MEDWIN PUBLISHERS
Committed to Create Value for Researchers

Open Access Journal of Astronomy
ISSN :2996-6701

Big Data in the Astronomical Field

Yuan P*

Guangdong Baiyun University, China

***Corresponding author:** PU Yuan, Guangdong Baiyun University Baiyun, Guangzhou, Guangdong Province, China, Email: puyuan@baiyunu.edu.cn

Mini Review

Volume 2 Issue 1

Received Date: June 05, 2024

Published Date: June 20, 2024

DOI: 10.23880/oaja-16000116

Abstract

From the beginning, astronomy is in the big data era. There are plenty of data sizes, types, sources, resolutions, etc. for the astronomical data. This paper gives a short description of astronomical data, and provides the links to some famous telescopes. Furthermore, this paper gives a short summary for AI (Artificial Intelligence) applications in astronomy. It calls on more AI researchers to join the research in astronomical data.

Keywords: Big Data; Astronomical Data; AI

Abbreviations: EM: Electromagnetic Radiation; GWs: Gravitational Waves; LAMOST: Large Sky Area Multi-Object Fiber Spectroscopy Telescope; IRIS: Interface Region Imaging Spectrograph; HERA: Hydrogen Epoch of Reionization Array; FAST: Five-hundred-meter Aperture Spherical radio Telescope; WISE: Wide-field Infrared Survey Explorer; SDSS: Sloan Digital Sky Survey.

Introduction

The word 'big data' is a commonly used term now, which has no an official definition but describes the explosive growth in data volume, velocity, and variety [1]. It becomes popular along with the growth of social media in internet. But in fact, the dawn of big data is with the operation of new technically advance telescopes at the very beginning. Hundreds of telescopes around the world are collecting different types of astronomical data day and night, so that 'big data' is firstly generated in human's history.

Now, with the fast development of techniques, more and more telescopes are built or being built or in planning. Multiwavelength and multi-messenger studies are becoming popular in current research [2].

From γ-ray to radio, from neutrinos to gravitational waves, the family of astronomical datasets is growing faster and faster [3]. And the volume of the datasets is also growing very fast. For example, the spectral amount of LAMOST (Large Sky Area Multi-Object Fiber Spectroscopy Telescope) has reached 46,941,395 for its medium resolution, and 11,939,296 for its low resolution [4].

Astronomical data has different types: spectrum looks like an electrocardiogram which is a kind of 1D signal; photometric image is from photometric survey and indeed a 2D image, with different bands (generally 5, u(305.5-403.0nm), g(379.8-555.3nm), r(541.8-699.4nm), i(669.2-840.0nm), z(796.5-1087.3nm)) [5]; infrared image is also a kind of 2D image with 4 bands (W1(2.8-3.8μm), W2(4.1-5.2μm), W3(7.5-16.5μm), W4(20-28μm)) [6]. Of course, there are more types of astronomical data, and this paper will give a summary about this.

Generally speaking, big data has 5 'V' characteristics: Volume, Variety, Velocity, Veracity, Value [1]. Sometimes there are more: Variability, Visualization. Astronomical data has all of these. This brings a lot of difficulties for astronomers to handle the huge datasets. Without AI, they may spend

~ 80% of their time for purely cleaning and preparing the observationally obtained data, which becomes a disaster for astronomers because this is not research for them. The tedious 'mechanical works' include checking data (types of columns, names, duplicate entries, outlier values, variables, scaling, normalizations, etc.), selecting data (classes, parameters, locations, etc.), finding useful/relative/special data etc.

AI is the right tool to handle these 'mechanical works'. 'Research in AI has focused chiefly on the following components of intelligence: learning, reasoning, problem solving, perception, and using language' [7]. And AI learns from data, especially big data. 'Although many AI technologies have been in existence for several decades, only now are they able to take advantage of datasets of sufficient size to provide meaningful learning and results' [8]. The 'must' techniques like clustering, classification in astronomical data processing are the basic ones in AI, as well as determining data type, data cleaning, exploration, visualization, feature selection, identification, etc. Therefore, with the fast growth of astronomical big data, AI is playing a more and more important role in astronomy. On the other hand, big data can help obtain correct results over human errors, and it is the fuel of AI engine.

