GSFC Activities in AI and Data Science

Nargess Memarsadeghi

Mark Carroll, Dan Duffy, Brian Thomas, Barbara Thompson, and Joe Foster

NASA GSFC

February 9, 2021

AI Center of Excellence Introduction

- Initial idea proposed to NASA Digital Transformation (DT) committee as a way to integrate AI/ML work across the agency
- Started in 2019 with seed funding from the Chief Technologist Office at GSFC
- Continued with funds from Computational Information Science and Technology Office (Code 606) funds
- Supplemented with funds from a Science Task Groups supported by code 600
- *Inaugurated in Spring 2020*

AI Center of Excellence (AICOE)

- Website: https://ai.gsfc.nasa.gov
- MS Teams: GSFC AI Center of Excellence , 145 members
- Focal point for information
  - GSFC AI/ML Communities of Practice
  - Goddard ML Academy
  - AI/ML Projects
  - Events (seminars, hackathons,..), conferences and workshops
  - Computational Resources (NCCS Scientific Computing, GSFC Cloud Computing program)
  - Funding opportunities
  - Publications

GSFC AI/ML Communities of Practice

- Computational and Information Science and Technology Office's (CISTO) Data Science Group
- The Intelligent Systems and Data Analysis Technologies (ISDAT)
- The Center for Helioanalytics
- The Goddard Cloud Computing Program (GCC)
- The Goddard Machine Learning Academy
- ML Science Task Group (STG)

AI Center of Excellence (AICOE)

- Forming *partnerships*
- Developing the next generation workforce
- *Accelerating scientific discovery*

AICOE Partnerships

- Across organizations
  - GSFC scientists and engineers (IRADs, GCIF, STGs, ROSES,….)

- Academia
  - EPSCoR and Space Grant (POC: James Harrington)
  - Seed money grants to University to work with a NASA partner
  - Opportunity to collaborate with Academia

- Industry
  - Small businesses: SBIRS
  - Computational hardware and software
    - AWS, Google, NVIDIA, IBM

AICOE Events and Activities

- Seminars since April 2020:
  - Hosted or co-hosted 24 seminars
  - Featuring various NASA programs: IRADs, ECF, NSTGRO, STGs, EPSCoR, SBIR,…
- *Workshop on AI/ML use in Strategic Planning and Science Prioritization* (May 2020)
- AWS Immersion Day (June 17, 2020)
- Helio Hackweek 2020 (August 20-28, 2020).
  - Enhance skills in applying AI/ML for scientific discovery
  - Raise awareness of High-End Computing (HEC), Graphics Processing Unit (GPU), Commercial Cloud.

Workshop on AI/ML use in Strategic Planning and Science Prioritization

- 3 Days, ~40 participants: scientists and AI/ML practitioners.
- Forecasting 2010, 2020 Astronomy Decadal Survey priorities {We have a poster paper at AAS, research paper in draft}
- Developed a tool, “TEV” (Topic Emergence Visualizer). Identified several Test Cases.
- Workshop report, *coming soon.*
- Point of Contact: Dr. Brian Thomas.

AWS Immersion Day and Projects

(*on going*)

- How to Use Amazon Web Services (AWS) for AI/ML Projects
- Information on AWS AI/ML Services
- Solicited Projects:
  - AI/ML problem that requires significant compute
  - Finite duration (completion within 3 – 6 months)
  - Potential to result in publication
  - Topics from Earth, Helio, Solar System and Astro science areas as well as engineering
- Selected Projects (*on going*)
  - Automated Detection of CME in SOHO/LASCO
  - Spatiotemporal Drought Estimation Quantification
  - Solar Irradiance Evaluation SDO/HMI – SORCE/TSIS-1
  - Remote Sensing Water Quality Parameters
  - Hurricane Prediction Deep Learning Mode

