Supplement to Radio Galaxy Zoo: Machine learning methods for radio source host galaxy cross-identification

M. J. Alger et al.

Accepted XXX. Received XXX

1 CLASSIFIERS

We use three different classifiers as our binary classification model: logistic regression, convolutional neural networks, and random forests.

1.1 Logistic Regression

Logistic regression is a binary classification model. It is linear in the feature space and outputs the probability that the input has a positive label. The model is (Bishop 2006):

$$f(\mathbf{x}) = \sigma(\mathbf{w} \cdot \mathbf{x} + b) \quad , \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^D$ is a weight vector, $b \in \mathbb{R}$ is a bias term, $\mathbf{x} \in \mathbb{R}^D$ is the feature representation of a candidate host, and σ is the logistic sigmoid function:

$$\sigma(a) = (1 + \exp(-a))^{-1} \quad . \tag{2}$$

The logistic regression model is fully differentiable, and the weight vector $\mathbf{w}$ can therefore be learned using gradient methods.

1.2 Convolutional neural networks

Convolutional neural networks (CNN) are a biologically-inspired prediction model for prediction with image inputs. The input image is convolved with a number of filters to produce output images called feature maps. These feature maps can then be convolved again with other filters on subsequent layers, producing a network of convolutions. The whole network is differentiable with respect to the values of the filters and the filters can be learned using gradient methods. The final layer of the network is logistic regression, with the convolved outputs as input features. For more detail, see LeCun et al. (subsection II.A 1998). We use Keras (Chollet et al. 2015) to implement our CNN.

CNNs have recently produced good results on large image-based datasets in astronomy (e.g. Dieleman et al. 2015, ; Lukic et al. in prep). We employ only a simple CNN model in this paper as a proof of concept that CNNs may be used for class probability prediction on radio images. The model architecture we use is shown in Figure 1.

1.3 Random Forests

Random forests are an ensemble of decision trees (Breiman 2001). It considers multiple subsamples of the training set, where each bootstrap subsample is sampled with replacement from the training set. For each subsample a decision tree classifier is constructed by repeatedly making axis-parallel splits based on individual features. In a random forest the split decision is taken based on a random subset of features. To classify a new data point, the random forest takes the weighted average of all classifications produced by each decision tree.

References

Bishop C. M., 2006, Pattern recognition and machine learning. Springer
Breiman L., 2001, Machine Learning, 45, 5
Chollet F., et al., 2015, Keras, https://github.com/fchollet/keras
Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
LeCun Y., Bottou L., Bengio Y., Haffner P., 1998, Proceedings of the IEEE, 86, 2278

This paper has been typeset from a TEX/LATEX file prepared by the author.

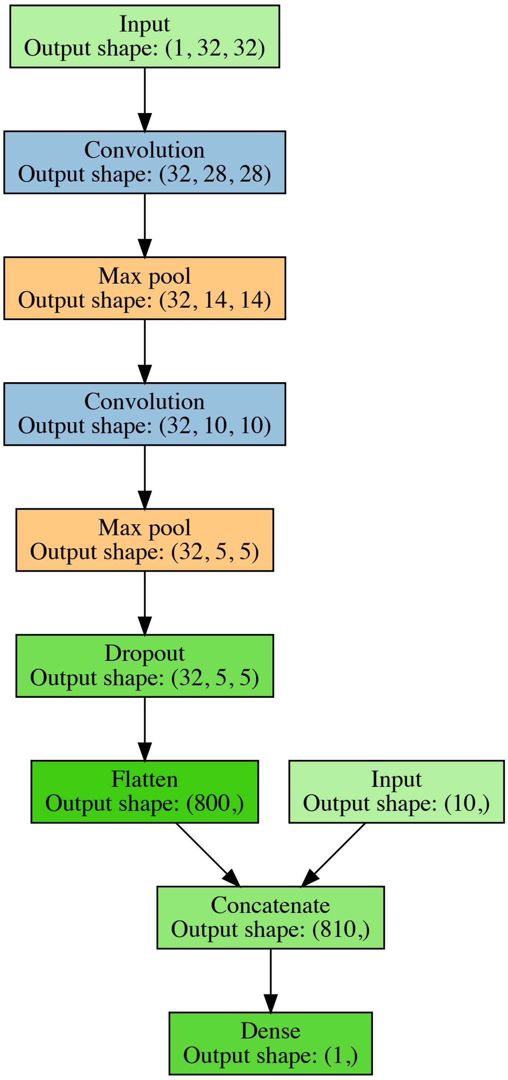

Figure 1. Architecture of our CNN. The concatenate layer flattens the output of the previous layer and adds the 10 features derived from the candidate host in SWIRE, i.e. the flux ratios, stellarity indices, and distance. The dropout layer randomly sets 25% of its inputs to zero during training to prevent overfitting. Diagram based on `https://github.com/dnouri/nolearn`.