Uncovering the Nature of Fermi LAT Unassociated Gamma-ray Sources

Fan Kwok Lung

Major:Decision analytics and Physics

Name:Fan Kwok Lung
University No.:3035272731
Student's Major:Decision analytics and Physics
Poster No.:A2

Research Colloquium for Science UG Students 2017-18

Introduction

Fermi Large Area Telescope(Fermi LAT) is the latest generation of gamma-ray telescope and it was launched in 2008 and it has operated almost ten years[1].Recently, Fermi LAT team had released the Preliminary sources catalog using eight years of data(FL8Y). It includes 5524 sources and 2130 sources are unassociated sources which mean we do not know the nature of the sources. The majority of these gamma-ray sources will be active galactic nucleus(AGN) or pulsar(PSR) .Our aim is to uncover the nature of those unassociated gamma-ray source with machine learning.

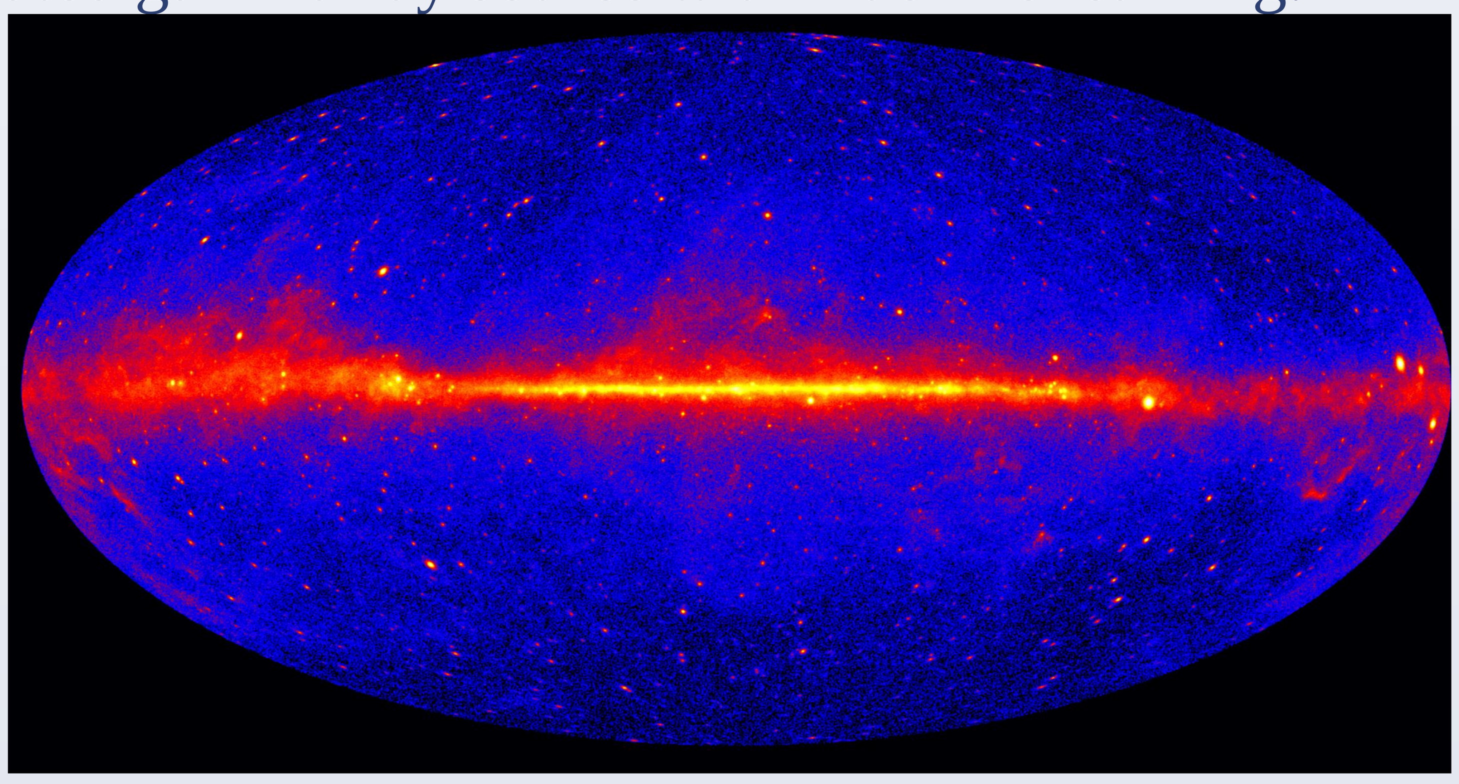


Figure 1. Fermi LAT 5 Years all-sky map(NASA)