Helio Hackweek 2020

- Projects were “pitched” to the participants in early August to determine which project each participant was interested in
- Event held virtually August 20 – 27, 2020
- Tutorials presented by ASTG and NVIDIA August 20 and 21
- Five teams worked on individual projects August 25 – 27
- Report back on August 28
- Over 40 participants from 16 institutions, domestic and international
- Feedback from organizers and participants
  - Good balance between Helio science disciplines
  - Participants fully engaged the process
  - Skill sets of participants was well balanced on each team
  - Great support from industry partner

GSFC AI/ML Projects

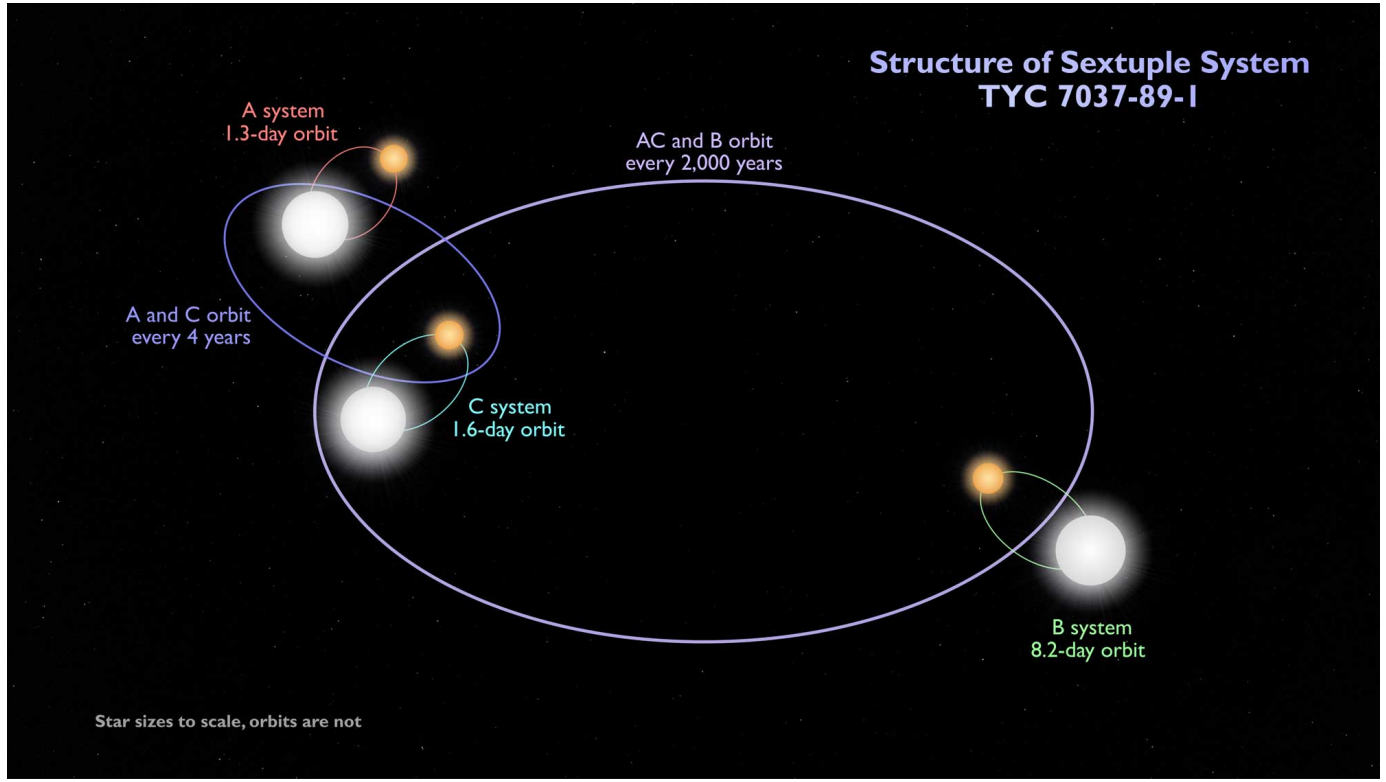
(A few examples)