Astronomical Datasets

According to the dimensions of data, astronomical datasets can be classified as 1D, 2D, 3D; according to time-dependency, there are time-variant, and time-invariant; according to multi-messenger, there are electromagnetic radiation (EM), gravitational waves (GWs), neutrinos and cosmic rays [9,10]. And astronomical datasets are often along with telescopes. For one telescope, generally there is at least one dataset published. And the type of dataset depends on the type of the telescope.

Cosmic rays are charged particles deflected by the magnetic fields between and within galaxies, so hard to be traced back to their origins.

Neutrinos are neutral and can be detected by the instruments like IceCube [11]. And the newest data release is in https://icecube.wisc.edu/data-releases/2023/07/icecube-hese-12-year-data-release/. GWs may come from compact binary merger. The most famous detector is from LIGO (Laser Interferometer Gravitational-wave Observatory) [12]. A dataset description file can be found by https://dcc.ligo.org/public/0009/M1000066/027/LIGO-M1000066-v27.pdf. And the data release link is https://gwosc.org/eventapi/html/.

As for EM, there are radio, microwave, infrared, optical, X-ray, gamma-ray. Interface Region Imaging Spectrograph (IRIS) obtains high, resolution UV spectra and images of the sun's chromosphere [13]. And its dataset is in https://iris.lmsal.com/data.html. One of the most famous radio datasets is from FAST (Five-hundred-meter Aperture Spherical radio Telescope), and one link is https://blinkverse.alkaidos.cn/ #/availability. The most famous infrared dataset is from WISE (Wide-field Infrared Survey Explorer), and its link is https://wise.ssl.berkeley.edu/astronomers.html. As for optical datasets, SDSS (Sloan Digital Sky Survey) is one of the most famous. And its link is https://www.sdss.org/dr18/. It has the biggest optical image dataset. When talking about spectrum, LAMOST has the biggest spectral dataset in the world, whose link is https://www.lamost.org/lmusers/. An X-ray dataset can be found in the link of https://www.cosmos.esa.int/web/xmm-newton/xsa, which is belonging to Max Planck Institute [14].

INTEGRAL is the International Gamma-Ray Astrophysics Laboratory of the European Space Agency [15]. It observes the Universe in the X-ray and soft gamma-ray band. And its dataset is in http://www.isdc.unige.ch/integral/archive.

The link https://www.cosmos.esa.int/web/esdc gives the 'ESAC SCIENCE DATA CENTRE', which integrates a lot of astronomical datasets from the European Space Agency's website. Inside it there is a famous plan called Gaia [16], which can collect 3D data for celestial bodies.

Currently there are hundreds of telescopes collecting huge data daily. And it's difficult to list all of them in one paper. This section just gives some famous ones, while Table 1 as an uncompleted table may be a short summary.

Name of Telescopes (Websites)	Approx. size of dataset	Models of AI used	Efficiency	Significant Result
IceCube (https://icecube.wisc.edu/)	4GB	A decoherence model	High	Searching for Decoherence from Quantum Gravity
LIGO (https://gwosc.org/data/)	~10 TB/year	DLHub / cuDNN	High	Detection of Gravity Wave
IRIS (https://iris.lmsal.com/)	~1GB/day	Cross-correlation	High	Coronal Loops
FAST (https://blinkverse.alkaidos.cn/)	7-10PB/year	HiFAST	High	FRB Detection

WISE (https://wise.ssl.berkeley.edu/)	15.6TB/14 mons	CNN	High	Mid-infrared Survey of the Entire Sky
SDSS (https://www.sdss.org/)	15TB	CNN	High	Star Catalog
LAMOST (https://www.lamost.org/)	7TB	Transformer	High	Spectra
('Models of AI used have a lot in fact, because different tasks need different models)				

Table 1: Datasets from various telescopes with AI models in relevant publications.

The Trend of using AI in Astronomical Data Processing

With the big data in astronomy and the fast development of AI, in recent years AI has become popular in astronomical data processing. This is because of 3 characteristics of astronomical data:

Huge Volume: Astronomy has come to the big data era, and the traditional manual measurement ways are losing supporters, even for the most conservative astronomers;

Difficult in the Definition of Useful Features: Color, shapes, texture, the traditional features have already been used to their physical limits, even with the most powerful telescopes. However, machine learning with data-driven methods can bypass the feature design step, as well as mine the unseen/ subtle features of the astronomical data;

Heavy Noises: Telescopes are always challenging mankind's technical limit. So the noises in the signals collected are often heavy especially focusing on the deep universe. Machine learning can enhance the results and help reaching the limits of mankind's techniques.