Methodology

We apply four different machine learning algorithms which is logistic regression, support vector machine , random forest and neural network. We normalized the data to avoid the effect caused by skew distribution and large value different between column. We also apply class weight method and cross-validation to obtain the optimal cut-off probability to tackle the imbalance class problem in our dataset(Number of AGN is more than PSR by an order of magnitude)[2].

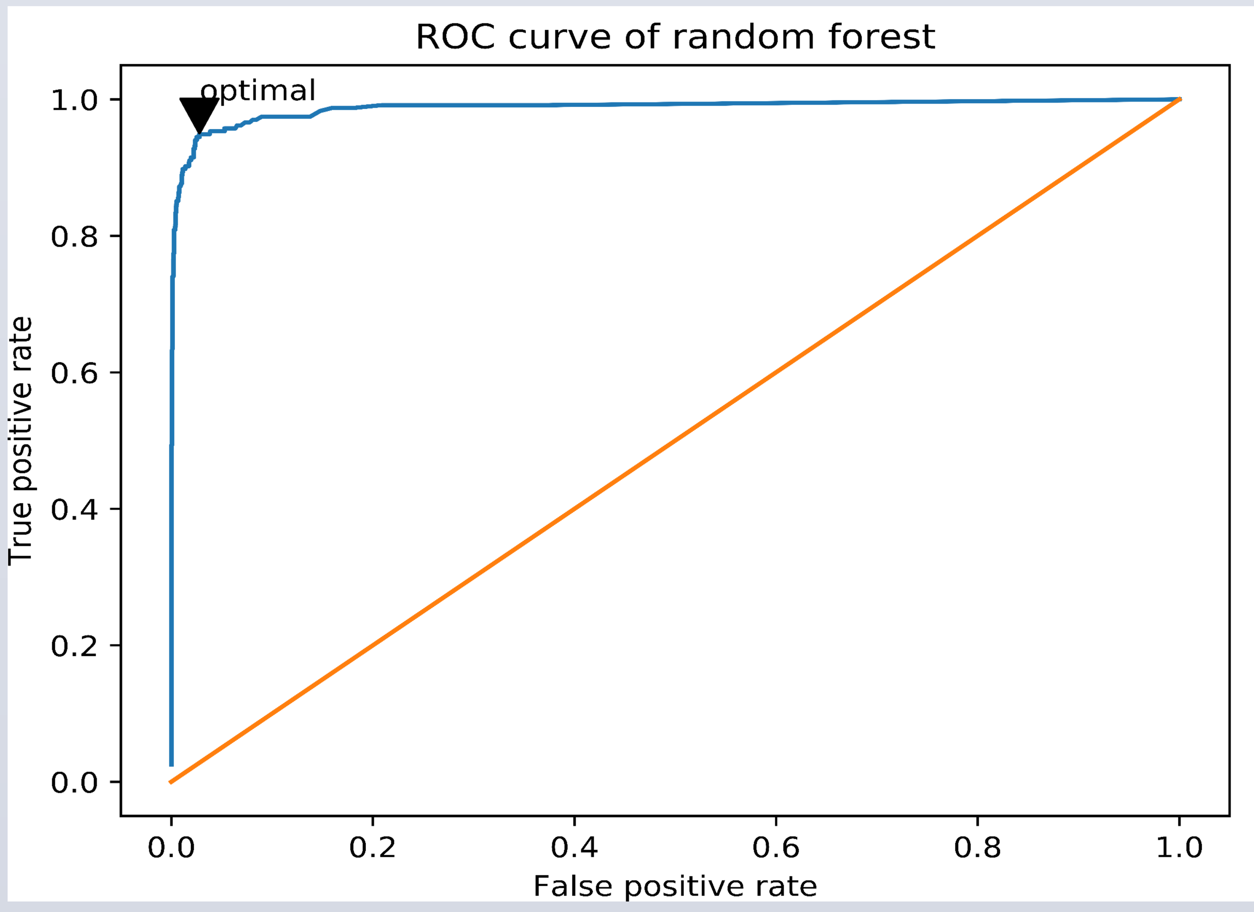


Figure 2. Receiver operating characteristic(ROC) curve of our random forest model. The optimal cut-off is the maximum of (true positive – false postive)

