Classification of pixel tracks to improve track reconstruction from proton-proton collisions

Halle Purdom1, Jobin Joseph1, Kebur Fantahun1,Nibhrat Lohia1

1 Master of Science in Data Science, Southern Methodist University,  
Dallas, TX 75275 USA

hpurdom@smu.edu, jobinj@mail.smu.edu, kfantahun@mail.smu.edu

nlohia@mail.smu.edu

**Abstract.** In this paper we hope to leverage the ability of data science to unlock secrets about how particle collisions work. There is a facility in Europe (CERN) that has built a collider underground which can produce collision of particles at high speeds. The problem is that there are many detectors on the collider which produce a great amount of data. In this paper we are trying to decipher what factors are important to find the particle trajectories and predict future collisions.

1 Introduction

1. What is the topic? Why should we care? Why is this topic important to investigate?
2. What have the other experts discovered about the topic? (High level overview)
3. What are the gaps? What is the problem?
4. How will this research solve this problem?
   1. This research aims to use \_\_\_\_\_\_\_\_\_ to solve \_\_\_\_\_\_\_\_\_\_.

Particle collisions can inform us about what the universe is made of and how it works. We could find new particles, look at the properties of resultant particle debris and hopefully discover new physics.

In order to analyze particle collision experiments, the particles’ positions must be tracked as they leave a collision on their specific path. The CMS Silicon Pixel Detector at CERN contains many pixels that track paths of particles resulting from a collision. The tracker records the different positions a particle passes through after exiting a collision with enough precision for the particle track to be reconstructed.

In this study we are going to utilize machine learning to assist in path reconstruction for particles that are leaving a particle collision. Because there is a huge combinatorial background in the data collected, machine learning can be utilized to parse out the noise in the data to isolate the signal. From the data this looks like a general classification problem that can be modeled via different machine learning techniques like random forests, decision trees, and neural networks. This research aims to use machine learning to assist in path reconstruction of protons from particle collisions.

2 Literature Review

This section is on background information to help to reader understand why this research’s method will work.

[1] Applications of silicon strip and pixel-based particle tracking detectors

This article describes how the silicon strip can be used to help provide extra data about the LHC. There are different types of silicon detector systems, some experiments use strips and others planar detectors. The LHC can benefit with a silicon detector system, and it can provide an extra data feature that can be used when building ML models.

[2] Electrostatic charging due to individual particle-particle collisions.

This article describes how there are different factors that can influence a particle collision and its trajectory. Using video recording and particle tracking velocimetry they could see how the angle and the velocity of the initial particles caused a clear effect on the final particles. One effect that can be seen is the resultant charge that is left on the particle. This can be useful in our research as particle charges should be noted as possibly being an important feature to look at in our dataset.

[3] Graph Neural Networks for Particle Tracking and Reconstruction

Graph neural networks can be useful in particle tracking and reconstruction for particle physics data. Deep learning algorithms were developed initially for other tasks and may not be the most useful for particle physics purposes. Particle physics produces data that can be visualized in graphs, so GNN is a good choice. It can provide 3 dimensional points in tracking its particles.

[4] THE CMS PIXEL DETECTOR

In the large hadron collider, there are thousands of particles that need to be tracked by the CMS. The silicon pixel detector is designed to track these particles after particle collisions at enough precision to reconstruct these paths.

[5] Convolutional Neural Network for Track Seed Filtering at the CMS High-Level Trigger

Improvements to the LHC in general and thus the CMS detector overtime result in a rise in luminosity. This follows by producing a significantly large and difficult task of sifting through more noise than before when considering the pixel tracks that are created from proton-proton collisions. Filtering and clustering are methods that can help with sifting through the large amount of data to improve pixel tracks and thus higher-level particle tracking algorithms. Convolutional neural networks used by the researchers will help to improve results in learning how to classify real vs fake pixel tracks.

[6] Examining the event-shape dependent modifications to charged-particle transverse momentum spectra and elliptic flow in p-Pb collisions at energies available at the CERN Large Hadron Collider

This article highlights how there is not enough knowledge about proton-proton and proton-nucleus collisions. Using caution, the researchers work to utilize the models from hydrodynamic models to help with shape engineering. One of the parameters that they thought could correlate with the particle collision was ellipticity. This could be useful to us when conducting our research as we will be using data from the LHC as well for data. This could give us the framework to organize the data that comes from the LHC

[7] Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at root s=13 TeV

The CMS muon detector runs at high large hadron collider energy and instantaneous luminosity. Proton-proton data is used with a center of mass energy root s=13 TeV to study the system. This article provides much insight into how the CMS muon detector functions and how it performs and is evaluated based on its performance.

[8] The HEP.TrkX Project: Deep Learning for Particle Tracking

Particle tracking and reconstruction is used in HEP experiments, but the techniques scale quadratically worse as the detector increases in occupancy. This study uses machine learning algorithms to attempt to solve this problem. The paper looks at recurrent and convolutional neural networks from 2D to more complex models. It also looks at how to scale these models to an applicable dimension and with enough sparsity.

[9] A novel technique for the reconstruction and simulation of hits in the CMS pixel detector

The CMS pixel detector deals with the full process of understanding how pixels begin and end as a result of proton-proton collisions. This article describes new techniques for particle hit reconstruction. A cluster generated by CMS software is proposed to use new cluster shapes or templates in order to simulate and learn about pixel hits from an irradiated detector.

[10] A Numerical Study on Droplet-Particle Collision: Lamella Characterization. Flow, Turbulence and Combustion

This article talks about droplet‑particle collisions. The look into factors such as surface tension, numerical validation, statistical learning methods and computer modeling. This is useful as many industries need to calculate forces and volumes when making droplets or even spraying droplets. We can see in the statistical learning methods where we can convert a material that is heavy in physics such as particle collisions to something that we can somehow predict.

