ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 216, 2008

*Marcin Wolter**

HOW TO RECONSTRUCT THE UNKNOWN PHYSICAL QUANTITIES USING NEURAL NETWORKS?[1]

ABSTRACT. In this article an application of neural networks to the reconstruction of unknown physical quantities in particle physics is presented. As an example the mass reconstruction of the hypothetical Higgs boson in the typical high energy physics experiment is used. Monte Carlo events are used to determine the probability distributions of observables (energies of two jets and the angle between them) for various Higgs boson mass, which are later fitted using a Neural Network. These distributions are used to determine the mass probability distribution of the measured particle. The mass is reconstructed without knowing the functional dependence between the observables and the measured quantity. The miscalibration of the measured quantities is automatically corrected in this method.

Key words: reconstruction, physics, Bayesian, neural network.

I. INTRODUCTION

In elementary particle physics we need frequently to reconstruct physical quantities while the functional dependence between them and the measured observables is not well known. The precision of particle property reconstruction can be improved by using data analysis methods that better explore all the accessible information. The Bayesian approach, based on an interpretation of probability as a conditional measure of uncertainty, provides such an opportunity.

In the study presented we consider the reaction $H \to bb \to 2\ jets$, in which the Higgs boson decays into two *b*-mesons producing two jets, i.e. bunches of particles. The obvious and simplest estimate of the Higgs boson mass is the invariant mass of the two jets. To calculate the mass estimate the energies of two jets (E_1, E_2) and the angle between them (θ_{12}) are needed:

* Ph. D., Institute of Nuclear Physics Polish Academy of Sciences, Kraków.

[1] Supported in part by Polish Ministry of Science And Higher Education grants 154/6.PRUE/2007 and PBS NR 132/CER/2006/03.

$$m^2_{Higgs} = (E_1 + E_2)^2 + (\vec{p}_1 + \vec{p}_2)^2 \approx 2E_1E_2(1-\cos(\theta_{12})). \tag{1}$$

The same three variables are used in our analysis, but we avoid using the invariant mass formula. Also the energies of reconstructed jets E_1 and E_2 are not calibrated, consequently the invariant mass of two jets is shifted towards lower values.

In the method presented here for every event the probability $P(x|m)$ of x (vector of measured variables) belonging to class m (mass) is computed. The event x is assigned to the class with the highest probability. For a sample of many events originating from the same, but unknown class, the probability is a product of the single event probabilities:

$$P(m \mid x_1,\ldots,x_n) \propto \left(\prod_{i=1}^{n} P(x_i, m)\right) \times P(m), \tag{2}$$

where $P(m)$ is a prior on the Higgs mass which we shall take to be constant. In the limit of a continuous probability function of the particle mass this method gives the best possible estimate of the mass, provided that the probability functions are well measured and the vector x describing the event contains all the necessary information. One should note that no explicit knowledge of the functional dependence of the mass estimate is needed.

In this study, neither physical nor combinatorial background is present. Only perfectly identified b-jets are used. We rely heavily on the Monte Carlo simulation, as in the most analysis using machine learning procedures. Events simulated with the given Higgs mass give us the relation between the true value of the unknown quantity and the observables.

Monte Carlo events were generated using PGS (Pretty Good Simulator) (Conway 2005, which was developed during the Fermilab Run II SUSY/Higgs Workshop, provides a fast simulation (thought to be accurate to about 15%) of the response of typical collider experiments to high energy collision events. Thirteen samples with Higgs masses between 95GeV and 155GeV with 5GeV increments were generated. All the simulated samples are divided into two subsamples: one is used for training the neural network, the second for further analysis.

II. METHOD OF RECONSTRUCTING THE UNKNOWN HIGGS MASS

Following (Garrido, Juste 1998) each of the simulated samples was used for the training of feed-forward neural network (Zell 1994, Zell 1995) against a sample with a flat distribution in all three observables. The probability $P(x|m)$ is obtained from the output NN_{out} of the adequately trained neural network:

$$P(x \mid m) \propto \frac{NN_{out}}{1 - NN_{out}}. \tag{3}$$

The network used for the fit consists of three input nodes, 50 nodes in the hidden layer and one output node. There are 13 independent networks, each of them trained using a sample with different Higgs mass.

