Łukasz Janyst*, Anna Kaczmarska**, Tadeusz Szymocha**, Marcin Wolter**, Andrzej Zemła**

OPTIMIZATION OF TAU IDENTIFICATION IN ATLAS EXPERIMENT USING MULTIVARIATE TOOLS

Elementary particle physics experiments, which search for very rare processes, require the efficient analysis and selection algorithms able to separate a signal from the overwhelming background. Four learning machine algorithms have been applied to identify τ leptons in the ATLAS experiment: projective likelihood estimator (LL), Probability Density Estimator with Range Searches (PDE-RS), Neural Network, and the Support Vector Machine (SVM).

All four methods have similar performance, which is significantly better than the baseline cut analysis. This indicates that the achieved background rejection is close to the maximal achievable performance.

Keywords: *multivariate methods, High Energy Physics, ATLAS, tau leptons*

OPTYMALIZACJA IDENTYFIKACJI LEPTONÓW TAU W EKSPERYMENCIE ATLAS Z UŻYCIEM METOD ANALIZY WIELU ZMIENNYCH

W eksperymentach fizyki wysokich energii, poszukujących bardzo rzadkich procesów, dużego znaczenia nabierają algorytmy umożliwiające separację sygnału od przeważającego tła. Cztery algorytmy uczące się na przykładach zostały zastosowane do identyfikacji leptonów tau w eksperymencie ATLAS: rzutowane rozkłady prawdopodobieństw (projective likelihood estimator – LL), PDE-RS (Probability Density Estimator with Range Searches), sieć neuronowa oraz maszyna wektorów wspierających (SVM).

Algorytmy te mają zbliżone wydajności znacząco lepsze od standardowej analizy z użyciem cięć. Sugeruje to, że osiągnięte wydajności są bliskie maksymalnej osiągalnej granicy.

Słowa kluczowe: *analiza wielu zmiennych, fizyka cząstek elementarnych, ATLAS, leptony tau*

* Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland, `Lukasz.Janyst@cern.ch`

** Institute of Nuclear Physics, Polish Academy of Sciences, Krakow, Poland, `Anna.Kaczmarska@ifj.edu.pl`, `Tadeusz.Szymocha@ifj.edu.pl`, `Marcin.Wolter@ifj.edu.pl`, `Andrzej.Zemla@ifj.edu.pl`

1. Introduction

Tau leptons play an important role in the physics to be observed at the LHC (Large Hadron Collider at CERN, Geneva). Their efficient detection is crucial for electroweak measurements, studies of the top quark and as a signature in search for new phenomena such as Higgs bosons, Supersymmetry (SUSY) and Extra Dimensions. However, the tau reconstruction and identification is not an easy task. The QCD multi jet events dominating the backgrounds have much larger cross section, therefore an efficient selection using multivariate analysis techniques is needed [1, 2].

In this contribution, we describe multivariate methods used for τ-jet identification in the ATLAS experiment: projective likelihood estimator (LL), Neural Network (NN), Probability Density Estimator with Range Searches (PDE-RS) and Support Vector Machine (SVM). The analysis is performed on the simulated ATLAS data, the channels $Z \to \tau\tau$, $W \to \tau\nu$ (with hadronic τ decays) are used as signal events and the events with QCD jets as background.

The ACK Cyfronet AGH "Zeus Cluster" was used to run the reconstruction jobs in order to prepare training samples and later on to validate the identification methods. It also participated, as part of the GRID, in the data simulation task.

2. Physics processes with τ leptons at ATLAS detector

The ATLAS experiment (A Toroidal LHC Apparatus) measures 22 m high, 44 m long and weights 7000 tons. The ATLAS detector is composed of a tracker, a calorimeter system (electromagnetic and hadronic) and of a large muon spectrometer. More details about the detector can be found elsewhere [3].

