Astronomical Data Analysis Software and Systems XXVI
ASP Conference Series, Vol. 521
Marco Molinaro, Keith Shortridge, and Fabio Pasian, eds.

Real Time Streaming Analysis of IACT Data

Kai Bruegge,[2] Jan Adam,[2] Max Ahnen,[3] Dominik Baack,[2] Matteo Balbo,[4] Matthias Bergmann,[5] Adrian Biland,Christian Bockermann,[1] Jens Buss,[2], [3] Michael Blank,[5] Thomas Bretz,[3] Anton Dmytriiev,[4] Daniela Dorner,[5] Alexey Egorov,[1] Sabrina Einecke,[2] Christina Hempfling,[5] Dorothee Hildebrand,[3] Gareth Hughes,[3] Lena Linhoff,[2] Karl Mannheim,[5] Katharina Morik,[1] Sebastian A. Mueller,[3] Dominik Neise,[3] Andrii Neronov,[4] Maximilian Noethe,[2] Aleksander Paravac,[5] Felicitas Pauss,[3] Wolfgang Rhode,[2] Tim Ruhe,[2] Amit Shukla,[3] Fabian Temme,[2] Julia Thaele, [2] and Roland Walter[4]

[1] *Lehrstuhl für Experimentelle Physik 5, TU Dortmund*
`firstname.lastname@tu-dortmund.de`

[2] *Lehrstuhl für künstliche Intelligenz, Informatik, TU Dortmund*

[3] *ETH Zurich, Institute for Particle Physics*

[4] *University of Geneva, ISDC Data Center for Astrophysics*

[5] *Universität Würzburg, Institute for Theoretical Physics and Astrophysics*

Abstract. FACT, The First G-APD Cherenkov telescope, is an Imaging Atmospheric Telescope (IACT) located at the Observatorio del Roque de los Muchachos on the Canary Island of La Palma. FACT is dedicated to monitor bright Blazars in the northern sky at TeV energies. FACT's continuous monitoring produces a dense sample of observation points for various active galactic nuclei (AGN). To help understand the mechanisms of cosmic ray acceleration, these sources need to be observed simultaneously over a wide range of wavelengths. A fast analysis process is needed to alert other experiments when the output of an observed source changes. This proceeding presents some details about FACT's real time analysis (RTA). A program which uses a streaming pipeline to apply machine learning models to the measured data in real time. All of the software presented here is freely available on GitHub at

`https://github.com/fact-project`

1. The First G-APD Cherenkov Telecope

FACT, the First G-APD Cherenkov Telescope, is an Imaging Atmospheric Telescope (IACT) located at the Observatorio del Roque de los Muchachos on the Canary Island of La Palma. At the time of its construction in 2011 it was the first IACT to use Silicon Photomultipliers (SiPMs) instead of the conventional Photomultiplier Tubes (Biland et al. 2014). Monitoring active galactic nuclei (AGN) over wide energy bands is essential for understanding the possible acceleration mechanisms of gamma-rays, UHE cosmic rays and extragalactic neutrinos in AGN. Many of the observed AGNs can show volatile behavior and sometimes dramatically change the amount of energy they output within minutes. Alerting other experiments, as fast as possible, of these flaring sources is an essential task for FACT.

2. Raw Data Processing

A raw FACT event contains 300 samples recorded in a 150 ns event for each of its 1440 pixels. This creates up to one terabyte of data per night. The FACT-Tools are a collection of tools build on top of the `streams-framework` (Bockermann 2015). They reduce FACT raw data into image parameters that can be used for supervised machine learning methods (Bockermann et al. 2015). The `streams-framework` is a data streaming environment written in the Java programming language. The purpose of the `streams-framework` is to provide an abstract way to model the data flow of an application. It comes with a number of fundamental building blocks to define a data flow graph using an Extensible Markup Language (XML) based description. The data source, in this case the telescope, continuously emits small packets of data, the data items. These elementary units of data are processed by small functional units called processors. After the processing has finished, the data can be written to a file or a database.

3. Machine Learning on FACT Data

When gamma rays or cosmic rays hit Earth's atmosphere they create extensive air showers. The Cherenkov light emitted by these showers is measured by IACTs. Machine learning is used for two major tasks when analyzing IACT data: energy estimation and gamma-hadron separation. The first uses the observed values to estimate the energy of the cosmic primary particle which started the air shower. The later classifies the events by the type of the primary particle to suppress the main background in FACT's data: cosmic ray induced air showers. The popular Python machine learning library scikit-learn (Pedregosa et al. 2011) is used to train and then store the models into the PMML (DMG 2016) format. The stored PMML model can be shared between programming languages and applications. A RandomForest classifier with 150 trees is applied to the data stream during the Real-Time-Analysis to perform background suppression. A total of 52 features, image parameters, are used to train the model. All are calculated using the FACT-Tools. Online energy estimation for the Real Time Analysis works in a similar fashion.

