Machine Learning to Classify Pulsar Candidates

Anna Stephens, Otho Ulrich, Mariia Kravtsova

Introduction

Pulsars are neutron stars that rotate with great regularity, seen from earth as periodic radio signals. Some millisecond pulsars are so precise they rival the best atomic clocks [5], and there are other applications. The regular period of each pulsar provides a unique signature, making them excellent guide-posts in universe mapping, and a tool for informing extraterrestrial intelligence how to locate Earth[2], but mainly they're interesting for what they can tell us about fundamental physics. Researchers have used the spectral and dynamic properties of pulsars to create catalogs of many thousands of these objects. Modern astronomical devices use Fourier transformations to identify candidates which are flagged for later classification by researchers. [1] Machine learning is recognized as a potential herald in this field, presenting the possibility of very accurate classification of pulsars with minimal human supervision. [1] We explore the feasibility of using machine learning classification techniques to classify pulsar candidates.

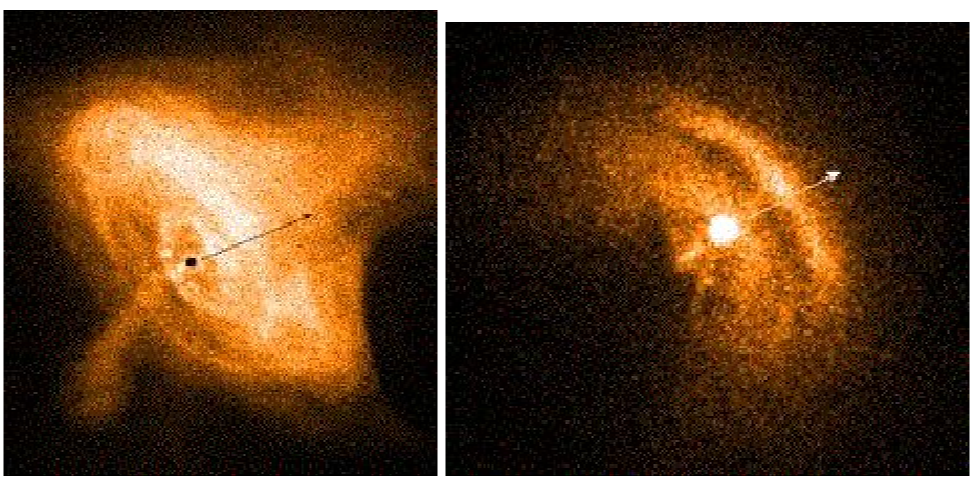

Figure 1. X-ray telescope observations from Chandra of two well-known pulsars surrounded by their accretion disks and showing signs of outward wind. Left: Crab Pulsar. Right: Vega Pulsar. [3]

The process of identifying a pulsar takes two steps. First, a pulsar candidate is identified by its fourier transform within the telescope or other pre-processing device. Since a sky survey doesn't look at the same place for long, a single pulse from the pulsar is observed at most. Figure 2 shows a typical emission profile of a pulsar candidate. The profile is saved for later analysis by a scientist. The candidate can then be identified as a pulsar or not by statistical measures of the emission profile or by watching the same position until more pulses are observed.

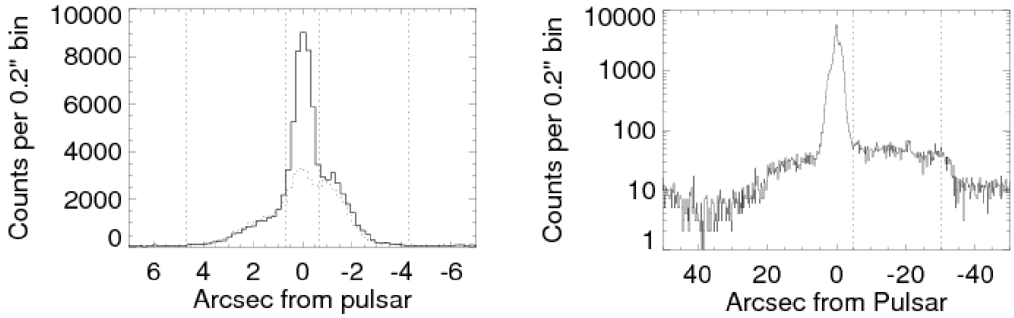

Figure 2. Emission profile of a typical pulsar candidate. Left: linear received signal strength. Right: logarithmic received signal strength, which more readily demonstrates the relative background. Statistical measures like kurtosis and skewness may be sufficient to identify a pulsar properly. [4]

