

Popular Abstract

One of the main goals of the experiments in the Large Hadron Collider is to find phenomena beyond what the Standard Model can explain. There are many theories predicting what phenomena beyond the Standard Model could be circulating in the theory community, and it is the job of the experimentalists to find the truth of these theories. And as you may have already guessed, it is not so easy to find something that can not be explained by the strongest theory in physics.

The challenge in searching for something that exciting is that it happens very rarely (if it doesn't, we would have already found it!). Even in the case that it happens, it is hidden in the abundance of the other not-so-exciting stuff. It is like trying to find a golden egg in a swamp of mud that is full of brown eggs, at night, with only a dim flashlight in your hand. Even if you find it, it would look like a brown egg and chances are, you would throw it back to the mud. So the challenge is, how does one develop a method to distinguish the golden egg and the brown egg? Or to say it in a more technical term: to discriminate the *signal* from the *background*.

In high energy particle collisions, particles are flying out of the collision point, created by transforming energy into mass, following Einstein's $E = mc^2$. Some of these particles cannot exist on their own, so new particles are created from the vacuum to bind with them. The detector would see this as a cone-shaped spray of particles coming from the collision point, called a *jet*. Some theories describing what lies beyond the Standard Model predict that there will be new undiscovered particles that would decay into jets with some particularity. This is why we use jets as our search object – our eggs.

To look for the jets we are interested in, physicists usually use quantities called the *jet substructure variables*. They are variables that describe some particular properties of a jet. Knowing the predicted value of these variables for the signal, we can use that value to make a cut around it to narrow our search. In our egg-finding analogy, if we predict the mass of the golden egg to be m , we can develop an algorithm like, "if the egg is lighter than $m - p$, or heavier than $m + p$, discard it!", where p is how tight you want your search algorithm to be. Make it too loose, you would accept too many brown eggs; make it too tight, you would lose some of the precious golden eggs. One can combine the algorithm with other variables, say, the shininess or the shape of the egg. A good search algorithm is the one that can reject as much background, while at the same time, keeping as much signal as possible.

Now, this is where the machine beats us: while humans can only process a limited number of variables, machine learning algorithm can learn from all the variables there are! Machine learning can extract the information contained in all of the variables and conclude whether the jet is more signal-like or background-like. In the studies in this thesis, machine learning techniques have been shown to perform roughly twice as good compared to a conventional method.

The implementation of machine learning in high energy physics will provide more powerful tools than the ones used in the past searches. These newly acquired tools will allow us to see hidden events that were previously undetectable and increase our chance for new discovery.