Fig. 1 shows the distributions of the three observables together with the projections obtained from the fitted 3-dimensional function. One of the advantages of a neural network fit is that data are not binned, which improves the quality of the fit while fitting small data samples. Also no analytical formula of the fitted function is needed. The complexity of the function shape is determined by the number of neurons in the neural network. The trained neural network can be easily converted into a C-language function, which later returns the probability at a very low CPU cost.

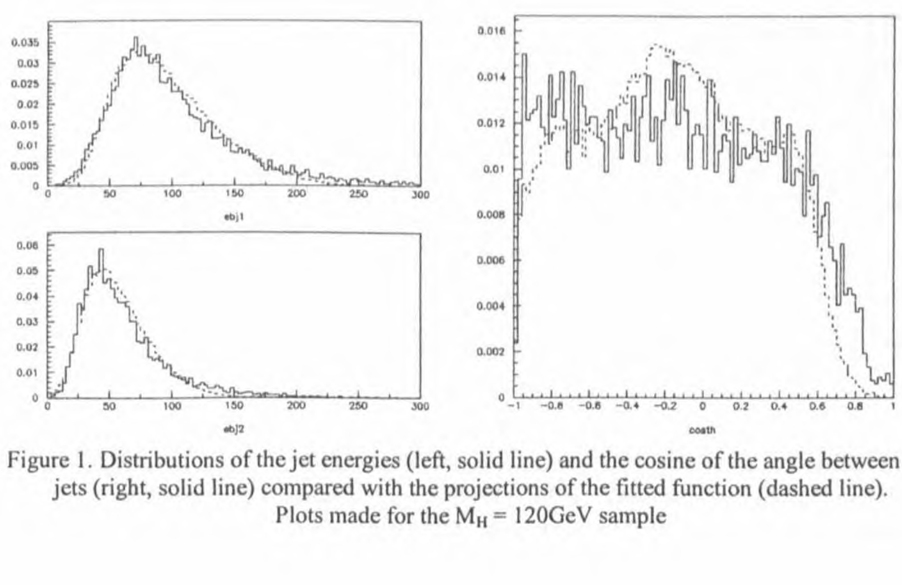

Figure 1. Distributions of the jet energies (left, solid line) and the cosine of the angle between the jets (right, solid line) compared with the projections of the fitted function (dashed line). Plots made for the M_H = 120GeV sample

For each event the trained neural networks return the set of probabilities corresponding to each Higgs mass. To obtain the mass estimate based on a sample of few events the probabilities are multiplied. If the true Higgs mass is in between two masses, for which the probability distributions were found, the probabilities for these two masses are approximately equal. Therefore the better mass estimation is obtained while using the mean instead of the mode. In practical application the size of averaging window should be optimized, here the average was taken over the entire mass spectrum.

III. RESULTS

The probability distributions as functions of the Higgs mass are shown in Fig. 2 for samples of 2300 events generated with four different Higgs masses. On the Y-axis the *ln(P(x|m))* is plotted, therefore heights of the bins differ by orders of magnitude. The plot shows, that the maxima are located at the true Higgs masses.

It is useful and necessary to test this method on a sample generated with a Higgs mass different from the masses used for training. Fig. 2 shows the distribution of masses reconstructed as means of the probability distributions for many smaller samples of events (129 events in a sample) and for two different generated Higgs masses of 120GeV and 137.5GeV. For the Higgs mass of 137.5GeV (not used in training procedure) two distinct maxima corresponding to two neighbor masses of 135GeV and 140GeV are observed.

