Astronomy & Astrophysics

A package for the automated classification of periodic variable stars*

Dae-Won Kim** and Coryn A. L. Bailer-Jones

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

ABSTRACT

We present a machine learning package for the classification of periodic variable stars. Our package is intended to be general: it can classify any single band optical light curve comprising at least a few tens of observations covering durations from weeks to years with arbitrary time sampling. We use light curves of periodic variable stars taken from OGLE and EROS-2 to train the model. To make our classifier relatively survey-independent, it is trained on 16 features extracted from the light curves (e.g., period, skewness, Fourier amplitude ratio). The model classifies light curves into one of seven superclasses – δ Scuti, RR Lyrae, Cepheid, Type II Cepheid, eclipsing binary, long-period variable, non-variable – as well as subclasses of these, such as ab, c, d, and e types for RR Lyraes. When trained to give only superclasses, our model achieves 0.98 for both recall and precision as measured on an independent validation dataset (on a scale of 0 to 1). When trained to give subclasses, it achieves 0.81 for both recall and precision. The majority of misclassifications of the subclass model is caused by confusion within a superclass rather than between superclasses. To assess classification performance of the subclass model, we applied it to the MACHO, LINEAR, and ASAS periodic variables, which gave recall/precision of 0.92/0.98, 0.89/0.96, and 0.84/0.88, respectively. We also applied the subclass model to HIPPARCOS periodic variable stars of many other variability types that do not exist in our training set, in order to examine how much those types degrade the classification performance of our target classes. In addition, we investigate how the performance varies with the number of data points and duration of observations. We find that recall and precision do not vary significantly if there are more than 80 data points and the duration is more than a few weeks.

Key words. methods: data analysis – methods: statistical – stars: variables: general – techniques: miscellaneous

1. Introduction

Periodic variable stars have played a central role in our learning about the Universe. RR Lyrae stars, for instance, have been used to determine distances to globular clusters, to study Galactic structure, and also to trace the history of the formation of our Galaxy (Vivas et al. 2001; Carretta et al. 2000; Catelan 2009; Sesar 2011). Cepheid variables are important standard candles for measuring distances to external galaxies (Freedman et al. 2001; Riess et al. 2011). Eclipsing binaries are vital for estimating stellar masses (Torres et al. 2010), and δ Scuti are used in asteroseismological studies to learn about the internal structure of stars (Brown & Gilliland 1994).

As the volume of astronomical survey data grows, the reliable automated detection and classification of objects including periodic variable stars is becoming increasingly important. LSST, for instance, will produce about 20 TB of data per night, requiring robust and powerful automated methods to classify the sources and light curves (see Ivezic et al. 2008 and references therein). The *Gaia* satellite (Perryman et al. 2001) is scanning the whole sky to monitor a billion stars, and is using machine learning and statistical analysis techniques to classify the sources and to estimate the astrophysical parameters (effective temperature, extinction, etc.; Bailer-Jones et al. 2013) as well as to analyze the light curves (Eyer et al. 2014).

In this paper, we introduce UPSILoN: AUtomated Classification for Periodic Variable Stars using MachIne LearNing. This software package classifies a light curve into a class of a periodic variable star in an automatic manner. It can be applied to light curves from optical surveys regardless of survey-specific characteristics (e.g., photometric accuracy, sampling rate, duration, etc.) as long as the light curves satisfy a few broad conditions:

– it contains at least a few tens of data points that sample the characteristic variability well;
– the observation duration is more than a few weeks;
– the light curve is obtained in an optical band.

If multiple optical bands are available, UPSILoN can classify them separately. It does not use – or rely on – color information. UPSILoN is currently trained to classify into six types of periodic variables: δ Scuti, RR Lyrae, Cepheid, Type II Cepheid, eclipsing binary, long-period variable, and their subclasses. UPSILoN can also separate non-variables from these periodic variables, which is important because the majority of light curves from most time-series surveys are non-variables.

Some previous studies have investigated the possibility of producing general purpose classifiers such as UPSILoN. For instance, Blomme et al. (2011) used a mixture of HIPPARCOS (Perryman et al. 1997), OGLE (Udalski et al. 1997), and CoRoT (Auvergne et al. 2009) sources to train a classification model, and then applied this to the TrES dataset (O'Donovan et al. 2009). Debosscher et al. (2009), Richards et al. (2012a) used HIPPARCOS and OGLE sources to train classification models and

---

applied them to CoRoT and ASAS (Pojmanski 1997) data, respectively. The latest work of this kind was done by Paegert et al. (2014) who developed a periodic variable classifier optimized for eclipsing binaries based on neural networks (Mackay 2003). Most of these studies confirmed that their trained models successfully classified target sources from other surveys that were not included in their training sets.

