Application of Machine Learning to Radio Transients

Yang Xuan

2026-08-18

ChinaRxiv

Abstract

Full Text

doi: 10.15940/j.cnki.0001-5245.2026.04.012

Selected Abstract of Doctoral Dissertation

Application of Machine Learning to Radio Transients

Yang Xuan†

(Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023)

Radio transients are transient signals detected in radio observations. Common examples include pulsars, rotating radio transients (RRATs), and fast radio bursts (FRBs). Pulsars have been discovered for 57 years; they are stable periodic pulse sources in the Milky Way. RRATs likewise exhibit periodicity, but pulses are not detectable in every period. FRBs are millisecond-duration bright radio pulses discovered in 2007 and are considered to originate outside the Milky Way; they are divided into repeaters and non-repeaters. In recent years, as observational instruments have been continuously upgraded, the data volume has risen sharply, bringing new challenges to data storage and real-time searches. Focusing on the application of machine learning to the search for and classification of radio transients, this paper carries out research in three aspects: supervised-learning-based FRB searches, unsupervised-learning-based FRB classification, and anomalous-signal mining.

First, we introduce the residual neural network (Residual Neural Network, ResNet) into the FRB search pipeline. ResNet is used to perform image recognition on large volumes of archived observational data, identifying 81 candidate FRBs from massive images, thereby demonstrating the effectiveness of deep-learning models in the classification of radio transients. On this basis, for follow-up observational data of FRB 20200120E, ResNet was again introduced to assist candidate screening, which significantly shortened the time required for manual inspection and ultimately led to the successful detection of a bright pulse. These two studies show that supervised-learning methods can effectively replace traditional manual screening in large-data-volume scenarios, greatly improving the efficiency of FRB searches.

Second, using the Uniform Manifold Approximation and Projection (UMAP) dimensionality-reduction algorithm, we perform unsupervised clustering analysis on known FRB samples, mapping high-dimensional radio-observation features into a low-dimensional space to reveal the natural groupings of the samples. The results show that repeaters and non-repeaters exhibit a certain degree of separation in the UMAP-reduced space but are not completely mixed. Further

analysis finds that repeaters tend more strongly to display narrow-band features in the frequency direction, whereas this feature is not significant among non-repeaters. This result suggests that the two classes of FRBs may differ in their radiation mechanisms or local environments, providing new observational constraints for FRB classification and origin models.

Finally, we extend the UMAP algorithm to the direct search for anomalous signals in radio-observation archival data. Unlike classification work targeting known types of objects, this method assumes no physical model in advance, but performs anomaly detection solely according to the high-dimensional structure of the data itself. From large-scale archival data, a total of 202 anomalous signals are detected. Subsequent analysis confirms that some of them are radio-frequency interference (Radio Frequency Interference, RFI) such as satellites and lightning, while others show features suggestive of an astrophysical physical origin, such as particular dispersion measures and pulse widths; follow-up observations have been arranged to further confirm their physical nature. This work shows that unsupervised anomaly detection can serve as a model-independent data-mining tool, assisting in the discovery of new transients or rare physical events.

In summary, this work demonstrates that supervised residual neural networks and unsupervised UMAP dimensionality-reduction methods each have advantages and complement each other in the search, classification, and anomaly discovery of radio transients. With the gradual commissioning of next-generation observational facilities such as the Square Kilometre Array (SKA), machine-learning methods are expected to become a core tool for addressing the challenges of processing massive astronomical data.

Manuscript received 2026-06-02

† Awarded the doctoral degree on 2024-12-03; supervisor: Wu Xuefeng, researcher, Purple Mountain Observatory; xfwu@pmo.ac.cn

Application of Machine Learning in Radio Transients

YANG Xuan

(Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023)

Radio transients are short-lived signals detected in radio observations, commonly including pulsars, Rotating Radio Transients (RRATs), and Fast Radio Bursts (FRBs). Pulsars, discovered 57 years ago, are periodic and stable Galactic sources. RRATs, first identified in 2006, are periodic but their pulses are not detectable in every period. FRBs were discovered in 2007 as bright millisecond-duration pulses thought to be extragalactic, with both repeating

and non-repeating populations. In recent years, the continuous upgrading of radio facilities has led to a dramatic increase in data rates, posing new challenges for data storage and real-time transient searches. This thesis focuses on the application of machine learning to radio transient studies, covering supervised FRB detection, unsupervised FRB classification, and unsupervised anomaly detection.

First, a Residual Neural Network (ResNet) is introduced into the FRB search pipeline. Applied to a large archival dataset for image recognition, ResNet identifies 81 FRB candidates from a massive number of images, demonstrating the effectiveness of deep learning in transient classification. The same approach is further employed to process follow-up observations of FRB 20200120E, significantly accelerating candidate screening and leading to the discovery of a bright pulse. These results show that supervised learning can effectively replace manual inspection in large-data-volume scenarios, substantially improving FRB search efficiency.

Second, the Uniform Manifold Approximation and Projection (UMAP) algorithm is employed for unsupervised clustering of known FRBs, mapping high-dimensional observational features into a low-dimensional space to reveal natural groupings. The results show that repeating and non-repeating FRBs are partially separable in the UMAP-reduced space. Further analysis reveals that repeaters tend to exhibit narrow-band features along the frequency axis, a characteristic not prominent in non-repeaters. This finding suggests differences in radiation mechanisms or local environments between the two populations, providing new constraints for FRB classification and progenitor models.

Third, the UMAP algorithm is extended to the direct search for anomalous signals in radio observation archives. Unlike the classification work targeting known FRB types, this method imposes no prior physical model and performs anomaly detection based solely on the high-dimensional structure of the data. A total of 202 anomalous candidates are identified. Subsequent analysis confirms that some originate from radio frequency interference (RFI), such as satellites and lightning, while others exhibit characteristics suggestive of astrophysical origin, including specific dispersion measures and pulse widths. Follow-up observations have been arranged for these candidates. This work demonstrates that unsupervised anomaly detection can serve as a model-agnostic data mining tool for discovering new types of transients or rare physical events.

In summary, this thesis demonstrates that supervised methods such as ResNet and unsupervised methods such as UMAP each offer distinct and complementary advantages in the search, classification, and anomaly detection of radio transients. As next-generation facilities such as the Square Kilometre Array (SKA) come online, machine learning is poised to become a core tool for addressing the massive data processing challenges in astronomy.

Note: Figure translations are in progress. See original paper for figures.