

CLAS 6GeV Data Mining: Jefferson Lab proposal for DOE NP DE-FOA-0002490

David Lawrence

March 2021

Abstract

Extracting meaningful and subtle physics results from large complex data sets has always been challenging. AI/ML provides a new tool that could be used to aid in this effort, potentially reaching signals that are practically impossible to attain using traditional methods with limited resources. In order to wield this new tool effectively, we must gain experience and develop the proper techniques using data sets where traditional methods have already been applied. The CLAS data set from the 6GeV Science program at Jefferson Lab provides a perfect resource for developing such technology.

1 Introduction

The CLAS 6GeV data set has produced over 200 publications and nearly 200 PhD Theses. It is well understood and readily available. CLAS was a 4π detector which took data with fairly loose triggering resulting in a data set ripe for mining any number of results. The data set is also fairly recent and much of the expertise on CLAS still exists at JLab and within the CLAS and CLAS12 collaborations. The proposed project would have the following goals:

- Reproduce 2-3 specific, published results using AI
- Produce a large curated data set well-formed for use with AI that can be shared with other facilities
- Produce a plan for applying the techniques developed to the CLAS12 data set currently being acquired

2 Potential Target Analysis

Early CLAS-6 data produced a set of 3 papers measuring the differential cross-section $d\sigma/dt$ of the ϕ [1], ρ [2], and ω [3]. The data set was based on 70M triggers with individual channels consisting of event counts around $\sim 100k$. The project would review the analysis notes for these three papers and investigate how AI might be used to simultaneously produce

the differential cross-sections for these channels. This is interesting since the channels are backgrounds for each other. An early objective could be simply having an AI identify how many classifiable channels are present. This would be an exercise in unsupervised learning.

3 Resources

The project would require some resources. Some of this would be absorbed into existing operations as small perturbations. A list of required items follows:

- At least one dedicated post-doc
- Portions of existing staff at JLab with expertise in the CLAS data set
- Experts in AI from the JLab staff and affiliated Universities
- Some dedicated equipment ($\sim 10k$ for personal computers)
- Additional Data Center capacity (to host curated data set)
- Support from the JLab CST division
- Support from JLab Library Services

References

- [1] E. Anciant et al. Photoproduction of $\varphi(1020)$ mesons on the proton at large momentum transfer. *Phys. Rev. Lett.*, 85:4682–4686, Nov 2000.
- [2] M. Battaglieri et al. Photoproduction of the ρ^0 meson on the proton at large momentum transfer. *Phys. Rev. Lett.*, 87:172002, Oct 2001.
- [3] M. Battaglieri et al. Photoproduction of the ω meson on the proton at large momentum transfer. *Phys. Rev. Lett.*, 90:022002, Jan 2003.