Matrix Factorization for Primary Vertex Reconstruction in LHCb

The Large Hadron Collider beauty detector (LHCb) aims to answer the question "Why is matter more prevalent than anti-matter in the observable Universe" by studying beauty/bottom quarks (b-quarks) and their anti-matter counterpart (b anti-quarks) which were abundant at the time after Big-Bang. The LHCb detector is expected to produce 4Tb/sec of data when it starts running in 2021 after the LS2 maintenance is completed [4]. This calls for efficient track and primary vertex (PV) reconstruction algorithms to analyze data effectively within a practical time frame using relatively low powered computers.

Existing methods of PV reconstruction mainly focus on using clustering algorithms such as k-means and Hierarchical Agglomerative Clustering (HAC) to locate PVs. Through research done as part of the IRIS-HEP project, it was observed that the distance matrix between reconstructed tracks displayed an interesting block structure [Fig 1]. Matrix factorization is a tool heavily used in the machine learning community, and algorithms such as non-negative matrix tri-factorization (NMTF) could be used to produce block-structure matrices [2], which could be used to search for interesting patterns within the data. Different implementations of NMTF also allows for different constraints to be added to the factor matrices, constraints such as orthogonality [1] which results in cleaner block structures.

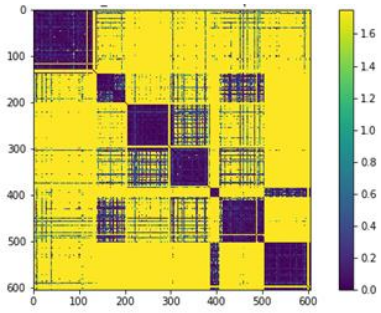

Figure 1. Distance matrix capturing distance between tracks

This project aims to develop a python workflow to test the viability of matrix factorization for PV identification, specifically Matrix tri-factorization with orthogonality constraints [1]. It has been shown that Non-negative Matrix Factorization (NMF) with orthogonality constraints performs better than k-means and spectral clustering [1][5], and NMTF with orthogonality constraints produces factors with less noise than NMF [1]. The algorithms developed through this project will be tested and compared against existing clustering algorithms used in the literature. This project will be carried out under the supervision of Gowtham Atluri (Asst. Professor, Dept. EECS, UC) and Mike Sokoloff (Professor, Dept. Physics, UC).

Background:

I am a computer science undergraduate student at the University of Cincinnati, entering my 2nd year of college. I have prior exposure to python and have implemented a linear algebra library and a simple neural network library in Java. I am also the country topper of India for the CAIE As level Physics examination.

Deliverables:

- Implementation of NMTF with orthogonality constraints in Python
- Visualization of PVs reconstructed using the algorithm developed
- Evaluation and comparison results of the developed algorithm against existing methods
- Final report with findings

All deliverables will be made publicly available via GitHub. A report of the deliverables and present findings will be made.

Timeline:

Weeks	**Planned activity**
1	Gain understanding of Matrix Factorization and run available implementations
2&3	Gain understanding of NMTF with orthogonality constraints and implementation in python
4 & 5	Modify and tune the algorithm to better suit PV reconstruction
6	Run performance and efficiency tests on the algorithm, and plot results
7	Perform comparison with existing PV reconstruction clustering algorithms
8 & 9	Implement an automated system to find the optimal number of clusters to factorize for
10	Prepare final report with findings

References:

[1] Del Buono, Nicoletta & Pio, Gianvito. (2015). Non-negative Matrix Tri-Factorization for co-clustering: An analysis of the block matrix. Information Sciences. 301. 13-26. 10.1016/j.ins.2014.12.058.

[2] Long, B., Zhang, Z., & Yu, P. S. (2010). *Relational data clustering: Models, algorithms, and applications*. Boca Raton: Chapman & Hall/CRC.

[3] Meloni, F. "Primary Vertex Reconstruction with the ATLAS Detector." *Journal of Instrumentation*, vol. 11, no. 12, 2016, doi:10.1088/1748-0221/11/12/c12060.

[4] Transforming LHCb: What's in store for the next two years? (n.d.). Retrieved June 03, 2020, from https://home.cern/news/news/experiments/transforming-lhcb-whats-store-next-two-years

[5] Xu, W., Liu, X., & Gong, Y. (2003). Document clustering based on non-negative matrix factorization. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval - SIGIR '03*. doi:10.1145/860435.860485