Photometric classification of BHB stars

Kester Smith, Coryn Bailer-Jones, Rainer Klement

Max-Planck Institute for Astronomy
Koenigstuhl 17
69117 Heidelberg, Germany

Email: smith@mpia-hd.mpg.de

Introduction

Blue horizontal branch (BHB) stars are useful as tracers of galactic structure because they have a narrow range of intrinsic luminosities, and so can provide good photometric parallaxes, and because they are intrinsically bright, so tracing a large volume. Photometric methods of identifying them can be reasonably efficient, since their colours are distinctive, but suffer from contamination from MS stars and from blue stragglers.

Spectroscopic identification is less prone to contamination, but also less reliable at fainter magnitudes, and can be used only for smaller samples.

We have developed a method for photometrically identifying BHB stars from SDSS colours. We start from a recent spectroscopic study (Xue et al. 2009), and use their results to build training sets with which we train models based on various standard machine learning techniques.

The Xue sample

The sample of Xue et al. contains 10224 objects, of which 2558 were identified as BHB stars. Figure 1 shows the $u-g$, $g-r$ colour-colour diagram, with BHB stars in red.

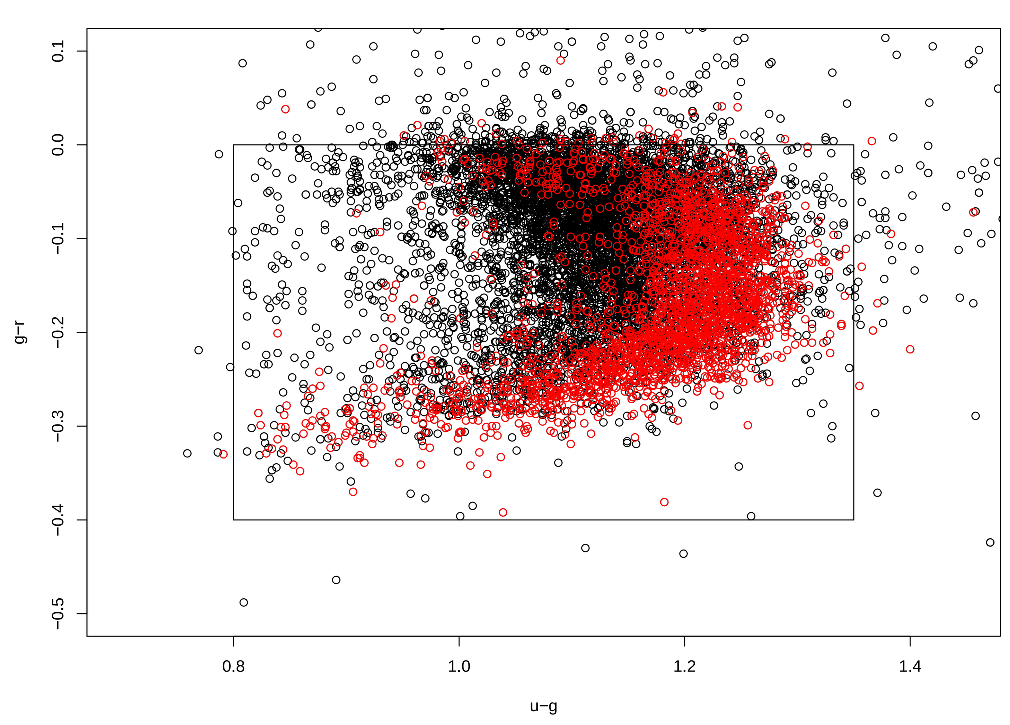

Figure 1: $u-g$, $g-r$ colours of Xue et al sample. BHB stars (from Xue) are red. The inner box is the selection box suggested by Yanny et al. (1999) and used by Xue.

The fraction of BHB stars in the sample has a strong dependence on g magnitude, as illustrated in Figure 2.

