Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 166 (2020) 534–538

Procedia Computer Science

www.elsevier.com/locate/procedia

3rd International Conference on Mechatronics and Intelligent Robotics (ICMIR-2019)

A Review of Research on Pulsar Candidate Recognition Based on Machine Learning

Cheng Jun Zhang, Zhen Hong Shang*[1],Wan Min Chen,Liu Xie and Xiang Hua Miao

Faculty of Information Engineering and Automation
Kunming University of Science and Technology, Kunming 650500, China

Abstract.

The pulsar is a highly magnetized rotating neutron star that provides the first indirect evidence for the existence of gravitational waves and also provides the possibility to reveal extreme phenomena in neutron star astrophysics. Therefore, the identification of pulsars in the universe is a prerequisite for the study of pulsars and gravitational waves. At present, a large number of pulsar searches have produced millions of pulsar candidates. In the face of these large-scale data, if only relying on manual visual classification by experts in related fields, it will be a huge project. Since the emergence of machine learning, its theory and technology have become increasingly mature, and has been successfully applied to astronomical research fields such as pulsar candidate screening. This paper introduces the related machine learning theory of pulsar candidate recognition firstly, and then reviews the research status of pulsar candidate recognition based on machine learning in recent years. Finally, we discuss and prospect the identification of pulsars in the future.

Keywords: Pulsar, Machine learning, Candidate identification.

1. Introduction

Since the first pulsar was discovered by Cambridge University's Jocelyn Bell and Antony Hewish in 1967 [1], researchers have gone through a long journey of pulsar search in related fields.

Searching for pulsar signals in radio data is a process of finding periodic pulsar signals with dispersion characteristics. Through the DM search, Fourier transform and periodic signal search, people get the candidate of

[1] Corresponding Author. Tel.+(86) 15912136296
*E-mail: shangzhenhong@qq.com

10.1016/j.procs.2020.02.050

real pulsar signal. Single-pulse searches can produce millions of pulsar candidates, most of which are contaminated by radio frequency interference (RFI) or other noise, making it difficult to use simple metrics such as signal-to-noise ratio (SNR). Pick the pulsar correctly from the candidate. Faced with millions of candidates, it is impractical for experts in the field to identify pulsars by looking at the characteristics of the pulsar signal. Although the actual pulsar signal can be screened out from the candidate by the graphical selection tools [2], [3], the use of these tools has certain limitations and is prone to errors. Therefore, in order to more effectively and accurately select true pulsars from the candidate, human intervention in the pulsar search process should be minimized. In recent years, machine learning methods have not only been widely practiced in the computer field, but have also been gradually applied to the search process of radio pulsars in the astronomical field. This paper reviews the application of machine learning methods to the identification of pulsars.

2. Machine Learning Principle of Pulsar Candidate Recognition

As an important direction of artificial intelligence technology, machine learning has been widely used in data mining, natural language processing, computer vision and other fields [4]. The machine learning method is a method in which a computer learns rules from training data, generates a model, and then uses this model to predict new data. The principle of pulsar candidate recognition based on machine learning is shown in Fig. 1.

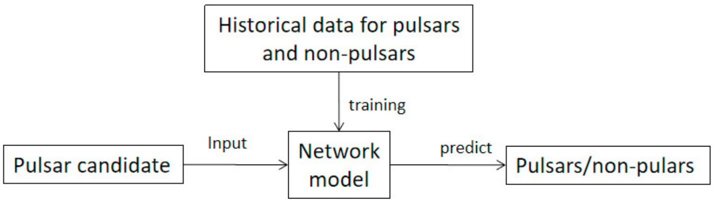

Figure. 1. Schematic diagram of pulsar candidate recognition based on machine learning

Since machine learning has been widely used in the computer field, researchers have applied various machine learning methods to the search project of radio pulsars. At present, the machine learning algorithms applied to pulsar candidate recognition mainly include Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree, Ensemble Learning, and other improved algorithms.

2.1 Artificial Neural Network Based Pulsar Candidate Identification Method

Artificial Neural Networks (ANN), abbreviated as neural networks, is a mathematical model based on the basic principles of neural networks in biology, abstracting human brain neural networks from the perspective of information processing [5]. At present, there are two main applications for identifying pulsar candidates based on artificial neural networks.

Eatough, et al. (2010) first applied ANN to identify pulsar candidates [6] and used this method to find a new pulsar J1926+0739 in the PMPS (Parkes Multi-beam Survey) [7] data. With the training of the ANN, the best ANN model can be obtained before the error rate of the verification set begins to increase, that is, before 120 cycles. In the 8:8:2 ANN, the analysis of 2.5 million pulsar candidates took a total of 50 CPU time, and the results showed that the eligible candidates accounted for 0.5%, about 13,000 candidates, and 92 of them. % of the pulsars have been found. In the 12:12:2 ANN, the pulsar recognition rate increased to 93% compared to the 8:8:2 ANN, but the calculation time doubled.

