Pulsar Astrophysics the Next Fifty Years
Proceedings IAU Symposium No. 337, 2017
P. Weltevrede, B.B.P. Perera, L.L. Preston
& S. Sanidas, eds.

doi:10.1017/S1743921317009000

Machine Learning for Pulsar Detection

Rebecca McFadden[1], Aris Karastergiou[2] and Stephen Roberts[1]

[1]Department of Engineering Science [2]Department of Physics
University of Oxford, Oxford, UK
email: rebecca.mcfadden@eng.ox.ac.uk

Abstract. The next generation of radio telescopes will have unprecedented sensitivity and time-resolution offering exciting new capabilities in time-domain science. However, this will result in very large numbers of pulsar and transient event candidates and the associated data rates will be technically challenging in terms of data storage and signal processing. Automated detection and classification techniques are therefore required and must be optimized to allow high-throughput data processing in real time. In this paper we provide a summary of the emerging machine learning techniques being applied to this problem.

Keywords. pulsars: general, methods: data analysis, methods: statistical

1. Introduction

Pulsars emit radio waves; the features of which are indicative of their emission and propagation mechanisms. These features can be exploited to aid in signal detection and classification, however, their complex signal properties also make pulsar signal processing computationally intensive. Computational requirements will increase with the next generation of radio telescopes. In particular, the Square Kilometre Array, SKA (https://www.skatelescope.org), will produce data at a rate of Gbits/s for pulsar and transient searches. Off-line signal processing and event detection is impractical due to immense data rates and therefore extreme optimization is required to automate the computationally intensive process of knowledge-extraction and classification, allowing high-throughput data to be processed for candidate detection in real time.

2. Related Works on Automated Selection Methods

Automated detection methods exploit the signal feature space to identify data representations which maximize separation between noise and candidate events. Features can be extracted from various stages of the signal processing pipeline and several authors (e.g. Morello *et al.* 2014; Lyon *et al.* 2016) provide critiques of the features used in existing machine learning solutions. In particular, parameters derived from the dispersion measure search stage and the final integrated pulse profile are already commonly used in classification algorithms. Table 1 shows the evolution of automated candidate selection techniques over time with the most recent publications representing the state of the art in machine learning applications. Early automated methods for candidate selection have included graphical selection tools (Faulkner *et al.* 2004; Keith *et al.* 2009), and scoring algorithms (Lee *et al.* 2013). While these methods provide a level of bulk processing they still typically also require a manual-processing stage.

Supervised machine learning classifiers have also been applied to the candidate selection problem via artificial neural networks, ANNs, (Eatough *et al.* 2010; Bates *et al.* 2012; Morello *et al.* 2014), and pattern recognition algorithms (Zhu *et al.* 2014), while an unsupervised tree-based classifier has also been implemented (Lyon *et al.* 2016) and found particularly suitable for survey-independent applications and continuous data-stream

Table 1. Automated Detection Methods for Pulsar Candidate Selection.

Publication	Method	Details
Faulkner 2004	Graphical Selection Tool	128 new pulsars
Keith 2009	Graphical Selection Tool and Scoring Algorithm	28 new pulsars
Eatough 2010	ANN	8 to 12 features, 1 new pulsar
Bates 2012	ANN	up to 22 features
Lee 2013	Scoring Algorithm	6 'quality factors' 47 new pulsars
Morello 2014	ANN	Feature-based
Zhu 2014	ANN, CNN and SVM	Image-based Algorithms combined in Deep Neural Network
Lyon 2016	Hellinger Decision Tree	Feature-based
Devine 2016	ANN, SVM, Direct Rule Learner, Standard Tree Learner, Hybrid Rule-and-Tree Learner and Ensemble Tree Learner	Algorithms combined optimally for binary and multi-class classification
Bethapudi 2017	ANN, Adaboost, GBC and XGBoost	Comparative Study of 4 algorithms

processing. More recently Devine *et al.* (2016) have looked at optimally combining six different algorithms for single dispersed pulse searching and Bethapudi *et al.* (2017) noted that most previous machine learning attempts in pulsar literature have involved a variant of ANN so they explored various metrics to compare three different techniques which they conclude perform better than previous ANN implementations. Historically, the success of these techniques has been measured by the number of new pulsars discovered but, since the High Time Resolution Universe survey data has been made public (Morello *et al.* 2014), algorithms are being tested on the same dataset and detection efficiency benchmarks are becoming more complex.

3. Concluding Remarks

Machine learning methods have reduced the amount of processing time required for pulsar discoveries, however, most are only applied at the candidate selection stage. This leaves scope for us to re-examine the signal processing chain and identify areas still to be optimised. We are investigating non-parametric data exploration techniques to examine interrelations and underlying structure in the data, and to identify signal features which will be effective in classification algorithms. We will also explore probabilistic methods as a promising alternative to previous ANN implementations. These methods may provide improved performance by characterising model bias and uncertainty.

References

Bates, S., et al. 2012, *MNRAS*, 427, 1052
Bethapudi, S. & Desai, S. 2017, *ArXiv e-prints*
Devine, T., Goseva-Popstojanova, K., & McLaughlin, M. 2016, *MNRAS*, 459, 1519
Eatough, R., et al. 2010, *MNRAS*, 407, 2443
Faulkner, A., et al. 2004, *MNRAS*, 355, 147
Keith, M., et al. 2009, *MNRAS*, 395, 837
Lee, K., et al. 2013, *MNRAS*, 433, 688
Lyon, R., et al. 2016, *MNRAS*, 459
Morello, V., et al. 2014, *MNRAS*, 443, 1651
Zhu, W., et al. 2014, *ApJ*, 781, 117