gallix, Elsevier North-Holland (1999).
- [3] L. Lönnblad, C. Peterson and T. Rognvaldsson, *Phys. Rev. Letters* 65 (1990) 1321.
 - [4] P. Mazzanti and R. Odorico, *Z. Phys. C*59 (1993) 273.
 - [5] C. Peterson, T. Rognvaldsson and L. Lönnblad, *Comp. Phys. Comm.* 81 (1994) 185.
 - [6] G.B. Orr and K.-R. Müller, (eds.) (1998), *Neural Networks: Tricks of the Trade*, Springer Heidelberg (1998), LNCS 1524.
 - [7] V. Vapnik, *The nature of statistical learning theory*, Springer Verlag, New York (1995).
 - [8] C. Cortes and V. Vapnik, *Machine Learning* 20 (1995) 273.
 - [9] B. Schölkopf, C.J.C. Burges and A.J. Smola (eds.), *Advances in Kernel Methods – Support Vector Learning*, MIT Press, Cambridge, MA (1999).
 - [10] For more information on SVM, consult <http://svm.first.gmd.de>
 - [11] OPAL Collaboration, G. Abbiendi et al., *Eur. Phys. J. C*8 (1999) 217.
 - [12] OPAL Collaboration, K. Ackerstaff et al., *Z. Phys. C*74 (1997) 1.
 - [13] P.P. Allport et al., *Nucl. Instr. and Methods A*346 (1994) 476.
 - [14] T. Sjöstrand, *Comp. Phys. Comm.* 82 (1994) 74.
 - [15] OPAL Collaboration, G. Alexander et al., *Z. Phys. C*69 (1996) 543.
 - [16] OPAL Collaboration, R. Akers et al., *Z. Phys. C*63 (1994) 197.
 - [17] J. Proriot, *Proceedings of AIHENP 1995*, eds. B. Denby & D. Perret-Gallix, World Scientific (1995).