Bates, et al. (2012) used the study of Eatough et al. (2010) to identify the pulsar candidate again using ANN [8] and used this method to find 75 in the HTRU (High Time Resolution Universe pulsar survey) [9] data. Pulsar. The ANN uses a 22:22:2 network structure with input parameters using 22 candidate parameters such as pulse period, DM, signal-to-noise ratio, and pulse width. The ANN training set and validation set use 270 data in the HTRU data (70 pulsar candidates and 200 non-pulsar candidates). In the ANN training process, the ownership weight parameters are randomly assigned, and the best ANN training can be achieved when the verification error reaches the minimum value. The ANN test identifies a test set containing 510 pulsar candidates, which correctly identifies

85% of pulsars. The ANN is capable of detecting pulsars in all cycle periods, but is clearly not good at identifying wide pulses (ie, duty cycles greater than or equal to 20%), pulsars with low signal-to-noise ratio and short period.

2.2 Support Vector Machine Based Pulsar Candidate Identification Method

Support Vector Machine (SVM) is a supervised machine learning model that analyzes data, identifies patterns, and uses classification and regression analysis [13]. SVM can be used for linear classification, and can also use the kernel function to map the input space to the high-dimensional feature space for nonlinear classification. Support vector machines show many unique advantages in solving small sample, nonlinear and high dimensional pattern recognition.

Devine et al. (2016) used a support vector machine algorithm to classify discrete pulsars [11]. During the training process, the SVM uses an iterative training algorithm to minimize the error function. When the error function reaches a minimum, an optimal hyperplane is obtained to achieve the purpose of discrete pulsar group classification.

2.3 Decision Tree Based Pulsar Candidate Identification Method

The Decision Tree is a basic classification and regression method and belongs to the supervised machine learning algorithm [14]. The decision tree is in a tree structure, in which each internal node represents a feature or attribute, and each leaf node represents a category, and its main advantage is that it is readable and fast.

Lyon et al. (2015) proposed an adaptive online classification algorithm [15], Gaussian Hellinger Very Fast Decision Tree (GH-VFDT) [16], and applied this method to LOTAAS. Four new pulsars were found in the data (LOFAR Tied-Array All-Sky Survey). GH-VFDT is based on the algorithm of the Fast Decision Tree (VFDT) developed by Hulten et al., which can maximize the classification of “unbalanced” candidate data (poor pulsar data and non-pulse data). performance. GH-VFDT uses eight new features to characterize typical pulsar candidates, using three separate data sets, HTRU1, HTRU2, and LOTAAS1, to test the performance of the recognition. Lyon et al. use the GH-VFDT method to process millions of candidates in a matter of seconds (1 million per 15 seconds) and achieve pulsar recognition rates of over 90%.

2.4 Integrated Learning Based Pulsar Candidate Identification Method

Ensemble Learning is a machine learning method that combines multiple learners and then integrates these learners with a certain rule to obtain better learning performance than a single learner, mainly including Bagging and Boosting. Method [17]. The structure of integrated learning is shown in Fig. 2.

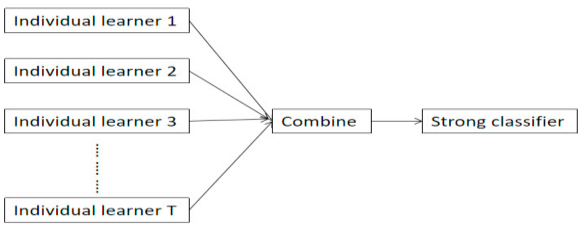

Figure. 2. Integrated learning

Devine et al. (2016) proposed a new method to identify and classify discrete bursts, in which a random forest machine learning algorithm was used in the classification process [11], which has obvious advantages over other machine learning algorithms.

2.5 Pulsar Candidate Recognition Method Based on Support Vector Machine and Neural Network

Zhu et al. (2013) used a deep neural network image pattern recognition method, PICS (Pulsar Image-based Classification System) [22] to automatically identify whether a pulsar candidate image is a real pulsar image, and use this method in PALFA. (Pulsar Arecibo L-band Feed Array) [23] Six new pulsars were found in the data.

3. Conclusion

Throughout the above several machine learning methods, the real pulsar can be successfully identified to a certain extent, and the recognition speed is improved, which has certain reference significance for the future pulsar candidate identification work, but the pulsar identification still faces certain The challenge. Faced with extremely high-sensitivity radio telescopes that capture billions or more of pulsar candidates, there are some exotic types of pulsars in the machine learning universe. For these possible types of candidate features, use machines. Learning methods will also be powerless. Future work can try to improve the pulsar search process, build a search database for machine learning training sets, and classify candidates when constructing machine learning training sets, such as MSPs, long periods. Pulsar, noise signal, and RFI signal.