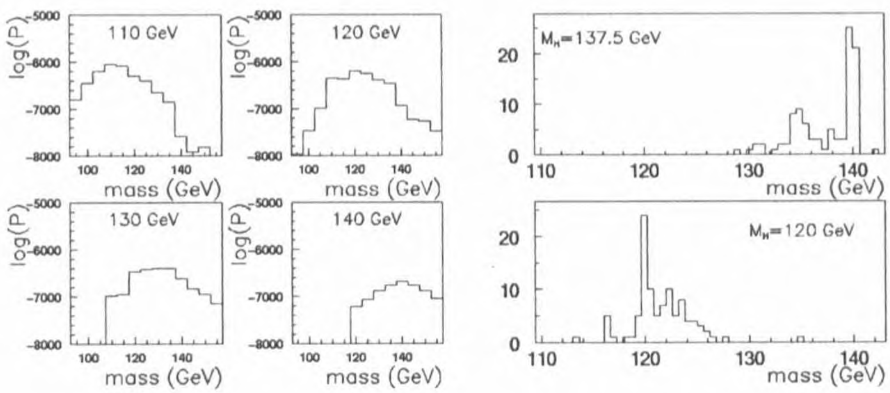

Figure 2. Probability distributions for the samples of 2000 events generated at four Higgs masses: 110 GeV , 120 GeV , 130 GeV and 140 GeV (left plot) . Reconstructed mass distribution for a sample size of 129 events and for M_H = 137.5 GeV (upper right plot) and for M_H = 120GeV (lower right plot)

Fig. 3 shows the reconstructed mass as a function of the number of events used for reconstruction (MC sample with Higgs mass 137GeV). On the Y-axis the mean mass over many small subsamples is plotted. The bias at low sample size is an "edge effect" due to the weighted average being taken over a finite Higgs mass range. The effect can be reduced by using a wider Higgs mass range.

The estimation of the mass resolution is obtained by analyzing the root mean square (*RMS*) of the reconstructed mass distribution as a function of the number of events in the sample. It should scale according to the formula $RMS(n) = RMS(1)/\sqrt{n}$, where n is a number of events in a sample. Fig. 3 shows this dependence for samples generated with Higgs masses 137.5GeV and 120GeV. The violation of the $1/\sqrt{n}$ dependence at small sample size is due to the fact, that the resolution is limited by the Higgs mass range (i.e. 95GeV to 155GeV). For greater numbers of events in a sample some violation of this dependence for $M_H = 137.5GeV$ sample is observed. Since the closest generated Higgs masses are 135GeV and 140GeV, the reconstructed mass tends to be equal to one of them, and therefore the RMS is approximately half of the difference between them, i.e. 2.5GeV (see Fig. 2). This effect can be reduced by generating MC samples with intermediate Higgs masses. Fig. 4 shows the dependence of the reconstructed mass (weighted mean for a sample of 57 events) and single event RMS as a function of the true Higgs mass. The dependence is fairly linear and the mass is properly reconstructed.

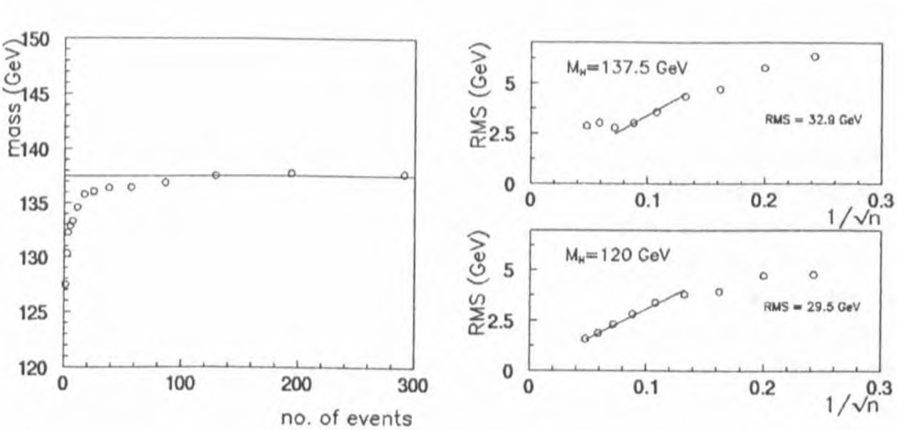

Figure 3: Mean reconstructed mass as a function of the sample size for a sample generated with Higgs mass of 137GeV (left plot). RMS as a function of $1/\sqrt{n}$, where n is a number of events in a sample. Upper right plot shows the dependence for a sample with $M_H = 137.5$GeV, the lower right for $M_H = 120$GeV.

