Using Machine Learning to Extract Properties of Systems of Particles

Ellen D. Giulian† ‡, Michael Kordell II§, Rainer J. Fries‡ ¶
†Department of Physics, University of Maryland - Baltimore County
‡Cyclotron Institute, Texas A&M University
§Department of Physics and Astronomy, Texas A&M University

In high energy and nuclear physics, one is often presented with complex final-state systems of particles that are products of many dynamical processes. With traditional analysis methods, it is hard to determine the specific processes that created these systems. Here, we explore the application of machine learning in analyzing such systems. We first discuss our Python simulations, which create systems of particles. Then, we apply various machine learning algorithms to pseudo-data generated by our models in order to predict features of the systems of particles.

Results

We analyze the daughter momenta in an event to extract parameters \( M, T \) and \( \alpha \). Using Python’s Scikit-learn library \([1]\), we trained several machine learning algorithms on the data generated by our model. Each algorithm was trained with 14,000 events and tested with 6,000 events. Each event consisted of 10 particles. We focused on four classifiers to obtain binary classifications:

1. random forest (RF): an ensemble classification algorithm with many individual decision trees; each tree in the forest yields a class prediction, and the class with the most number of votes is the winning prediction.
2. adaptive boost (ADA): an ensemble classification algorithm that fits weak learners to weighted data in order to create a strong learner; very sensitive to noise and outliers.
3. gradient boost (GB): an ensemble classification algorithm that builds a prediction model using weak estimators (typically decision trees); allows for extensive optimization.
4. multilayer perceptron (MLP): a neural network classifier that uses back-propagation (a tool that increases prediction accuracy) to train.

### Panel 1. Performance of ensemble classifiers in predicting targets in relation to the closeness in value between two binary target values.

- **RF**
- **ADA**
- **GB**

### Panel 2. Performance of ensemble classifiers in predicting targets as a function of the number of estimators.

We include results for the following machine learning algorithms:

- multilayer perceptron (MLP)
- random forest (RF)
- adaptive boost (ADA)
- gradient boost (GB)

While our work is a first step in training machine learning algorithms to understand systems of particles, further work is needed to develop more sophisticated simulation code. Ongoing work involves adjusting the model to account for particles that undergo \( N \)-body decays \((N \geq 2)\). In addition, further research is needed into optimizing the parameters to the classifiers. In the future, after exploring the ability of machine learning algorithms in our simplified models, we plan to use established simulation codes like JETSCAPE \([2]\) and PYTHIA \([3]\) which create more complex and realistic systems. If proven feasible, one possible application of our work is to increase our understanding of the hadronization process and the phenomenon of confinement by analyzing experimental data with machine learning.

### Discussion and Analysis

1. As expected, the performance of all classifiers improves with increasing contrast between parameter choices (Panel 1).
2. The algorithms can easily infer the masses of decayed particles. However, they have slightly more trouble predicting temperature and are least efficient in predicting the collective flow velocities.
3. Increasing the number of estimators, in general, results in higher accuracy. However, the impact of additional estimators appears to decrease at some point, and one can imagine that eventually there is a trade-off between accuracy and efficiency: adding more estimators to the classifier can increase runtime significantly. Other options, such as increasing the size of the training data set, can also improve accuracy.
4. The multi-layer perceptron classifier in Panel 3 is shown only for comparison. With default parameters, it appears to perform worse than the ensemble classifiers. While neural networks are very powerful, they are also very complex and require a significant amount of parameter tuning. Given the performance of the ensemble classifiers, we chose not to study the parameter optimization of the multi-layer perceptron classifier here.

### Remarks

While our work is a first step in training machine learning algorithms to understand systems of particles, further work is needed to develop more sophisticated simulation code. Ongoing work involves adjusting the model to account for particles that undergo \( N \)-body decays \((N \geq 2)\). In addition, further research is needed into optimizing the parameters to the classifiers. In the future, after exploring the ability of machine learning algorithms in our simplified models, we plan to use established simulation codes like JETSCAPE \([2]\) and PYTHIA \([3]\) which create more complex and realistic systems. If proven feasible, one possible application of our work is to increase our understanding of the hadronization process and the phenomenon of confinement by analyzing experimental data with machine learning.

This research was supported by NSF grants PHY-1550221, PHY-1812431, and PHY-1550221.

### References

