

Machine Classification of Transient Images

Lise du Buisson¹†, Navin Sivanandam²‡, Bruce A. Bassett^{1,2,3}¶
and Mathew Smith⁴

¹Department of Mathematics and Applied Mathematics, University of Cape Town,
Cross Campus Rd, Rondebosch 7700, South Africa

²African Institute for Mathematical Sciences, 6–8 Melrose Rd, Muizenberg 7945, South Africa

³South African Astronomical Observatory, Observatory Rd, Observatory 7925, South Africa

⁴University of the Western Cape, Robert Sobukwe Rd, Bellville 7535, South Africa

Abstract. Using transient imaging data from the 2nd and 3rd years of the SDSS supernova survey, we apply various machine learning techniques to the problem of classifying transients (e.g. SNe) from artefacts, one of the first steps in any transient detection pipeline, and one that is often still carried out by human scanners. Using features mostly obtained from PCA, we show that we can match human levels of classification success, and find that a K-nearest neighbours algorithm and SkyNet perform best, while the Naive Bayes, SVM and minimum error classifier have performances varying from slightly to significantly worse.

Keywords. methods: data analysis, methods: statistical, techniques: image processing, techniques: photometric, surveys.

1. Introduction

Next-generation of survey experiments, like the Large Synoptic Survey Telescope (LSST) and the Dark Energy Survey (DES) will produce a huge photometric data deluge that will require new statistical and machine learning techniques for processing. This data deluge create problems in the supernovae (SNe) detection pipeline. Difference images are scanned for objects of interest (e.g. potential SNe) and artefacts are removed to avoid overwhelming spectroscopic follow-up capabilities with an unnecessary flood of data.

This classification of images into real objects and artefacts are carried out by human hand scanners, leading to large amounts of images having to be scanned each night. Additionally, human classifiers have different sets of biases that change with mood and circumstances, something that cannot be characterised in a systematic way and leads to irreproducible results.

Using machine learning techniques instead of human scanners for this classification is therefore an important next step in achieving the goals of transient surveys in the next decade. We here compare various machine learning algorithms in order to see if a desired classification accuracy can be achieved. It should be noted that this contribution is a condensed version of work that is soon to be published (du Buisson *et al.* 2015).

2. Data

We use g , i and r -band difference images from the 2nd and 3rd years of the Sloan Digital Sky Survey (SDSS) supernova survey (Frieman *et al.* 2008, Sako *et al.* 2014). The survey's transient detection algorithms consisted of producing difference images, which were then

† Email: lisedubuisson@gmail.com
‡ Email: navin.sivanandam@gmail.com
¶ Email: bruce.a.bassett@gmail.com

Table 1. The three hierarchical classification systems and how they relate to one another. A description of the different original classes (Frieman *et al.* 2008) can be found in Bassett *et al.* (2005).

Original class	Visual class	Real/Not-Real
Artefact	Artefact	Not-Real
Dipole	Dipole/Saturated	Not-Real
Saturated Star	Dipole/Saturated	Not-Real
Moving	Real	Real
Variable	Real	Real
Transient	Real	Real
SN Other	Real	Real
SN Bronze	Real	Real
SN Silver	Real	Real
SN Gold	Real	Real

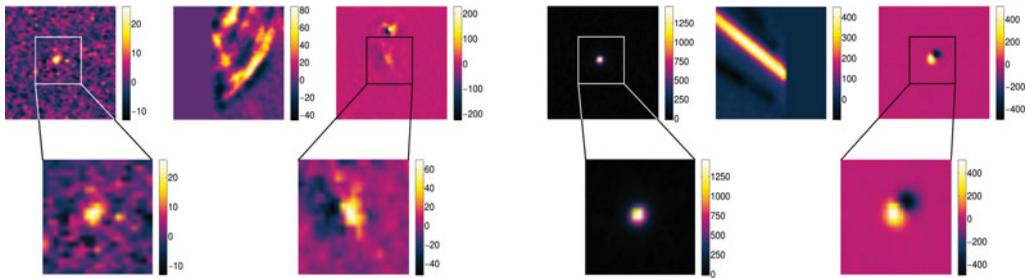


Figure 1. The three visual classes in the *r*-band. **(Left)** Representative (SNR < 20) examples of (left to right) real, artefact and dipole/saturated with zooms. **(Right)** High-quality (SNR > 40) examples of (left to right) real, artefact and dipole/saturated with zooms.

labelled as candidates for human hand scanning if they passed the necessary threshold cuts. Because this process is not perfect, it introduced a large number of artefacts into the candidate group.