Data

The High Time Resolution Universe Pulsar Survey is a sky survey that identifies pulsar candidates using the Parkes Radio Telescope. [6] We use a dataset from that survey containing 17,898 pulsar candidates already classified by inspection, of which 16,259 are not pulsars and 1,639 are pulsars. The included features are:

X1. Mean of the integrated profile
X2. Standard deviation of the integrated profile
X3. Excess kurtosis of the integrated profile
X4. Skewness of the integrated profile
X5. Mean of the dispersion measure-signal-to-noise-ratio curve
X6. Standard deviation of the DM-SNR curve
X7. Excess kurtosis of the DM-SNR curve
X8. Skewness of the DM-SNR curve
X9. Pulsar binary class

These are statistical measures of the pulsar emission integrated profile and its dispersion measure relationship to the signal-to-noise ratio. The integrated profile is received signal strength across all monitored wavelengths. The dispersion measure is a measure of the spread of the pulse, and its relationship with the signal-to-noise ratio is of particular interest to astronomers because of dispersion spread over spatial distances. [4] The excess kurtosis is a measure of how long the tail of the pulse is, and the skewness is a measure of how large one tail of the pulse is relative to the other.

We visualized the data by plotting all relationships between the features. Many relationships are correlated because they are related statistical quantities, such as skewness and kurtosis. No one feature was immediately apparent as a superior predictor of the pulsar class, but inspection did indicate that the pulsar class had some meaningful distribution across some features, especially the excess kurtosis in the integrated profile and the dispersion measure-signal-to-noise ratio. This told us that we may be able to find a strongly predictive model.

We also plotted a correlation matrix across all features, seen in Figure 3. Here we saw the strongest correlation with pulsar class to be the integrated profile excess kurtosis. The skewness of the integrated profile was also strongly correlated, but we recall that the kurtosis and skewness are expected to correlate. This indicates the skewness feature may be a reducible variable.

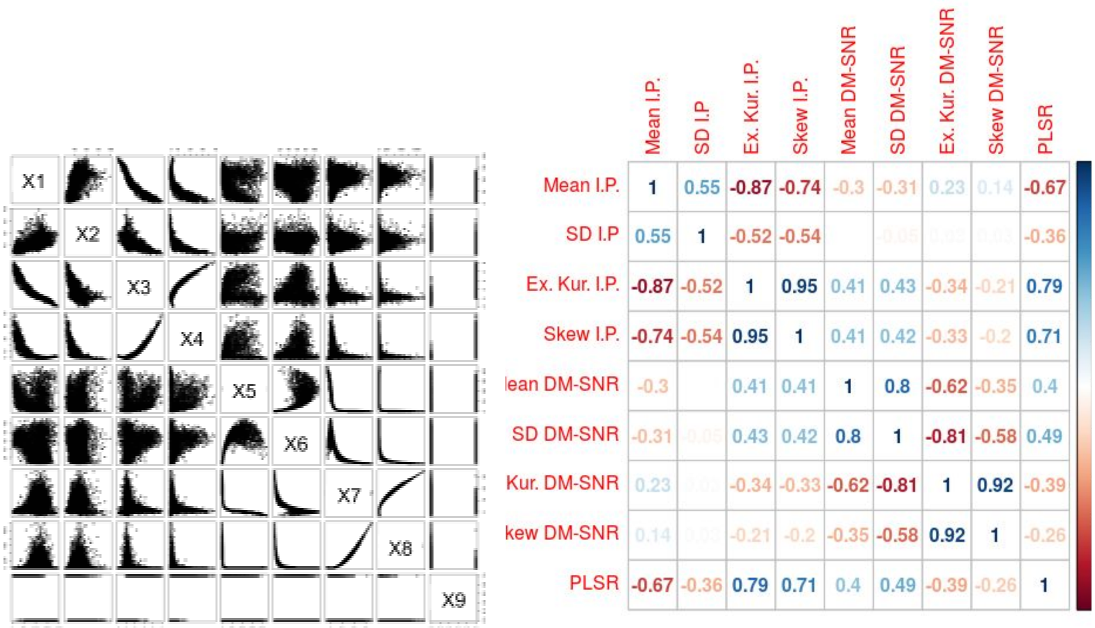

Figure 3. Left: all feature relationships plotted. Many correlations are due to the relationships between statistical quantities. The plots hint at a relationship between pulsar class and excess kurtosis in both statistics. Right: the correlation matrix for this dataset. PLSR is highly correlated with kurtosis and skewness, but skewness and kurtosis are also strongly correlated.