4. Monitoring Sources

IACTs record so-called light curves which show the gamma ray flux versus time. One of FACT's main goals is the long-term monitoring of bright AGN type gamma ray sources. These sources can show variable behavior in terms of both total brightness and spectral properties. If a source shows a sudden increase in flux, it is considered to be in a flaring state. In case a flare is detected, FACT can alert other experiments to trigger multi-wavelengths observations. Figure 1 shows excess rates of the AGN Markarian 501 over time. A flare on the 24th of June is clearly visible and was confirmed by other experiments as well (Stegmann 2014).

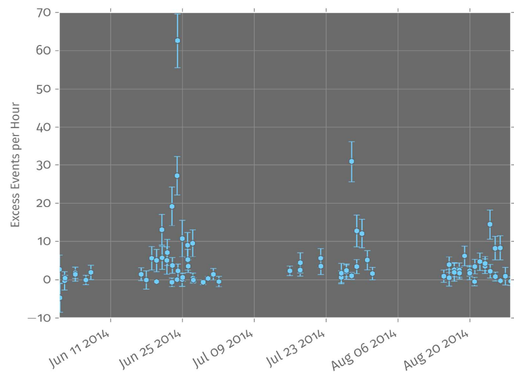

Figure 1. Observations of the active galactic nucleus Markarian 501 during 2014. This plot was produced by applying the RTA process to archival data (Bruegge 2016).

5. The Real Time Analysis

The RTA combines all the steps needed to measure excess rates from observed sources. Starting with the raw telescope data, the RTA produces high level analysis results, such as excess rates, which can be stored in a database and displayed on a website. The single steps can be summarized as follows:

Telescope Data The telescope emits 50 events per second on average. This leads to a mean data rate of 43.2 MB per second and up to 1TB of collected raw data per night.

Calibration Calibration data is applied to the data stream to convert raw sensor output to physical units. This is the most time consuming step and is parallelized to several CPUs.

Image Parameter The number of photons and their respective arrival times are extracted from the raw data. Camera pixels not containing any Cherenkov photons are dropped and parameters describing the final image are calculated.

Energy Estimation Use pre-trained regression models on the parameters to estimate the energy of signal events.

Classification Use pre-trained models to separate signal from background events.

Storage Results of the analysis are stored in an SQL Database. A web server provides public access to high-level results in real time.

The whole RTA process, including the web server, is running in single Java executable running on a machine at the telescope site. Figure 2 shows a screenshot of the RTA website presenting results live in a web browser.

6. Summary & Conclusion

The Real-Time-Analysis for FACT has strong requirements on usability, scalability, maintainability and processing speed. Results need to be available as soon as possible.

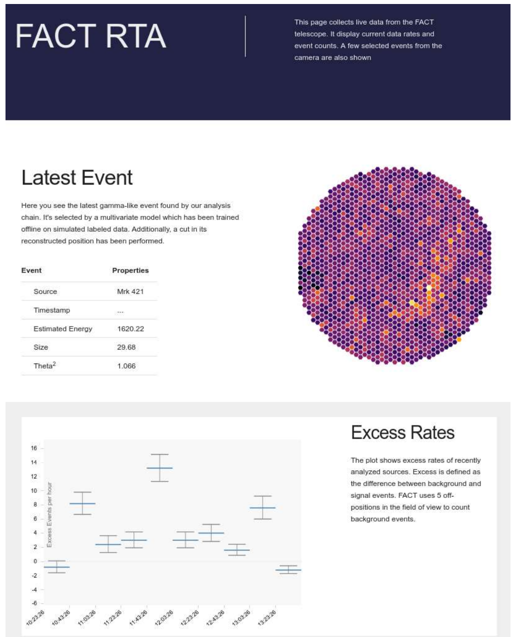

Figure 2. A screenshot of the web interface for the RTA. Excess rates are plotted live in the browser. It can be found at `www.fact-project.org` (Bruegge 2016)

At the same time they mus be persisted in a reliable way which can be used for further analysis. This was achieved by applying pre-trained models to the data stream from the telescope. Additionally the RTA features a web interface showing live analysis results and data rates and serves as a cross-check for other analyses applied to FACT data.

Acknowledgments. This work has been supported by the DFG, Collaborative Research Center SFB 876, project C3 (`http://sfb876.tu-dortmund.de/`).

References

Biland, A., et al. 2014, JINST, 9, P10012

Bockermann, C. 2015, Ph.D. thesis, TU Dortmund University. URL `https://eldorado.tu-dortmund.de/handle/2003/34363`

Bockermann, C., et al. 2015, in Proceedings of the European Conference on Machine Learning (ECML) (Springer Berlin Heidelberg)

Bruegge, K. 2016, Analysis of fact data from online processing to real time excess rates. URL `http://app.tu-dortmund.de/cms/de/home/Masterarbeiten/Bisherige-Masterarbeiten/index.html`

DMG 2016, Pmml 4.2 - general structure. URL `http://dmg.org/pmml/v4-2-1/GeneralStructure.html`

Pedregosa, F., et al. 2011, Journal of Machine Learning Research, 12, 2825

Stegmann, C. 2014, Increased vhe activity from mrk 501 detected with h.e.s.s. URL `www.astronomerstelegram.org?read=6268`