Human hand scanners then sorted each candidate object into one of ten original classes. We decided to re-group the original classes into three visual classes (see Fig. 1) based on the similarities in appearance between some of the original classes, and used these three classes to carry out single-class PCA (see section 4). Ultimately, we are concerned with being able to correctly predict whether an object is real or not, and this is therefore the main classification used by the classifiers. To see how these hierarchical classification systems relate to one another, see Table 1.

3. Testing and Performance Measures

Testing: 25% of our data set were held apart for final testing after the various classifiers have been trained using the remaining 75% of the data, of which a further 30% were set apart for preliminary testing and parameter optimisation of the learners.

Performance Measures: For preliminary testing, optimisation and final testing we used two common performance measures, namely the accuracy (A) and recall (R), defined in terms of true/false positives/negatives (t_p , t_n , f_p and f_n) as follows:

$$A = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad R = \frac{t_p}{t_p + f_n} \quad (3.1)$$

A Receiver Operating Characteristic (ROC) curve shows a binary classifier's perfor-

mance as its discrimination threshold is varied. It is a plot of the true positive rate (TPR), or recall, versus the false positive rate (FPR) (Eq. 3.2) at various different threshold values.

$$FPR = \frac{f_p}{f_p + t_n} \quad (3.2)$$

The “Area Under Curve” (AUC) is often used for classifier comparison, and is equal to the probability that a classifier will correctly classify a randomly chosen instance (Hanley & McNeil 1982). We used the AUC to compare classifiers.

4. Feature Extraction

Multi-class PCA: We carry out Principal Component Analysis (PCA) on the full training set and then characterise individual objects by including a certain number of principal components (PCs) in their feature vectors.

Single-class PCA: We apply PCA independently on the three different visual classes (see Table 1), resulting in each class having a unique eigenspace. An image’s features is obtained by reconstructing it using the PCs of each individual class in turn, and then finding the error between each reconstructed image and the original. The three calculated errors are then features for an image.

LDA: We use Linear Discriminant Analysis (LDA) components together with the PCA components as features.

5. Machine Learning Techniques

Minimum Error Classification (MEC): MEC takes only the three error features (section 4) as input, and classifies an object to be of the class corresponding to the minimum error.

Naïve Bayes (NB): NB is a practical Bayesian classifier, resting on the assumption that, given the class, the feature values are conditionally independent.

K-Nearest Neighbours (KNN): KNN finds the K nearest training objects in feature space to a test item, averages their classes (with a uniform or distance weight) and classifies the test item accordingly.

Support Vector Machine (SVM): An SVM is a maximum margin classifier – it aims to find a decision boundary in such a way that the largest possible separation between classes is obtained. The kernel trick can be applied to transform from the linear decision boundary to a non-linear one.

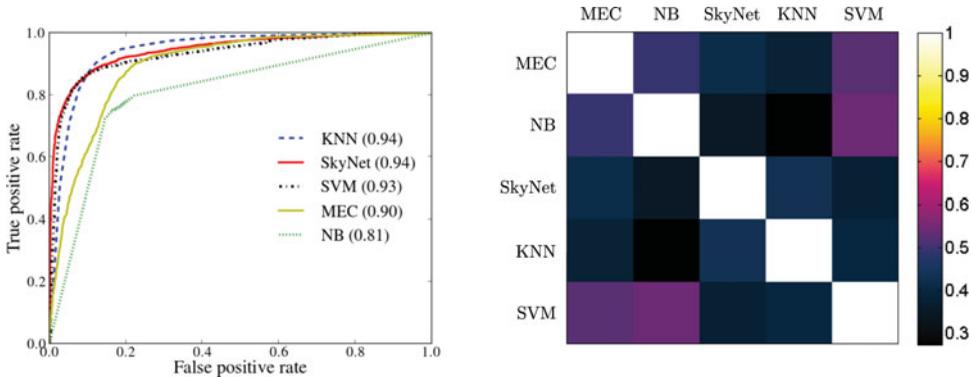
Artificial Neural Network (ANN): ANNs have interconnected nodes that processes information and passes it on to other nodes along weighted connections that are directed from an input layer of nodes to an output one. An ANN is trained to learn a mapping between the input and output layer, and it then predicts the outputs for new input data. We use SkyNet as our ANN (Graff *et al.* 2014).

