

Classification of Physics Events using Machine Learning

Capstone Proposal for Udacity's Machine Learning Engineer Nanodegree

Gabriel Santucci

May 2017

Proposal

1 Domain Background

There is a type of theories in particle physics called Grand Unified Theories (GUTs). The main idea behind all variations of GUTs is to have a unified description of 3 fundamental interactions of nature: the strong, weak and electromagnetic forces. There are many types of GUT models [1] each of them with different predictions, but one thing they all have in common is nucleon decay.

Nucleons is the generic name for protons and neutrons, particles that live inside the nucleus of an atom. As far as we know, the proton is a stable¹ particle and so is the neutron inside a nucleus (although free neutrons can decay). Since the typical energy of GUTs is far beyond the reach of any particle physics experiment, nucleon decay is the best bet for searching for evidence of GUTs.

One particular class of GUTs is the so called SUSY GUTs, when supersymmetry is present in the theory. In some models, the preferred channel for proton decay is $p \rightarrow \bar{\nu}_\mu K^+$, where the proton decays to two other particles: an anti-neutrino and a charged Kaon. Searching for this kind of event is very challenging, but some experiments were built in the 80's and 90's to search for this kind of event. One of the main difficulties to detect this type of signal is due to neutrinos that are created in Earth's atmosphere and then interact inside the detector, leaving a signature very similar to the one left by the decay of a proton.

In physics the interesting class that is being studied is called 'Signal' and all other possible interactions that produce similar data in the detector is called 'Background'. These can be mapped into $\{1,0\}$ classes for machine learning algorithms to perform binary classification². This study tries to improve the identification of these two classes: proton decay and atmospheric neutrino events.

¹Being stable means that the particle will not spontaneously decay into other particles.

²The physics and ML names will be used interchangeably.

2 Problem Statement

To date, the biggest experiment looking for proton decay signal is the Super-Kamiokande (SK) experiment [2]. SK is a big water tank containing 50 kton of ultra pure water. When a charged particle travels at extremely high speeds inside the tank, it produces light, this is the Cherenkov effect [3]. In the walls of the tank there are photo-multiplier tubes (PMTs), which essentially are light detectors [4]. Using these PMTs we know when a particle passed through the detector leaving a trace and producing light. In the SK tank there are more than 11 thousand PMTs to collect this light. We then have a data set with the time and charge of each PMT that received a hit. A reconstruction algorithm is then applied on this data so that physics quantities can be studied, like momentum, energy, direction and position of each particle that participated in the event.

As described in section , the proton decay mode that we are interested here consists of an anti-neutrino and a charged Kaon. The anti-neutrino has no electric charge, so it does not leave a trace in the detector and the Kaon also can not be seen, even though it has charge. That is because it does not have enough energy to be traveling at very high speeds, so it does not produce any light. But since the Kaon also decays, the hope is to see the decay products of the Kaon. Most of the time, it decays to a neutrino (which can not be seen) and an anti-muon³, which is charged and can be detected. Since this is a 2-body decay (the initial particle decays into 2 particles), and we know the masses of the final state particles, using energy and momentum conservation we also know the exact energy and momentum of the muon we are looking for.

The problem is that after all this, the only visible particle that we can detect is the mono-energetic muon⁴. This would make the search impossible, due to the amount of background events coming from atmospheric neutrino interactions. The solution is to look only for protons that decay inside the Oxygen nucleus of the water molecules, but not for the ones coming from the Hydrogen atoms. The difference is that when a proton decay inside the Oxygen, the remaining nucleus is left in an excited state and can emit a low energy photon (also called gamma) from nuclear de-excitation. If we look for this photon in coincidence with a mono-energetic muon, we then have a very particular signature in the detector that can be used to differentiate signal and background events.

Insert pdecay figure here!

Even though the situation is better now, it is still a challenge to identify if a signal comes from a proton that decayed or a neutrino that entered the detector and interacted with some nucleus. There are billions⁵ of neutrinos entering the detector every second.

³This is the only Kaon decay mode that will be used in the analysis here. But there is another decay mode that can be seen by SK.

