

# Backtracing of the Impact Parameter through Pattern Recognition Analysis on Heavy Ion Reaction Data

Jacopo De Sanctis\*, Matteo Masotti† and A. Bonasera\*\*

\*Department of Physics and INFN, viale Irnerio, 40126 Bologna, Italy

†Medical Imaging Group and Department of Physics, viale B. Pichat, 40127 Bologna, Italy

\*\*INFN Laboratori Nazionali del Sud via Santa Sofia, 95123 Catania, Italy

**Abstract.** One of the problems in the analysis of nucleus-nucleus collisions is to get information on the value of the impact parameter  $b$  through the knowledge of some features of each event. The experimental values of  $Z$ , transverse energy and multiplicity of fragments are good candidates to get hints on the values of  $b$ . Motivated by this consideration, this work consists in the application of pattern recognition techniques aimed at associating values of  $b$  to groups of events. To this end, a support vector machine (SVM) classifier is adopted to analyze multifragmentation reactions. This analysis consists of mainly two different steps. In the first one, known as the training phase, SVM learns how to discriminate between peripheral, semi-peripheral and central reactions. These examples are taken from a large number of events generated with classical molecular dynamics (CMD) and heavy ion phase-space exploration (HIPSE) models for each value of  $b$ . In the second one, known as the test phase, what has been learned is tested on new events generated by the same models. Our tests demonstrate that, by following this approach, central, semi-peripheral and peripheral reactions are correctly classified for about 85% of the cases.

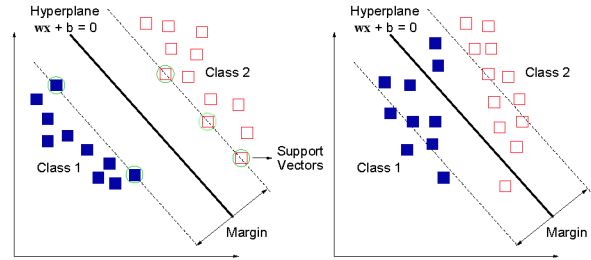
**Keywords:** Nuclear reactions  $Ni^{58}(Ca^{48}, X)$ , Nuclear fragmentation, Support Vector Machine

**PACS:** 24.10.-i; 25.70.-z; 25.70.Pq

## INTRODUCTION

Experimental heavy-ion collision data become more and more difficult to analyze without the help of theoretical calculations and new data analysis tools. For example, the knowledge for each event of the impact parameter value  $b$  is very important since it is strongly correlated to the *violence* of the reaction. Unfortunately, its experimental event-by-event extraction is still difficult to achieve. Up to now, the only way to achieve the value of  $b$ , is by analyzing the correlation between a single observable and the *violence* of the collision. Some observables can be charged particle multiplicity, transverse energy, etc. However, all these single observable methods, which are relatively efficient for peripheral reactions, fail for central collisions due to the saturation of the considered observables in the energy range between 20 and 100 AMeV.

The purpose of this paper is to approach this problem differently, namely, by applying pattern recognition techniques [1, 2]. Our goal is to discriminate objects that belong to different classes, by means of the knowledge of several observable characteristics of each event, i.e. features. Classes are represented by nuclear reactions arising from different values of the impact parameter  $b$ . A pattern is a pair of variables  $\{X, y\}$ , where  $X$  is a collection of observations or features and  $y$  is the label that represents the class to which the event belongs.



**FIGURE 1.** Left: the support vector machine training phase. Right: the support vector machine test phase. For sake of simplicity the case with only two classes is shown.

## SVM ON CMD AND HIPSE MODELS

Let us suppose that some samples  $\{X_i, y_i\}$ ,  $i = 1, \dots, d$  taken from some data distribution are given:

- $y_i \in \{1, 2\}$  are the labels representing the class membership of each sample,
- $X_i \in R^d$  are the features characterizing each sample.

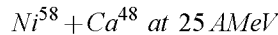
In order to discriminate between different classes of objects we use the support vector machine (SVM) classifier. This algorithm is based on two main phases (see Fig. 1). In the first one, known as the training phase, SVM finds the hyperplane  $W \cdot X + b = 0$  that maximizes the margin between the classes (see left of Fig. 1). In the second one, known as the test phase, once SVM has been trained,

each new sample  $X$  is classified according to:

$$f(X) = \text{sign} \left( \sum_{i=1}^d \alpha_i v_i K(X, X_i) + b \right), \quad (1)$$

where  $f(X)$  represents the class to which the objects belong, while  $K(X, X_i) = (\gamma X \cdot X_i + r)^d$  is known as the polynomial kernel of degree  $d$  (see right of Fig.1). This last one is very important because it determines the degree of hyperplane used in order to maximize the margin between different classes.