Detection of many processes depends on the efficient reconstruction of hadronic τ decays: light Standard Model (SM) Higgs production in Vector Boson Fusion (VBF) $qqH \to qqH$, charged SUSY Higgs production $H \to \tau\nu$, neutral SUSY Higgs $H/A \to \tau\tau$ at large $\tan\beta$, SUSY signatures with τ in the final state as well as Extra Dimensions. The well known SM processes $Z \to \tau\tau$, $W \to \tau\nu$ will be also used to calibrate the calorimeters. Tau leptons decay to hadrons in 64.8% of the cases and to electron or muon the rest of the time. In about 77% of hadronic τ decays only one charged track is produced: $\tau \to \nu_\tau + \pi^\pm + n\pi^0$ and in about 23% there are 3 charged tracks: $\tau \to \nu_\tau + 3\pi^\pm + n\pi^0$. The τ candidates with a single charged track are called 1-prong, with three tracks 3-prong. $3\pi^\pm$ candidates with one track missing or candidates with a single $\pi^\pm$ and one fake track are referred as 2-prong.

A τ lepton decaying hadronically generates a narrow τ jet. The background misidentified as τ candidates is mainly a QCD multi jet event, but also electrons that shower late or with strong Bremsstrahlung and muons interacting in the calorimeter are contributing. A τ-jet can be identified through the presence of a well collimated calorimeter cluster with a small number of associated charged tracks. In ATLAS two algorithms of τ reconstruction and identification are used: TauRec [4], which is based on calorimeter clusters and Tau1P3P [4, 5] starting from a good quality leading hadronic track and creating a τ-jet candidate based on tracks and also on an additional calorimeter information. All of the multivariate identification methods presented here refer to the Tau1P3P algorithm.

For Tau1P3P algorithm several discriminating variables to separate real τ jets from the background are defined [6]:

1. variables using the tracking information:
   - **numTrack:** number of associated tracks in a narrow cone (number of prongs), defines which version of an algorithm to use,
   - **rWidth2Trk3P:** weighted width of track with respect to axis τ (for candidates with more than one track),
   - **massTrk3P:** invariant mass of tracks (for candidates with more than one track),
2. variables using calorimetry information:
   - **numStripCells:** number of calorimeter strips with energy deposits over a given threshold,
   - **stripWidth2:** energy weighted width in strips,
   - **isolationFraction:** ratio of the uncalibrated transverse energy within $0.1 < \Delta R < 0.2$,
   - **emRadius:** energy weighted radius of the jet in the electromagnatic part of the calorimeter,
3. variables using both the calorimetry and the tracking information:
   - **m:** invariant mass calculated using the energy-flow technique,
   - **etIsolFrac:** ratio of transverse energy in $0.2 < \Delta R < 0.4$ to total transverse energy at electromagnetic scale,
   - **nAssocTracksIsol:** associated tracks in isolation region,
   - **etChrgHADoverPttot:** ratio of energy in the hadronic part of the calorimeter to the total energy of tracks,
   - **et:** visible transverse energy.

The variables are not independent and no single variable provides a really good signal and background separation (see Fig. 1 for three prong data), which emphasizes a need for efficient selection algorithms. Beside the standard cut analysis and projective likelihood estimator, three multivariate algorithms are applied to select τ candidates: Probability Density Estimator with Range Searches (PDE-RS), Neural Network (NN) and Support Vector Machine (SVM). For testing these techniques the data are split into two parts: one is used for training and the other one for validation.

3. Projective likelihood estimator

The method of maximum likelihood consists of building a model out of probability density functions (PDFs) that reproduces the input variables for signal and background.

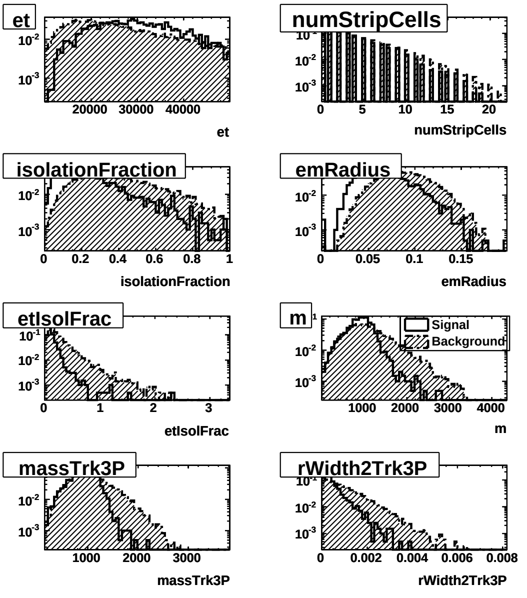

Fig. 1. Discriminating variables for 3-prong τ candidates in the transverse energy range 10 GeV–50 GeV (solid line – signal, dashed with hashed filling – background)