6. Results and Discussion

The performance results of the various classifiers upon final testing can be seen in Table 2, and the ROC curve for each of the classifiers can be seen in Fig. 2. It is interesting that the simple KNN performs better than (or as well as) the most intricate classifiers, like SkyNet. Furthermore, it can be seen that KNN’s, SkyNet’s and MEC’s classification performance on fake SNe equals that of human scanners (0.956 ± 0.010 fake SN-tag recall (Kessler 2007)). In order to determine how much overlap there is between the performance of the different classifiers on the test set we calculated the Cohen’s Kappa

Table 2. The classification performance of the various classifiers in final testing (see section 3) together with the performance on the fake SNe in the dataset.

Machine Learning Technique	AUC	Accuracy	Recall	Fake SN-tag	Recall
KNN	0.94	0.89	0.90	0.96	
SkyNet	0.94	0.88	0.89	0.96	
SVM	0.93	0.86	0.90	0.94	
MEC	0.90	0.84	0.92	0.96	
NB	0.81	0.78	0.77	0.85	

**Figure 2.** (Left) ROC curve for the five classifiers, listed in order of performance based on the AUC statistic (section 3), given in brackets. (Right) The Cohen's Kappa value (κ) for pairs of classifiers. Agreement: $0.0 < \kappa \leq 0.2$ – Slight; $0.2 < \kappa \leq 0.4$ – Fair; $0.4 < \kappa \leq 0.6$ – Moderate; $0.6 < \kappa \leq 0.8$ – Substantial; $0.8 < \kappa \leq 1.0$ – Almost Perfect (Landis & Koch 1977).

statistic κ (Cohen 1960) for each pair of classifiers. Fig. 2 shows that no pair of classifiers is highly correlated, which suggests that there may be significant gain from combining classifiers. Further improvements will also likely come from adding host galaxy and multi-epoch information, studies that we leave to the future.

Acknowledgements

LdB is supported by an NRF/SKA scholarship, NS is funded by a Claude Leon Foundation fellowship and BB and MS is funded by the SKA and the NRF.

References

- Bassett, B. A., Richmond, M., & Frieman, J. 2005, [url:http://spiff.rit.edu/richmond/sdss/sn_survey/scan_manual/sn_scan.html](http://spiff.rit.edu/richmond/sdss/sn_survey/scan_manual/sn_scan.html)
- Cohen, J. 1960, *Educational and Psychological Measurement*, 20, 37
- du Buisson, L., Sivanandam, N., & Bassett, B. A., Smith M., 2015, submitted to MNRAS, arXiv:1407.4118 [astro-ph.IM]
- Frieman, J. A., Bassett, B., Becker, A., & Choi, C., *et al.* 2008, *AJ*, 135, 338
- Graff, P., Feroz, F., Hobson, M. P., & Lasenby, A. 2014, *MNRAS*, 441, 1741
- Hanley, J. & McNeil, B. J. 1982, *Radiology*, 143, 29
- Kessler, R. 2007, Private Communication
- Landis, J. R. & Koch, G. G. 1977, *Biometrics*, 33, 159
- Sako, M., Bassett, B., Becker, A. C., & Brown, P. J., *et al.* 1977, *ArXiv e-prints*, [adsurl:http://adsabs.harvard.edu/abs/2014arXiv1401.3317S](http://adsabs.harvard.edu/abs/2014arXiv1401.3317S)