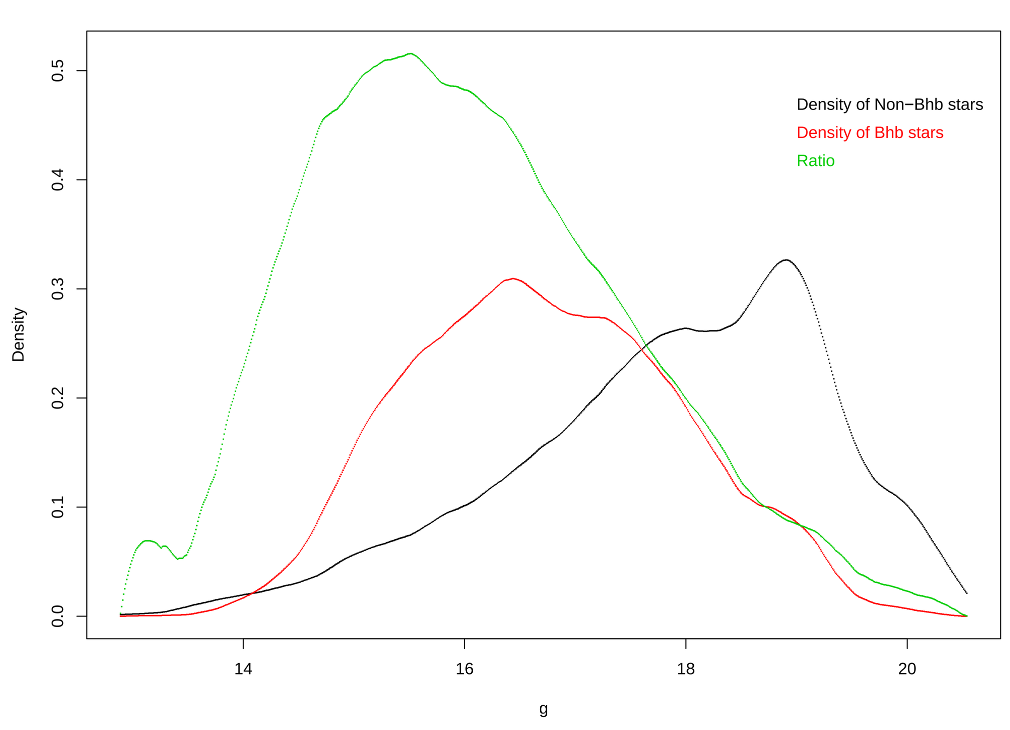

Figure 2: Density functions of BHB stars and Non-BHB stars with g magnitude. The ratio (green line) is weighted for the relative numbers.

Prior probabilities

- The Xue sample contains 2500 BHB stars out of 10000 sources, suggesting a prior of 0.25. The selection function is however unknown
- We obtained from T. Beers and coworkers a randomly sampled set of spectra from stripe 82.
- 48 of the Xue sources were identified in the Beers sample. 7 of these were BHB stars
- Since the Beers sample has very different g magnitude distribution to the Xue sample, we constructed a magnitude weighted prior for the Xue sources in stripe 82. This prior was 0.186
- We propose that the Beers prior (7/48=0.15) is appropriate for stripe 82 with the Beers magnitude distribution. We use the Xue predicted stripe 82 prior (0.19) to obtain a correction factor and apply this to the whole (SDSS) sky Xue prior of 0.25.
- The final prior is then $0.25 \times 0.15/0.17 = 0.22$

Support vector classifier

We consider here two classification algorithms, a Kernel density estimator (KDE) and a support vector machine (SVM). We used an SVM implementation called libSVM available online at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`. The results are shown in Table 1. The completeness (fraction of input sources correctly classified) and contamination (fraction of an output class that are false positives) are shown as a function of g magnitude in Figure 3. The performance is obviously degraded strongly for $g > 18$.

Flat prior		
	bhb star	Non-bhb star
BHB STAR	1046	222
NON-BHB STAR	264	1004
With prior		
BHB STAR	715	553
NON-BHB STAR	86	1182
completeness	0.564	
contamination	0.107	

Table 1: Confusion matrix for SVM method. Rows are true input classes, columns show output class. Last panel shows the completeness (fraction of input sources correctly classified) and contamination (fraction of output sample that is incorrect)