The likelihood ratio $y_L(i)$ for event i is defined by:

$$y_L(i) = \frac{L_S(i)}{L_S(i) + L_B(i)} \tag{1}$$

$$L_{S(B)}(i) = \prod_k p_{S(B),k} \tag{2}$$

where $p_{S(B),k}$ is the signal (background) PDF for the k-th input variable. In this approach correlations among the variables are ignored.

Since the parametric form of the PDFs is generally unknown, the PDF shapes are empirically approximated from the training data by nonparametric functions like polynomial splines of various degrees or unbinned kernel density estimators. In our application the polynomial splines are fitted to binned histograms.

4. PDE-RS method

The implementation of the PDE-RS [7] used for the τ identification is based on the publication [8]. As well as most of the standard multivariate algorithms, the technique combines the input observables into a single one, called a discriminant, on which a cut separating signal from background is applied. The calculation of the discriminant is based on sampling the signal and background densities in a multidimensional phase space built out of discriminating variables. Taking the number of signal events n_S and number of background events n_B in a small volume $V(x)$ around point x in the multidimensional space, a discriminant defined as:

$$D(x) = \frac{n_S}{n_S + cn_B} \tag{3}$$

is a good approximation of probability that given candidate is a signal. Parameter $c = N_S/N_B$ is the ratio of the total number of signal events N_S to the number of generated background events N_B. The event counting is done using multidimensional binary trees. As stated in [8], this method is supposed to give significantly better results than the cut analysis and comparable to other multivariate techniques.

5. Neural Network

Neural network is a non-linear discriminating method (we refer reader to [9] for detailed description of the neural network techniques). The Stuttgart Neural Network Simulator [10] is used for the identification of τ candidates. In the feedforward network, as used for the τ identification, the information propagates from input to output without any loops. To each neuron j in the hidden layer n inputs x_k and one output variable (the answer of the neuron) z_j are associated. For the first hidden layer the inputs are the discriminating variables, for next layers the inputs are the outputs of the preceding layer.

In the process of training, patterns are presented to the network which generates an output. The output is compared with the desired output from the training sample and the cost function is calculated. Then the weights in nodes are adjusted to decrease the value of the cost function. The errors are propagated backward using the current weights (the backpropagation algorithm [11, 12]).

The architecture of the network is optimized to give the proper classification of signal and background and to avoid over-fitting at the same time. The neural network used for the τ identification is built with 9 (1-prong candidates) or 11 (2 or 3-prong) input nodes and two layers of hidden nodes, each with 14 nodes.

6. Support Vector Machine

In the early 1960s, the linear support vector method was developed to construct separating hyperplanes for pattern recognition problems [13, 14]. The position of the hyperplane is defined by the subset of all training vectors called support vectors. The extension into non-linear SVM [15, 16] is performed by mapping input vectors into a high dimensional feature space in which data can be separated by a linear procedure using the optimal separating hyperplane.

A detailed description of SVM formalism can be found for example in [17], here only a brief introduction is given. Consider a simple two-class classifier with oriented hyperplanes. If the training data is linearly separable, then such a set of $(\vec{w}, b)$ pairs can be found which provides that the following constraints are satisfied:

$$\forall_i y_i(\vec{x_i} \cdot \vec{w} + b) - 1 \geq 0 \tag{4}$$

where x_i are the input vectors, y_i the desired outputs $y_i = \pm 1$ and $(\vec{w}, b)$ define a hyperplane. Intuitively, the classifier with the largest margin will give a better generalization. Hence, in order to maximize the margin, one needs to minimize the cost function W:

$$W = 1/2|w^2| \tag{5}$$

with the constraints from Eqn. (4). The training data points laying on the margins, which are called the support vectors (SV), are the data that contribute to defining the decision boundary (see Fig. 2). If the other data are removed and the classifier is retrained on the remaining data, the training will result in the same decision boundary.