Modeling and Results

Method 1: Trees

The first method that we selected for this project is decision trees. Our data is linearly separated into two defined classes. Comparing the mean, standard deviation, kurtosis, and skewness of integrated profile versus the DM-SNR curve. This method lets us avoid worrying about outliers and analyse the information gain with some variance trade off. Since decision trees are highly interoperable machine learning algorithms, we chose to test this model first.

The tree method was trained using 80% of the dataset, leaving 20% as test data. We generated a tree five times using randomly selected training and test sets. All of the tree models first split the data using the “Excess Kurtosis of the Integrated Profile”. That first split showed a significant drop in error rate. There were a few smaller separations made on the “Standard deviation of the DM-SNR curve” and “Skewness of the integrated profile”. Those branches, however, were relatively insignificant and varied between the trained models. All of the trees had five leaf nodes without pruning. See figure 4.

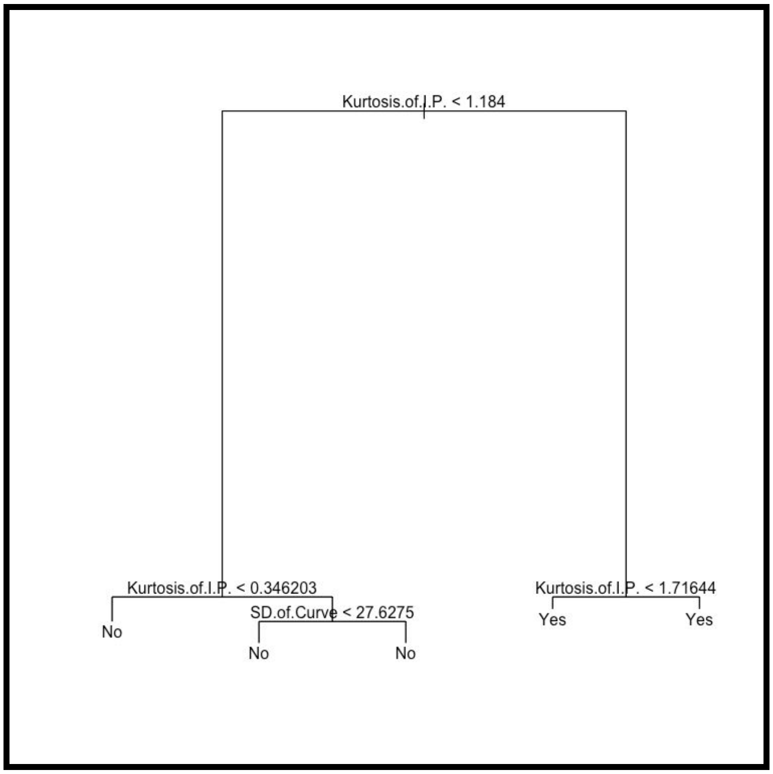

Figure 4. Plot of a decision tree trained on one of our training sets. Excess kurtosis of the integrated profile may be the most important predictor of the 8 features, by far. We see the effect of node purity here with leaves that don’t affect the classification error.

Trial #	Actual/Predicted				% Correct Classification
	No/No	No/Yes	Yes/No	Yes/Yes	
1	3212	35	53	279	97.54%
2	3209	37	55	278	97.43%
3	3221	22	66	270	97.54%
4	3223	24	57	275	97.74%
5	3234	37	45	263	97.71%
Average					97.59%

Table 1. Results for five separate trials using the tree method. The correct classification rate ranged from 97.43% to 97.74% and averaged 97.59%.

Method 2: Support Vector Machines (SVM)

Support vector machines work great for classification problems, so this was our second choice. SVM does not allow us to easily interpret our data as the decision tree. However, we relied on the classification score, and by using this method our goal was to reduce variance, and perhaps increase accuracy.