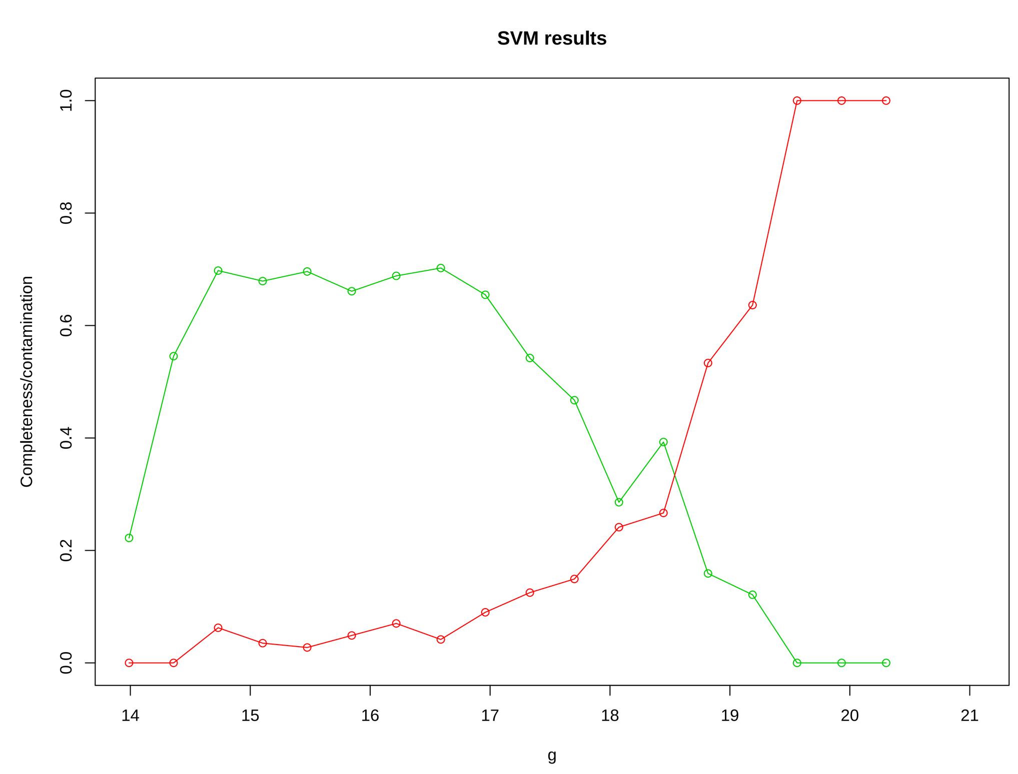

Figure 3: Performance of the SVM classifier for different g magnitudes. The completeness is shown in green and the contamination in red. The performance is strongly degraded at faint magnitudes.

Kernel density estimator

The KDE classifier works by modelling the density of each class in the data space and comparing the density functions to find the probability that a test point belongs to a particular class.

The results of a KDE classifier are shown in Table 2, and the completeness and contamination as functions of g are plotted in Figure 4. The contamination rises strongly beyong $g = 18$, as for the SVM, but the completeness is always substantially worse than the SVM.

Flat prior		
	bhb star	Non-bhb star
BHB STAR	1013	255
NON-BHB STAR	338	930
With prior		
BHB STAR	245	1023
NON-BHB STAR	67	1201
completeness	0.193	
contamination	0.214	

Table 2: Confusion matrix for KDE method. Rows are true input classes, columns show output class. Last panel shows the completeness and contamination

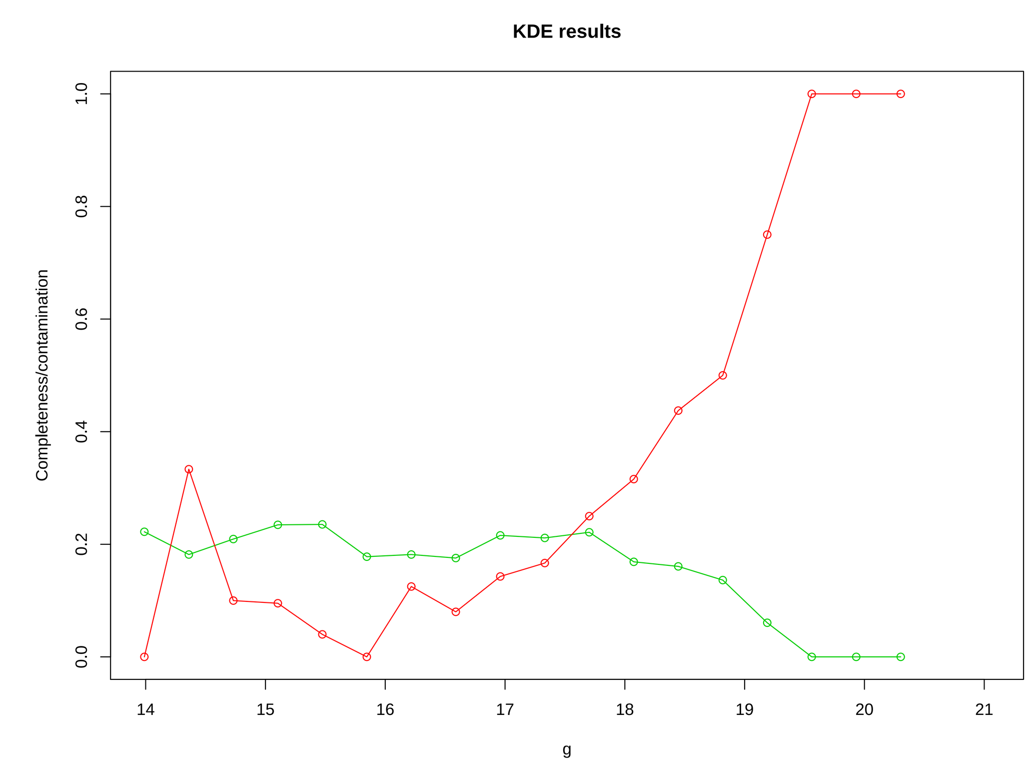

Figure 4: Performance of the KDE classifier for different g magnitudes. The completeness is shown in green and the contamination in red