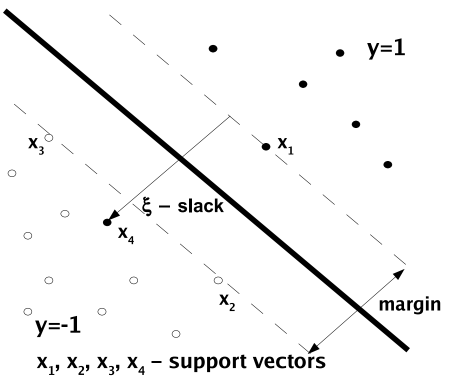

Fig. 2. Hyperplane classifier in two dimensions. Points x_1, x_2 and x_3 define the margin, i.e. they are the support vectors

For non-separable data the correct classification constraints in Eqn. (4) are modified by adding a slack variable ξ_i to it ($\xi_i = 0$ if the vector is properly classified, otherwise ξ_i is a distance to the decision hyperplane). The training algorithm needs to minimize the cost function, i.e. a trade-off between maximum margin and classification error:

$$W = 1/2|\vec{w}|^2 + C\sum_i \xi_i \tag{6}$$

The selection of C parameter defines how much a misclassification increases the cost.

The formulation of SVM presented above can be further extended to build a nonlinear SVM, which can classify nonlinear data. Consider a function Φ, which maps the training data from $\Re^n$ to some higher dimensional space $\Re^N$. In this high dimensional space, the data can be linearly separable, hence the linear SVM formulation can be applied.

In the SVM formulation data appear only in the form of dot products $\vec{x_i} \cdot \vec{x_j}$ [15]. The dot product in the high dimensional feature space is replaced by a kernel function:

$$K(\vec{x_i}, \vec{x_j}) = \Phi(\vec{x_i}) \cdot \Phi(\vec{x_j}) \tag{7}$$

By using the kernel function, one avoids the explicit mapping $\Phi(\vec{x})$. This is desirable, because $\Phi(\vec{x})$ can be tricky or impossible to compute. The most frequently used kernel functions are the Gaussian, polynomial and linear.

The optimization problem becomes well defined convex quadratic programming problem, which assures us that there exists a global minimum. This is an advantage of SVMs compared to neural networks, which may fall into one of the local minima.

We have implemented the SVM algorithm in the CERN ROOT framework [18], so that it became a part of the TMVA package [19] and became available for the whole High Energy Physics community.

7. Application to the identification of τ particles in the ATLAS experiment

The projective likelihood estimator, PDE-RS, Neural Network and the SVM have been used for the identification of τ leptons in the ATLAS experiment. The distributions of discriminants are shown in Figure 3 and the results in Figure 4. The signal efficiency is defined as a ratio of accepted and all signal events $\epsilon_S = (n^{sig}_{accepted})/(n^{sig}_{all})$ and background rejection as a ratio of all background events to the number of accepted background events $R = (n^{bkg}_{all})/(n^{bkg}_{accepted})$.

All multivariate algorithms perform significantly better than the basic cut analysis (background rejection is more than two times higher depending on the algorithm applied). Also the methods taking into account correlations between variables give a better background rejection than the projective likelihood ratio. All these methods have very similar performance, which might indicate that the achieved background rejection is close to the Bayesian limit.

8. Data processing at ACK Cyfronet AGH

For the production of the training samples for the multivariate analysis about one milion of simulated events were used. The simulation of events was performed centrally using the World Wide GRID environment, to which the ACK Cyfronet AGH "Zeus Cluster" also belongs. The reconstruction of all samples used for analysis was performed exclusively at ACK Cyfronet AGH using the "Zeus Cluster". The CPU needed to simulate and to reconstruct a single event is given in Table 1.