In FL8Y catalog, information related to the source is provided. Variables like energy flux and energy sprectrum etc. can potential predict whether the source is AGN or Pulsar. We use metrics like f-score and AUC score to find out the optimal features.

Beside the feature which provided in the catalog, we also generated new feature for better performance. Base on the physical property of pulsar, we know it is a very steady source[3].We use aperture photometry technique to generate light curve of the sources and then compute the variability measure. Variability together with the variables selected from the catalog, it will be the features fed into the models.

Result

We split the known pulsars and AGNs data set to 70%/30% and use it as training set and test set. We use PSR accuracy(TPR) as the main metric to evaluate the performance of a model.

RF_Overall	RF_PSR	RF_AGN
0.962305	0.948426	0.963381
LR_Overall	LR_PSR	LR_AGN
0.959709	0.908122	0.963684
ANN_Overall	ANN_PSR	ANN_AGN
0.96054	0.936045	0.962478
SVM_Overall	SVM_PSR	SVM_AGN
0.957996	0.90451	0.962214

Table 1. The summary of the performance of four different model after fine-tuning the hyper-parameter, Notice that the result is the average result of ten different random seed splitting

Random forest and neural network is the two best algorithm for our problem, having ~0.94 use PSR accuracy. However, logistic regression and support vector machine still have a satisfactory performance even we used linear kernel. This mean the data is fairly linear.

We try to find out the underlying pattern in our model after the evaluation of our model. We select the a few most important features using mean decrease in impurity and then project the data into the 2D plane.