New sample

We obtained a selection of SDSS DR7 sources within the Yanny colour box (see Figure 1) and classified them with the SVM classifier. 11940 sources were identified as BHB sources. We obtained photometric parallaxes using the method outlined in Sirko et al. (2004). Figure 5 shows the distances for sources in the direction $342 < l < 352$, $49 < b < 59$ (red histogram). A section of the Sagittarius stream is visible at 50kpc. The grey histogram shows a control region ($162 < l < 172$, $30 < b < 45$) for comparison.

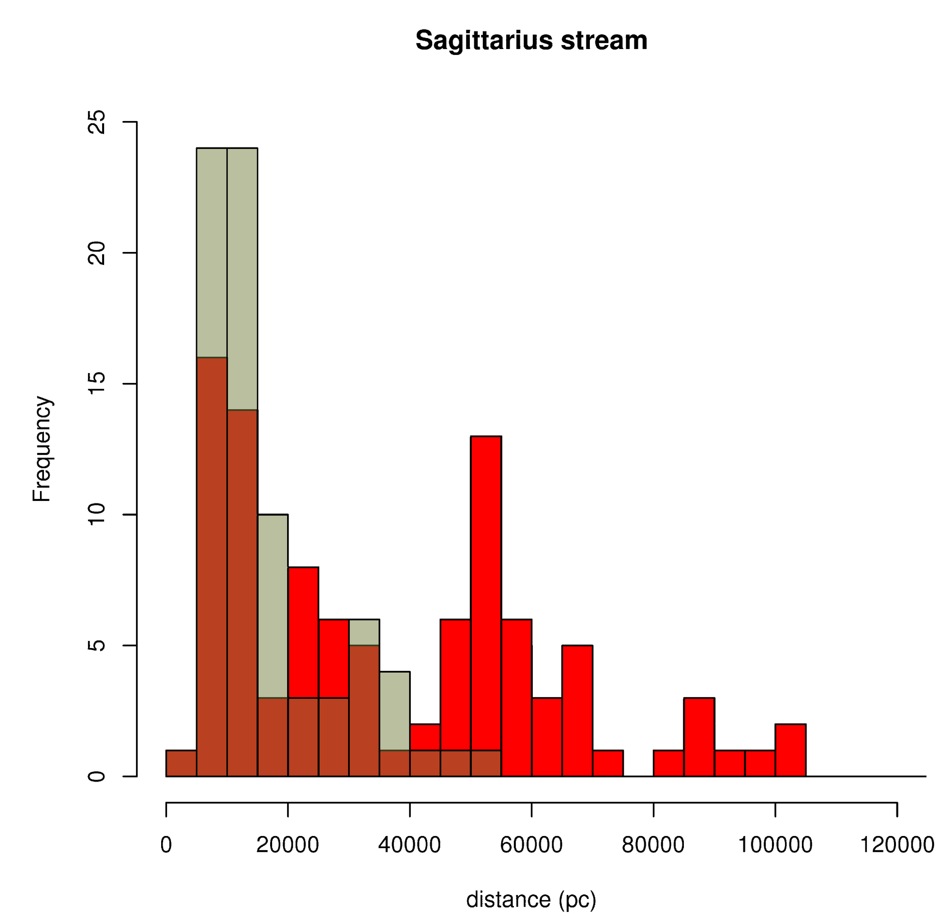

Figure 5: distances to newly identified BHB stars in the direction of part of the Sagittarius stream, and to a control field (grey).

Summary

It is possible to identify BHB stars from photometric colours alone at reasonable levels of purity and completeness for magnitudes $g < 18$. The support vector machine method apparently provides the best completeness at a low level of contamination. The KDE method, as employed here, achieves a low contamination but at the expense of large losses in completeness.

Both methods are limited at faint magnitudes, perhaps artificially, by the loss of sensitivity to BHBs in the training data, i.e. the loss of sensitivity of Xue's spectroscopic method. We need to carry out a full investigation of the nature of this effect in order to assess the prospects for extending the method to magnitudes $g > 18$.

References

References

[1] Sirko, E., et al., 2004, *ApJ* 127, p. 899.

[2] Vapnik, V., *The nature of statistical learning theory* (New York, Springer Verlag, 1995)

[3] Xue, X. X., et al., 2009, *ApJ* 684, p. 1143.

[4] Yanny et al. 1999, *ApJ* 540, p.825