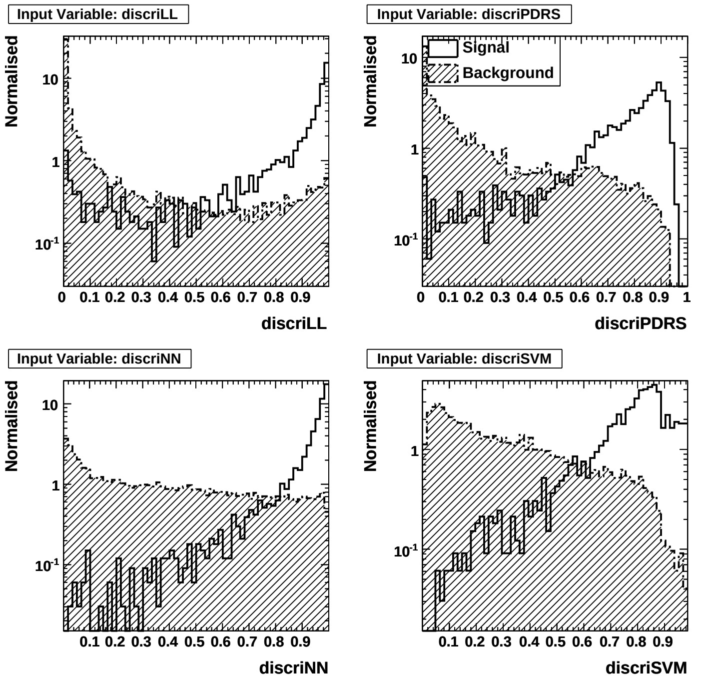

Fig. 3. Discriminant distributions for a signal and for the background in the visible transverse energy interval 10 GeV–50 GeV. Distributions are shown for the projective likelihood estimator (discriLL), PDE-RS method (discriPDRS), Neural Network (discriNN) and SVM (discriSVM). The solid line corresponds to the signal, the dashed line with a hashed filling to background

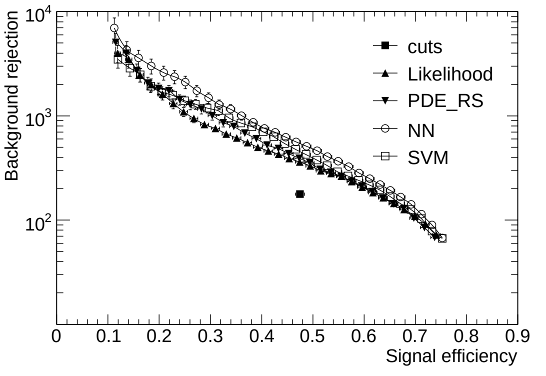

Fig. 4. The rejection of background as a function of signal efficiency for 3-prong τ candidates for the visible transverse energy interval 10 GeV–50 GeV. Results are shown for the simple cuts (single point), projective likelihood estimator, PDE-RS method, Neural Network and SVM

Table 1
The assumed event data sizes and the corresponding processing times. Processing times should be reduced to meet the ATLAS requirements (marked as "target") [20]

Item	Size of output event	Processing time
Simulation	2 MB	400 (100 target) $kSI2k - sec/event$
Reconstruction	0.5 MB	30 (15 target) $kSI2k - sec/event$

The CPU time needed for reconstruction is about 30 $kSI2k - sec/event$, which means, that to process one million events about $3 \cdot 10^7$ $kSI2k - sec$ (about 1 year) of Pentium IV 2.8 GHz is needed. The total disk space needed to store one milion of events is 2 TB.

Also for the validation of the multivariate method at least a partial reprocessing (about 20% of data) is needed. This procedure was repeated many times using exclusively the "Zeus Cluster".