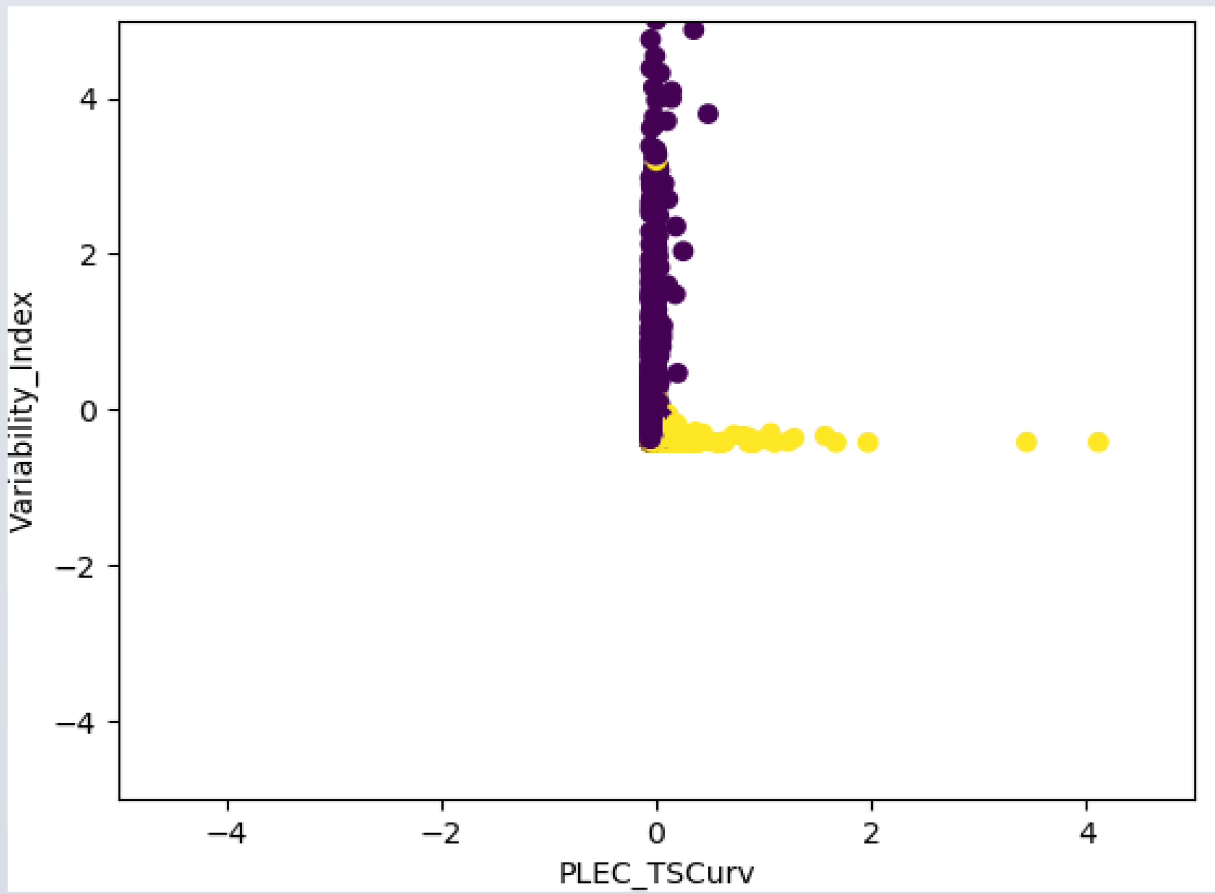
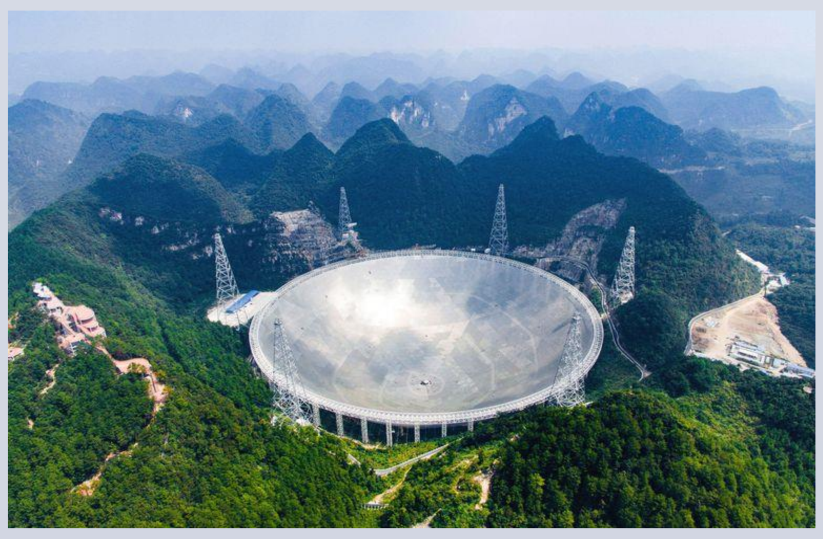


Figure 3. The data points projected into features 2D-plane(data had been normalized). A clear pattern can be seen from the two graph. Yellow dot represent pulsar and black dot represent AGN.

Further discussion

We train four new models using all data and use it to predict on the unassociated sources. We currently collaborating with FAST team and try to the radio counter-part of the pulsars predicted by our models.

Figure 4. Five-hundred-meter aperture spherical radio telescope(FAST) is the largest single dish radio telescope.



References

[1] WB Atwood, Aous A Abdo, Markus Ackermann, W Althouse, B Anderson, M Axelsson, L Baldini, J Ballet, DL Band, Guido Barbiellini, et al. The large area telescope on the fermi gammaray space telescope mission. The Astrophysical Journal, 697(2):1071, 2009.

[2] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning, volume 1. Springer series in statistics New York, 2001.

[3] PM Saz Parkinson, H Xu, PLH Yu, D Salvetti, M Marelli, and AD Falcone. Classification and ranking of fermi lat gamma-ray sources from the 3fgl catalog using machine learning techniques. The Astrophysical Journal, 820(1):8, 2016.