The SVM model was trained using 80% of the data for training, leaving 20% as test data. We tuned the model with randomly-chosen training and test sets to determine the ideal cost and gamma. The results for this first phase can be seen in Table 2. These results led us to choose a cost of 100 and gamma of 0.1.

Trial #	Tune Results		Actual/Predicted				% Correct Classification
	Cost	Gamma	No/No	No/Yes	Yes/No	Yes/Yes	
1	1	0.5	3225	18	63	273	97.74%
2	10	0.5	3248	23	43	265	98.16%
3	100	0.1	3216	20	60	283	97.76%
4	100	0.1	3238	24	42	275	98.16%
5	100	0.1	3242	18	52	267	98.04%

Table 2. Parameters and results of running the Support Vector Machine Algorithm five times.

In the next phase we wanted to see whether using feature subsets would influence the accuracy of the model. We trained SVM models on various subsets using the cost and gamma determined in the original tuning. Notable ones are listed in Table 3.

Parameters	Average	Min	Max	# of Trials
All	98.00%	97.76%	98.32%	10
X1-X4	97.96%	97.65%	98.24%	10
X3, X4, X6	97.69%	97.63%	97.79%	3
X3	97.53%	97.29%	97.71%	3
X5-X8	95.44%	95.42%	95.45%	3

Table 3. Parameters X1-4 refer to the four statistical measures of the integrated profile: mean; standard deviation; excess kurtosis; and skewness. Likewise X5-8 refer to the features of the dispersion measure-signal-to-noise ratio curve.

Conclusion

The decision tree classifier consistently provided an accuracy of approximately 97.6%. This is far better than we expected, and we were able to interpret from the model that excess kurtosis of the integrated profile is the most significant feature for predicting pulsar class, as predicted from our initial inspection of the data.

For the support vector machine model, the best accuracy was obtained by retaining all parameters. A model using X1-X4, the integrated profile features, was only 0.04% less accurate, compared to a loss of >0.3% for other parameter set choices. In fact, the accuracy of the X1-X4 model surpassed that of the model with all of the parameters in some trials. There may be no significant difference between these choices of feature sets. Overall, the SVM model performed slightly better than the decision tree model. Since SVM is less prone to overfitting than the decision tree, this suggests the other features of the integrated profile may have predictive significance.

Discussion

The problem that astrophysicists would like to solve is to have a machine algorithm analyse the data from telescopes and make accurate estimation if it is a pulsar and predict its distance. However, from the techniques we have learned in class we did not have enough tools to satisfy those conditions.

We have done some personal research, and can suggest that a neural network will be able to solve such problem. Neural Network is an adaptive algorithm therefore it has an ability to greatly improve the 98% accuracy, by using the weights to save results, and backpropagation for continuous improvement. This might also become a real time solution which will save a lot of waiting time.

That being said, a 97-98% prediction accuracy is quite good, and a strong improvement over current fourier transform methods. This technique could certainly be considered as a way of trimming the pulsar candidates before inspection by a scientist.

References

[1] R. J. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, J. D. Knowles, Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach. MNRAS, 2016.
[2] Marissa Cevallos, Science News,"HOW TO USE A PULSAR TO FIND STARBUCKS"[2], Discovery News,Wed Nov 24, 2010 10:21 AM ET.

http://www.seeker.com/how-to-use-a-pulsar-to-find-starbucks-discovery-news-1766499711.html
[3] P. Kaaret et al. "Chandra Observations of the Young Pulsar PSR B0540-69". In: apj 546 (Jan.
2001), pp. 1159–1167. doi: 10.1086/318287. eprint: astro-ph/0008388.
[4] M. J. Keith et al. "The High Time Resolution Universe Pulsar Survey - I. System configuration
and initial discoveries". In: mnras 409 (Dec. 2010), pp. 619–627. doi: 10 . 1111 / j . 1365 -
2966.2010.17325.x. arXiv: 1006.5744 [astro-ph.HE].
[5] D. N. Matsakis, J. H. Taylor, and T. M. Eubanks. "A statistic for describing pulsar and clock
stabilities." In: aap 326 (Oct. 1997), pp. 924–928.
[6] R. P. Mignani, P. A. Caraveo, and A. de Luca. "HST As a Powerful Astrometric Tool for
Pulsar Astronomy". In: A Decade of HST Science : Poster
[7] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, "An Introduction to Statistical Learning: with Applications in R"