9. Summary

Identification of τ candidates by the Tau1p3p algorithm is significantly improved by using multivariate analysis tools. All of the applied classification methods are performing well giving similar results. The analysis based on cuts is robust, transparent for users and doesn't require CPU consuming training. Also the projective likelihood estimator, which ignores correlations between variables, is a stable and fast classification method. The neural network is giving a very good performance. After a costly training it allows a very fast classification while the trained network is converted to a C code. PDE-RS is robust and transparent for users, but large samples of reference candidates are needed, also the classification is slower than for other methods. The SVM algorithm wasn't, up to now, commonly used in HEP. We have shown that Support Vector Machine can be successfully used to analyze High Energy physics data. The implementation described above is included in the ROOT package, therefore it is easily available to the entire particle physics community. This implementation extends the range of multivariate analysis tools available within the ROOT framework.

Acknowledgments

The authors would like to thank the entire ATLAS Tau WG for their help and support. Special thanks go to D. Cavalli and E. Richter-Wąs for their support and useful suggestions. We own thanks also to A. Höcker for introducing us to the TMVA package and for his help while dealing with TMVA problems.

The work was supported in part by Polish Ministry of Science and Higher Education grants 154/6.PRUE/2007 and PBS NR 132/CER/2006/03.

References

[1] Wolter M.: *Multivariate analysis methods in physics*, Physics of Particles and Nuclei, Volume 38, Issue 2, pp. 255–268, 03/2007

[2] Wolter M., Zemla A.: *Optimization of tau identification in ATLAS experiment using multivariate tools*, ACAT 07,April 23–27, 2007 Nikhef, Amsterdam, `http://pos.sissa.it/`

[3] ATLAS collaboration: *Detector and Physics Performance Technical Design Report, Volumes 1 and 2*, CERN/LHCC/99-14, ATLAS TDR 14, (1999)

[4] Bechtle P. *et al.*: *Identification of hadronic tau decays with ATLAS detector*, ATL-PHYS-INT-2008-003, ATL-COM-PHYS-2007-066

[5] Kaczmarska A., Richter-Was E., Wolter M., Janyst L.: *Performance of the tau1p3p algorithm for hadronic tau decays identification with release 12.0.6*, ATLAS Note ATL-PHYS-INT-2008-004; ATL-COM-PHYS-2007-039

[6] `https://twiki.cern.ch/twiki/bin/view/Atlas/Tau1P3P`

[7] Janyst L., Richter-Was E.: *Hadronic tau identification with track based approach: optimisation with multi-variate method*, ATL-COM-PHYS-2005-028; Geneva: CERN, 03 Jun 2005

[8] Carli T., Koblitz B.: *A multi-variate discrimination technique based on range-searching*, Nucl. Instrum. Meth. A, (501), p. 576, 2003

[9] Bishop C. M.: *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, 1995

[10] Zell A. *et al*: `http://www-ra.informatik.uni-tuebingen.de/SNNS/`

[11] Werbos P. J.: *Beyond regression: new tools for prediction and analysis in the behavioural sciences*, Ph.D. thesis, Harvard University, Boston, MA, 1974

[12] Rumelhart D. E., Hinton G. E., Williams R. J.: *Learning internal representations by error propagation. Vol. 1* of Computational models of cognition and perception, Cambridge, MA, MIT Press, chap. 8, pp. 319–362, 1986

[13] Vapnik V., Chervonenkis A.: *A note on one class of perceptrons.* Automation and Remote Control, (25), 1964

[14] Vapnik V., Lerner A.: *Pattern recognition using generalized portrait method.* Automation and Remote Control, (24), 1963

[15] Boser B., Guyon I., Vapnik V.: *A training algorithm for optimal margin classifiers.* Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, ACM Press, pp. 144–152, 1992

[16] Cortes C., Vapnik V.: *Support vector networks*, Machine Learning, (20), pp. 273–297, 1995

[17] Burges C.: *A tutorial on support vector machines for pattern recognition.* Data Mining and Knowledge Discovery, 2(2), pp. 1–47, 1998

[18] Brun R., Rademakers F.: *ROOT – An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81–86. See also `http://root.cern.ch`

[19] Höcker A., Voss H., Voss K., Stelzer J.: *TMVA (Toolkit for MultiVariate Analysis)*, `http://tmva.sourceforge.net`

[20] Atlas Collaboration: *Atlas Computing Technical Design Report*, ATLAS TDR–017, CERN-LHCC-2005-022, 18 March 2005