Inspired by the review, future work can try to use the method based on Bagging and neural network to automatically identify the pulsar candidate. On the other hand, in recent years, artificial intelligence has developed rapidly. Among them, deep learning technology covers all fields of artificial intelligence. Future work can consider applying deep learning to pulsar star selection.

Machine learning is the core of artificial intelligence, and it is the fundamental way to make computers intelligent. Its application spans all fields of artificial intelligence. The introduction of machine learning into the search project of radio pulsars has promoted the development of pulsar search and achieved remarkable results. This paper reviews the method of pulsar candidate recognition based on machine learning, discusses the problems that need to be solved, and proposes other identification methods.

4. References

1. Hewish A, Bell S J, Pilkington J D H, Scott P F, Collins R A. Observation of a Rapidly Pulsating Radio Source[J]. Nature, 217 (5130):709-713, 1968.
2. Edwards R T, Bailes M, van Straten W, et al. The Swinburne intermediate - latitude pulsar survey[J]. MNRAS, 326(1): 358-374, 2001.
3. Burgay M, Rea N, Israel G L, et al. Search for radio pulsations in four Anomalous X - ray Pulsars and discovery of two new pulsars[J]. MNRAS, 372(1): 410-416, 2006.
4. PETER F. Machine Learning: The Art and Science of Algorithms that Make Sense of Data[M]. Cambridge: Cambridge University Press, 2012.
5. Bishop C.M., Neural Networks for Pattern Recognition[M]. Oxford University Press,1995.
6. Eatough R P, Molkenthin N, Kramer M, et al. Selection of radio pulsar candidates using artificial neural networks[J]. MNRAS, 407(4):2443-2450, 2010.
7. Manchester R N, Lyne A G, D Amico N, et al. The Parkes Southern Pulsar Survey I:Observing and data analysis and initial results. MNRAS, 279(4):1235-1250, 1996.
8. [7] Manchester R N, Lyne A G, D Amico N, et al. The Parkes Southern Pulsar Survey I:Observing and data analysis and initial results. MNRAS, 279(4):1235-1250, 1996.
9. Keith M J, Jameson A, van Straten W, et al. The High Time Resolution Universe Pulsar Survey I: System configuration and initial discoveries[J]. MNRAS, 409(2): 619-627, 2010.
10. Morello V, Barr E D, Bailes M, et al. SPINN: a straightforward machine learning solution to the pulsar candidate selection problem[J]. MNRAS, 443(2):1651-1662, 2014.
11. Devine et al. Detection of dispersed radio pulses: a machine learning approach to candidate identification and classification. MNRAS, 459(2):1519-1532, 2016.
12. Bethapudi et al. Separation of pulsar signals from noise using supervised machine learning algorithms. Astronomy and computing,23(2018):15-26,2018.
13. Schoelkopf B, Burges C J C, Smola A J, Advances in Kernel Methods - Support Vector Learning. Cambridge: MIT Press, 1999.
14. Safavian S R, Landgrebe D. A survey of decision tree classifier methodology[J]. IEEE Transactions on Systems Man & Cybernetics, 21(3):660-674, 2002.
15. Lyon R J, Stapper B W, Cooper S, et al. Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach[J]. MNRAS, (1104-1125)2016.
16. Haykin S. Neural Networks and Learning Machines, 3rd edition[M], Upper Saddle River: Prentice Hall, 2009.
17. Thomas G. Dietterich. Ensemble Methods in Machine Learning[J]. International Workshop on Multiple Classifier Systems, 1857(1):1-15, 2000.
18. Mayr A, Binder H, Gefeller O, Schmid M. The Evolution of Boosting Algorithms - From Machine Learning to Statistical Modelling[J]. Methods of Information in Medicine, 53(6):419-427, 2014.
19. Friedman J H. Greedy function approximation: A gradient boosting machine[J]. Annals of Statistics, 29(5), 1189-1232, 2001.
20. Tramacere A, Paraficz D, Dubath P, Kneib J P, Courbin F. ASTErIsM:application of topometric clustering algorithms in automatic galaxy detection and classification[J]. MNRAS, 463(3):2939-2957, 2016.
21. Chen T, He T. Higgs boson discovery with boosted trees [C] // Proc of International Conference on High-Energy Physics and Machine Learning. 2015.

22. Zhu W W, Berndsen A, Madsen E C, et al. Searching for Pulsars Using Image Pattern Recognition[J]. ApJ, 781(2): 117, 2014.
23. Cordes J M, Frelre P C C, Lorimer D R, et al. Arecibo Pulsar Survey Using ALFA. I. Survey Strategy and First Discoveries[J]. ApJ, 637(1): 446-455, 2006.
24. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324,1998.