⁴This means that the energy of the muon is always the same. Also, since the SK can not detect the sign of the particle's charge (+ or -), we will not make a distinction between muons and anti-muons.

⁵Neutrinos come from many different sources, the main ones are the atmosphere and the Sun. Most of them cross our bodies, the Earth and all else and leave intact, since the chance of a neutrino interacting with something is incredibly small. Only from the Sun, 65 billion neutrinos cross a person's thumb every

The chance of a neutrino interacting with a nucleus inside the tank is extremely small, but since the flux of incoming neutrinos is so high, we are bound to see some neutrino events⁶ every hour or so.

Some of these events leave a trace in the detector that is very similar to our signal, so distinguishing between signal and background events is challenging. SK has searched for this decay mode before and the strategy used was to apply simple cuts in reconstructed variables such as momentum and number of particles present in the event. By looking at many distributions of these variables (features) and applying hard boundary cuts on them, a selection criteria was used to determine if an event is coming from signal or background. Details can be found here [6].

3 Datasets and Inputs

In this section, the dataset(s) and/or input(s) being considered for the project should be thoroughly described, such as how they relate to the problem and why they should be used. Information such as how the dataset or input is (was) obtained, and the characteristics of the dataset or input, should be included with relevant references and citations as necessary. It should be clear how the dataset(s) or input(s) will be used in the project and whether their use is appropriate given the context of the problem.

4 Solution Statement

In this section, clearly describe a solution to the problem. The solution should be applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, describe the solution thoroughly such that it is clear that the solution is quantifiable (the solution can be expressed in mathematical or logical terms), measurable (the solution can be measured by some metric and clearly observed), and replicable (the solution can be reproduced and occurs more than once).

5 Benchmark Model

In this section, provide the details for a benchmark model or result that relates to the domain, problem statement, and intended solution. Ideally, the benchmark model or result contextualizes existing methods or known information in the domain and problem given, which could then be objectively compared to the solution. Describe how the benchmark model or result is measurable (can be measured by some metric and clearly observed) with thorough detail.

second.

⁶The detection of neutrinos in SK was awarded the Nobel prize in physics in 2015 for proving that neutrinos have mass [5].

6 Evaluation Metrics

In this section, propose at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model. The evaluation metric(s) you propose should be appropriate given the context of the data, the problem statement, and the intended solution. Describe how the evaluation metric(s) are derived and provide an example of their mathematical representations (if applicable). Complex evaluation metrics should be clearly defined and quantifiable (can be expressed in mathematical or logical terms).

7 Project Design

In this final section, summarize a theoretical workflow for approaching a solution given the problem. Provide thorough discussion for what strategies you may consider employing, what analysis of the data might be required before being used, or which algorithms will be considered for your implementation. The workflow and discussion that you provide should align with the qualities of the previous sections. Additionally, you are encouraged to include small visualizations, pseudocode, or diagrams to aid in describing the project design, but it is not required. The discussion should clearly outline your intended workflow of the capstone project.

8 Before submitting your proposal, ask yourself. . .

- Does the proposal you have written follow a well-organized structure similar to that of the project template?
- Is each section (particularly Solution Statement and Project Design) written in a clear, concise and specific fashion? Are there any ambiguous terms or phrases that need clarification?
- Would the intended audience of your project be able to understand your proposal?
- Have you properly proofread your proposal to assure there are minimal grammatical and spelling mistakes?
- Are all the resources used for this project correctly cited and referenced?

...

References

- [1] W. de Boer, **Grand Unified Theories and Supersymmetry in Particle Physics and Cosmology**. [arXiv:9402266](#)

- [2] Super-Kamiokande collaboration, **The Super-Kamiokande detector**, Nuclear Instruments and Methods in Physics Research A 501 (2003) 418-462. [SK detector](#)
- [3] For more see: [Wiki Cherenkov Radiation](#)
- [4] For more see: [Wiki Photomultipliers](#)
- [5] For more see: [2015 Nobel](#)
- [6] Super-Kamiokande collaboration, **Search for Proton Decay via $p \rightarrow \nu K^+$ using 260 kiloton-year data of Super-Kamiokande**, [arXiv:1408.1195](#)
- [7] For more: [2015 Nobel](#)