**Hypothesis:** We predict convolutional neural networks will be the most effective at track reconstruction because of their applications with image classification.

**FOR THE OUTLINE ONLY – You do an annotated bibliography (1 paragraph summary of each article that will be used in the paper).**

2.1 Theme

2.2 Citations -READ ME

The list of references is headed “References” and is not assigned a number. The list should be set in small print and placed at the end of your contribution, in front of the appendix, if one exists. Please do not insert a pagebreak before the list of references if the page is not filled. An example is given at the end of this information sheet. For citations in the text please use square brackets and consecutive numbers: [1], [2], [3], etc. Use **APA format** in the reference section. You can choose to either have it alphabetical order or order of which it is shown in the paper.

**Hypothesis at the end of your literature Review**

3 Methods

1. Data
2. Where are you getting the data? Or where are you thinking you can find the data? We are planning to get our data from CERN Open Data <https://opendata.cern.ch/record/12303>
3. Methods we plan to use MVA methods including Manova, PCA, LDA and Multidimensional scaling.

4 Results

1. What you hope to find in your research? Accept or reject the hypothesis

\*\*This Section is for statistical jargon and tables/Figures. Results are facts.

We hope to find the best algorithm for the path reconstruction of particles. This will look like an optimization problem for a classification task. We will look at different machine learning methods, find the best algorithm to classify the data, then optimize the parameters of that algorithm.

5 Discussion

\*\*\*Do not add New Results. This section is to apply and interpretate the results into lay terms.

\*\*\* Write questions you hope to answer in your research.

1. Interpretations: What do the results mean?
2. Implications: Why do the results matter? How should the reader apply these findings?
3. What stood out as interesting/unique/unexpected?
4. Limitations
   1. What challenges occurred during analysis?
5. Ethics
6. Future Research
   1. Are there areas of research where others can pick up and go deeper?

Some questions that we hope to get answered are as follows. How do machine learning models work with the data sets from the LHC? Can particle collisions follow a trajectory that is predictable? Do factors such as the charge and size of the particles make a difference on how a collision will occur. These results matter as this could provide proof of why ML models should be deployed in real time during physics experiments.

6 Conclusion

2 paragraphs max on the overall findings and summary of the research.

We have not actually conducted research yet outside of the literature review. Once research is finished, conclusion will go here.

Acknowledgments. The heading should be treated as a 3rd level heading and should not be assigned a number.

References

1. Allport. (2019). Applications of silicon strip and pixel-based particle tracking detectors. Nature Reviews Physics, 1(9), 567–576. <https://doi.org/10.1038/s42254-019-0081-z>
2. Chowdhury, Ray, M., Passalacqua, A., Mehrani, P., & Sowinski, A. (2021). Electrostatic charging due to individual particle-particle collisions. *Powder Technology*, 381, 352–365. <https://doi.org/10.1016/j.powtec.2020.12.012>
3. Duarte, & Vlimant, J.-R. (2020). *Graph Neural Networks for Particle Tracking and Reconstruction*.
4. ERDMANN. (2010). THE CMS PIXEL DETECTOR. *International Journal of Modern Physics. A, Particles and Fields, Gravitation, Cosmology*, 25(7), 1315–1337. <https://doi.org/10.1142/S0217751X10049098>
5. Florio, Pantaleo, F., & Carta, A. (2018). Convolutional Neural Network for Track Seed Filtering at the CMS High-Level Trigger. *Journal of Physics. Conference Series, 1085*(4), 42040–. <https://doi.org/10.1088/1742-6596/1085/4/042040>
6. Kar, Choudhury, S., Zhang, X., & Zhou, D. (2020). Examining the event-shape dependent modifications to charged-particle transverse momentum spectra and elliptic flow in p-Pb collisions at energies available at the CERN Large Hadron Collider. <https://doi.org/10.1103/PhysRevC.102.044901>
7. Madlener, Van Remortel, N., Tavernier, S., Marinov, A., Bruno, G., Krintiras, G., Chinellato, J., Ahuja, S., Zhang, F., Cabrera, A., Godinovic, N., Ather, M. W., Dewanjee, R. K., Siikonen, H., Grenier, G., Schulz, J., Borras, K., Kleinwort, C., Zenaiev, O., … Heindl, S. M. (2018). Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at root s=13 TeV. *Journal of Instrumentation,* 13. <https://doi.org/10.1088/1748-0221/13/06/P06015>
8. Tsaris, Anderson, D., Bendavid, J., Calafiura, P., Cerati, G., Esseiva, J., Farrell, S., Gray, L., Kapoor, K., Kowalkowski, J., Mudigonda, M., Prabhat, Spentzouris, P., Spiropoulou, M., Vlimant, J.-R., Zheng, S., & Zurawski, D. (2018). The HEP.TrkX Project: Deep Learning for Particle Tracking. *Journal of Physics. Conference Series*, 1085(4), 42023–. <https://doi.org/10.1088/1742-6596/1085/4/042023>
9. Swartz, M., Chiochia, V., Fehling, D., Maksimovic, P., & Giurgiu, G. (2007). *A new technique for the reconstruction, validation, and simulation of hits in the CMS pixel detector* (No. CMS-NOTE-2007-033, p. 035). CERN-CMS-NOTE-2007-033. DOI: <https://doi.org/10.22323/1.057.0035>
10. Vilela, & de Souza, F. J. (2020). A Numerical Study on Droplet-Particle Collision: Lamella Characterization. Flow, Turbulence and Combustion, 105(4), 965–987. <https://doi.org/10.1007/s10494-020-00153-x>

Appendix

Use if needed for additional information