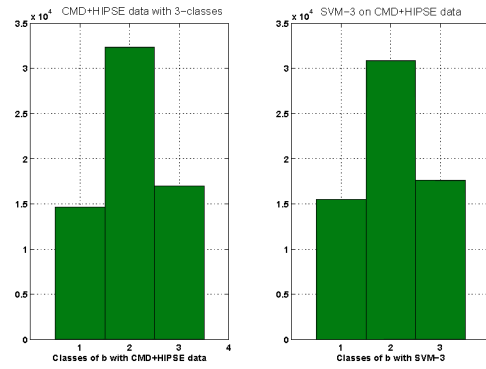
We apply the SVM of 3rd degree (SVM-3) on the classical molecular dynamics (CMD) model and on the heavy ion phase-space exploration (HIPSE) model [3, 4, 5] data for the following reaction:



In CMD and HIPSE model we have chosen the impact parameter  $b$  from 0.5 fm to 10 fm with a step of 0.5 fm, following a triangular distribution. For each event generated by CMD and HIPSE we have chosen some of the global observables, like the charged particle multiplicity ( $N_c$ ), transverse energy ( $E_t$ ), the biggest fragment ( $Z_{big}$ ), the number of intermediate mass fragments with  $Z \geq 3$  ( $N_{imf}$ ), the 0-th, 1st, and 2nd moment of the charge distribution and the total parallel momentum along the beam axis ( $q_z$ ), the total charge ( $Z_{tot}$ ) and the velocity along the beam axis of the biggest fragment ( $V_{Z_{big}}$ ). For CMD and HIPSE output data we used the same filter in order to simulate the experimental apparatus CHIMERA [6] and we selected for subsequent analysis the events with  $N_c \geq 2$ , as in the experimental case.

## RESULTS

The classification accuracy is defined as the number of objects correctly classified over the total number of objects. We have done many tests on our model's data in order to understand which was the best combination between the training and test data that could better reproduce the experimental data. In other terms, we used five different approaches: the first analysis was to apply the SVM-3 on CMD data, some of this data is used for the training phase while on the remaining part we checked the classifier accuracy (test phase). In this case we obtained, with a separation in 3-classes, a global accuracy of 89.91%. We called the first class central collisions (from 0.5 fm to 3.5 fm), the second class semi-peripheral collisions (from 4 fm to 6.5 fm) and the third class peripheral collisions (from 7 fm to 10 fm). In the second case we applied the classifier on HIPSE model data in the same way as CMD and we had a global accuracy of 80.97%, just a little bit less than CMD. It depends



**FIGURE 2.** Classification of  $b$  in 3 different classes. Learning on CMD plus HIPSE data and test on CMD plus HIPSE data with SVM-3. Global accuracy of 84.78%.

on the different structure of the interaction potential  $V$ . In this way we developed the learning phase of SVM-3 on CMD (HIPSE) data and then we checked it on HIPSE (CMD) data output obtaining a global accuracy of 71.91% (78.09%). The most important result is the case in which we took for the learning phase a dataset with a part of CMD data output plus a part of HIPSE data output, while on the remaining part of both we checked the SVM-3 accuracy obtaining a value of 84.78%, as is shown in Fig.2. This combination between the CMD and HIPSE data could be the most similar compared to the experimental data. In conclusion, pattern recognition analysis is a good and new approach to extract the impact parameter value from the CMD (HIPSE) model data by using many features at the same time. With respect to traditional methods, this technique also works well when used to discriminate central collisions. Furthermore, by means of the training phase that is an adaptable process, it will be possible to apply it on the experimental data [7] measured by the Nucleon-collaboration [6]. This will be object of study in the near future.

## REFERENCES

1. V.N.Vapnik, *The nature of statistical learning theory*, Springer Verlag, New York, 1995.
2. K.R.Müller, S. Mika et al., *IEEE Transactions on Neural Networks*, Vol.12, no.2, March, 2001.
3. Latora V., Belkacem M. and Bonasera A., *Phys. Rev. Lett.*, 73 (1994) 1765.
4. Latora V., Belkacem M. and Bonasera A., *Phys. Rev. C*, 52 (1995) 271.
5. D. Lacroix et al. *Physical Review C*, 69 054604 (2004).
6. www.bo.infn.it/nucl-ex
7. The analysis of experimental data is model dependent from CMD (HIPSE) model or from any model used in the training phase.