- Identification of AI/ML expertise at GSFC and the Agency via published papers
- From Data to Discovery in Astronomical Datasets
- Application of Random Forest to Quantify Lake Depth in Arctic Lakes
- Object Identification in Sub Meter Commercial Very High-Resolution Data Using Convolutional Neural Networks
- MERRA/Max: Harnessing the Potential of Climate Model Outputs in Studies of Ecosystem Change
- Machine Learning for 3D Lunar Data Analysis
- STEREO Space Weather Beacon Imagery Quality Enhancement
- Community Concept Mapping for Data-to-Decisions

GSFC Data Science News!

"The discovery: TYC 7037-89-1 is the first six-star system ever found where all of the stars participate in eclipses, a discovery made by NASA's Transiting Exoplanet Survey Satellite (TESS). The system is located about 1,900 light-years away in the constellation Eridanus."

"The discoverers: An international team, led by data scientist Brian P. Powell and astrophysicist Veselin Kostov at Goddard, made the discovery using TESS data."



Reference and Image Credit: https://exoplanets.nasa.gov/news/1672/discovery-alert-first-six-star-system-where-all-six-stars-undergo-eclipses/

AICOE Partnerships: Academia

- Established the Academic Research Collaborative (ARC) led by *James Harrington*
- Identified research areas of interest to GSFC scientists and engineers, drafted an Appendix for the EPSCoR 2020 Solicitations for CISTO, Code 606
    - **Artificial Intelligence and Machine Learning (AI/ML)**
    - Augmented Reality/Virtual Reality/Mixed Reality (AR/VR/MR)
    - **High Performance Computing; Applications to Exascale**
    - Citizen Science
- Proposal reviews
- Providing Technical Monitors
- Providing Computational Resources

AICOE Partnerships: Academia

- 2020 EPSCoR Awardees, paired with a GSFC Collaborator/Technical Monitor
  - *Machine Learning for Environmental Feature Recognition (Wyoming)*
  - *A Generic Data-Driven Framework via Physics-Informed Deep Learning (South Carolina)*
  - *Ray Events Detected by Radio Probes (Delaware)*
  - *Framework for Increasing Generalizability of Machine Learning Models (Kansas)*
  - *AI-based framework for real time RFI detection in radio astronomy data (West Virginia)*
  - *Other: A Citizen Science Game EPSCoR R2 proposal (University of Vermont)*

2021 EPSCoR Solicitation

- Computational and Information Sciences and Technology Office (CISTO) Computational and Technological *Advances for Scientific Discovery via AI/ML Modeling and Development Hackathons.*
  - Design a year of AI/ML education, training, networking and building of machine learning models in close collaboration with NASA project senior science and engineering mentors
  - "Hack Weeks as a Model" (https://www.pnas.org/content/115/36/8872)
  - SMD FDL Model
- Assessing and qualifying Citizen Science Labeling for Training Data for GLOBE Observer Mosquito and Land Cover protocols improving data quality within the GLOBE Observer community.
- Onboard Satellite Fault Diagnosis using Machine Learning
- Application of Machine Learning to High-Resolution Earth System Model Data
- *Responses are currently under review,* were due Feb. 5, 2021.

Looking Ahead

- Goddard has internally supported many cross-cutting data science activities through IRADs and Science Task Groups.
- Machine Learning for Everyone: A Model-Based Approach
  - Led by Dr. Barbara Thompson
  - FY21 with GSFC's Science Task Group Funding
  - Capture the full "Life Cycle" of a data science/machine learning project. The factors behind each decision point will be used to produce a MBx deployable methodology and toolkit

Summary and Next Steps

Seeking:

- Proposals
- Reviewers
- Speakers

Offering Opportunities to:

- Receive consultation services
- Access High-end computing environments
- Form partnerships
- Learn about various computational resources
- Learn about various funding opportunities

Questions?

Thank you!

Please do not hesitate to contact us