The results are compared to the mean and RMS of the invariant mass distribution, which gives the masses significantly lower than the true Higgs boson mass. The invariant mass is scaled by a factor of 1.25 and compared to the results of the Bayesian method.

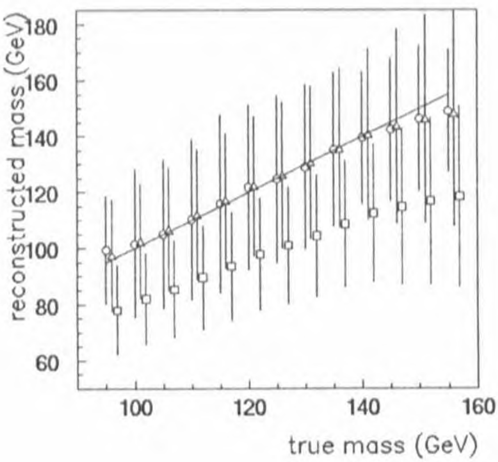

Figure 4. Reconstructed Higgs mass vs. the true Higgs mass for the Bayesian method (circles), mean of the invariant mass distribution (squares) and corrected mean of the invariant mass distribution (triangles). The error bars represent RMS of the distributions. The data points are shifted for better visualization.

IV. CONCLUSIONS

The Bayesian approach to the problem of the reconstruction of the particle mass or other particle properties can be performed without any knowledge of the functional dependence of the particle property on measured quantities. For this method, as for other methods based on learning algorithms, a good Monte Carlo simulation of the physical processes and the detector is essential. In the example presented here, where the Higgs boson mass is measured, the method gives a mass resolution similar to the one obtained using the standard invariant mass analysis, which leads to the conclusion, that in our simplified example no more information can be extracted from the jet energies and angle between them beyond that encoded in the invariant mass.

It was also shown, that the neural network is an excellent tool not only for signal and background discrimination, but also to perform multidimensional unbinned fits.

REFERENCES

Conway (2005), PGS (Pretty Good Simulator) is a fast approximate detector simulation program, developed by John Conway et al (http://www.physics.ucdavis.edu/~conway/)

Llus Garrido and Aurelio Juste (1998), On the determination of probability density functions by using Neural Networks , *Computer Physics Communications* 115, p. 25

Zell, A. (1994), Simulation Neuronaler Netze. Addison-Wesley (Germany).

Zell, A. (1995), SNNS user manual, version 4.1. *Technical report, University of Stuttgart, IPVR*. Report Nr. 6/95.

Marcin Wolter

REKONSTRUKCJA WIELKOŚCI FIZYCZNYCH Z UŻYCIEM SIECI NEURONOWYCH

W artykule zaprezentowane jest zastosowanie sieci neuronowych do rekonstrukcji nieznanych wielkości w fizyce cząstek elementarnych. Jako przykład użyta jest rekonstrukcja masy hipotetycznego bozonu Higgsa oparta na symulowanych danych. Dane te zostały użyte do wyznaczenia rozkładów prawdopodobieństwa mierzonych wielkości (energie dwóch dżetów oraz kąt pomiędzy nimi) dla różnych mas cząstki Higgsa. Rozkłady te zostały następnie sparametryzowane za pomocą sieci neuronowych oraz wyznaczenia rozkładu prawdopodobieństwa masy mierzonej cząstki. Masa jest wyznaczona bez użycia zależności funkcyjnej pomiędzy mierzonymi wielkościami a rekonstruowaną masą. Kalibracja wielkości pomiarowych jest automatycznie korygowana poprzez rozkłady prawdopodobieństwa.