GWs are searched from neutron stars using Hough transform, Bayes' theorem, Hidden Markov Models, et al. [17].

A radio interferometer in HERA (Hydrogen Epoch of Reionization Array) can generate over 50 terabytes (TB) of data each night [18]. Deep learning methods are more and more applied in radio signal processing, especially for FRB (Fast Radio Burst) searching [19].

Infrared images from WISE are mined to do celestial bodies' classification [20]. Photometric data from SDSS are used to estimate redshifts of quasars [21].

A comprehensive review about AI applied in astronomy can be seen in [22].

Conclusions

In big data era, AI researchers are hunting for good application fields to show their muscles. Astronomy is not a popular 'rich' field but a real big data field. All the newest AI technologies could try to find a suitable application in astronomical data processing. It's not a new interdisciplinary topic but rarely to be noticed by AI circle before. This paper gives some classic datasets in astronomy, and calls on the attention from AI field. Astronomy has enough data for AI to try different algorithms.

Astronomical society has fell into AI field for quite some time. They even use ChatGPT to process astronomical papers to generate an omniscient 'sage' of astronomy. And in the famous astronomical journals like Nature Astronomy, ApJs, MNRAS, etc., there have been plenty of papers using AI algorithms.

It can be believed that, with the new telescopes' big data growing larger and larger, AI will play a key role in the future in astronomy.

References

1. Diebold FX (2021) What's the big idea. Big Data and its origins. Significance 18(1): 36-37.
2. Ajit Kembhavi (2018) Big Data in Astronomy and Beyond. Studies in Big Data 38: 59-66.
3. Mickaelian AM (2020) Big Data in Astronomy: Surveys, Catalogs, Databases and Archives. Communications of the Byurakan Astrophysical Observatory 67: 159-180.
4. National Astronomical Observatories (2023) Phase 3 Survey. Chinese Academy of Sciences.
5. Astronomy (2016) What is the ugriz magnitude system. Stack Exchange Network.
6. Payload (2012) WISE Flight System and Operations. Wise.
7. Copeland (2024) Artificial Intelligence. Britannica.
8. Bean R (2017) How Big Data Is Empowering AI and Machine Learning at Scale: Big Data is powerful on its own. So is artificial intelligence. What happens when the two are merged. MIT Sloan.

9. Imre B, Marek K (2017) Multimessenger Astronomy. Bristol, IOP Publishing, UK, pp: 1-28.

10. ZiJian W, Yu LJ, Fan Z (2022) Overview of the multimessenger astronomy on the moon. Scientia Sinica Physica, Mechanica & Astronomica 52(8): 289505.

11. Nutrino Observatory (2018) Data Releases. IceCube.

12. LIGO Caltech (2024) Laser Interferometer Gravitational-Wave Observatory.

13. National Aeronautics and Space Administration (2013) Interface Region Imaging Spectrograph (IRIS) fills a crucial gap in NASA's ability to advance Sun-Earth connection studies. NASA.

14. XMM Newton (2020) X-Ray Multi Mirror. MPE.

15. Ferrigno C, Savchenko V, Coleirob A, Panessac F, Bazzano A, et al. (2020) Multi-messenger astronomy with Integral 92: 101595.

16. European Space Agency (2024) Gaia. Science & Technology.

17. Wette K (2023) Searches for continuous gravitational waves from neutron stars: A twenty-year retrospective. Astroparticle Physics 153: 102880.

18. Paul LP, Peter KGW, Kolopanis M, Dillon S, Beardsley AP, et al. (2021) A Real Time Processing System For Big Data In Astronomy: Applications To Hera. Astronomy and Computing 36: 100489.

19. Mengyao L, Bo Q, Li LA (2024) RSSNet: A novel network model for FRBs search. IEEE International Conference on Computing and Complex Data.

20. Pan ZR, Bo Q, Liu CX, Luo AL, Jiang X, et al. (2024) Morphological Classification of Infrared Galaxies Based on WISE. Research in Astronomy and Astrophysics 24(4): 045020.

21. Yao L, Bo Q, Luo AL, Zhou J, Wu K, et al. (2023) Photometric redshift estimation of quasars with fused features from photometric data and images. Monthly Notices of the Royal Astronomical Society 523(4): 5799-5811.

22. Snigdha S, Sonali A, Chakraborty P, Singh KP (2022) Astronomical big data processing using machine learning: A comprehensive review, Experimental Astronomy 53(1): 1-43.