Contribution ID: **567**   Type: **Sectional talk**

Anticipating Data Demand in HEP: A Transformer Approach

Tuesday 8 July 2025 18:00 (15 minutes)

Modern high-energy physics (HEP) experiments generate and store vast volumes of data, which users access through complex and irregular patterns. Efficient data management in such environments requires accurate forecasting of dataset popularity to optimize storage, caching, and data distribution strategies. In this work, we propose an approach for predicting future dataset access patterns using transformer-based deep learning models. By leveraging historical logs of user interactions with HEP datasets, our method captures temporal dependencies and contextual signals to forecast both short- and medium-term data demand.

We evaluate our approach on real HEP access logs and conduct a comparative analysis of the accuracy of the proposed transformer-based method with previously used methods, including Facebook Prophet, Random Forest, and LSTM. Our results suggest that transformer architectures are a powerful tool for proactive data management in large-scale scientific computing environments. Although the proposed method is demonstrated using user analysis data access patterns, it is equally applicable to production data popularity forecasting.

Additionally, we implement a custom evaluation metric focused on the total sum of future accesses compared to the sum of predicted accesses, rather than relying on traditional day-by-day accuracy metrics.

Authors: GRIGORIEVA, Maria (Moscow State University);   SHUBIN, Mikhail (Lomonosov Moscow State University);   POPOVA, Nina (Lomonosov Moscow State University)

Presenter: SHUBIN, Mikhail (Lomonosov Moscow State University)

Session Classification: Computing for MegaScience Projects