UPSILoN differs from these previous works in several ways:

– General purpose classifier
  UPSILoN has been trained to achieve the best classification quality for a broad set of periodic variables, rather than having been optimized for a specific type of variable.

– Rich training set
  We use a mixture of the OGLE and EROS-2 periodic variable stars to provide a relatively complete set of periodic variables for training (see Sect. 2.1).

– Simple and robust variability features
  We employ features that have been shown to quantify variability of both periodic and non-periodic light curves (Kim et al. 2011, 2012, 2014; Long et al. 2012). Most of these features are simple and easily calculated (see Sect. 2.2).

– Random forest (Breiman 2001)
  We use the random forest algorithm, which is one of the most successful machine learning methods in many astronomical classification problems (e.g., Carliles et al. 2010; Pichara et al. 2012; Masci et al. 2014; Nun et al. 2014 and references therein). This is described in Sect. 2.3.

– Application to MACHO, LINEAR, ASAS, and Hipparcos datasets
  To assess classification quality, we applied UPSILoN to the MACHO, LINEAR, ASAS, and Hipparcos periodic variables, as described in Sects. 3 and 4.

– Experiments with resampled MACHO and ASAS light curves
  We resampled the MACHO and ASAS periodic variables' light curves in order to build samples containing different number of measurements over different observation durations. Classification results using these samples are given in Sect. 5.

– Ready-to-use Python package
  The UPSILoN package including the random forest model is available from the GitHub repository[1].

In Sect. 2 we introduce the training set and variability features that we used to train the UPSILoN classification model, and we explain the training process. Sections 3–5 summarize the classification performance using the periodic-variable light curves collected from the MACHO (Alcock et al. 1996b), LINEAR (Stokes et al. 2000), ASAS (Pojmanski 1997), and Hipparcos (Perryman et al. 1997) surveys. Section 6 gives a summary.

2. UPSILoN classifier

2.1. Training set

As the quality of classification strongly depends on the training data, it is important to construct a rich and clean training set. We used a mixture of periodic variable stars from the OGLE (Udalski et al. 1997) and EROS-2 (Tisserand et al. 2007) Large Magellanic Cloud (LMC) databases. We chose these because they provide the most complete and well-sampled set of

Table 1. Acronyms of the variable types.

Variable type	Acronym
δ Scuti star	DSCT
RR Lyrae	RRL
Cepheid	CEPH
Type II Cepheid	T2CEPH
Eclipsing binary	EB
Long-period variable	LPV
Non-variables	NonVar

light curves in the optical for several types of periodic variable stars. We identified the OGLE periodic variables from several literature sources: Soszyński et al. (2008a,b, 2009a,b), Poleski et al. (2010), Graczyk et al. (2011). For the EROS-2 periodic variables, we used the training set of the periodic variable stars from Kim et al. (2014). As described in that paper, this data set was visually inspected and cleaned. From these source we complied a catalog of periodic variables comprising δ Scuti stars, RR Lyraes, Cepheids, Type II Cepheids, eclipsing binaries, and long-period variables (these are the superclasses). Table 1 gives the acronyms of each superclass. We also compiled subclasses (e.g., RR Lyrae ab, c, d, and e) of each class. We then added non-variable sources to the catalog, which is important because a given set of survey data that we want to classify will generally include many such sources. For these, we selected 5000 EROS-2 non-variables from Kim et al. (2014) and 5000 OGLE light curves randomly selected around the Large Magellanic Cloud[2] (Szymanski 2005). We visually examined these light curves and removed light curves showing variability. Note, however, that it remains difficult for any general classifier to identify non-variables on account of the wide range of survey-dependent characteristics of such stars (e.g., due to systematic trends, variable noise levels). Thus although we added the OGLE and EROS-2 non-variable sources to our training set, this does not ensure that the classifier will efficiently exclude non-variables during the classification of light curves from other surveys.

Before extracting variability features (Sect. 2.2) from these training-set light curves, we first refined each light curve by

– removing data points with bogus photometry (e.g., 99.999 in EROS-2). The EROS-2 light curves have many such measurements.
– removing data points that lie more than 3σ above or below the mean (no iteration), where σ is the standard deviation of the light curve. This cut does not significantly alter light curves because it removes generally less than 1% of the data points per light curve.

In order to build the cleanest possible training set, we then excluded light curves from the training set if

– the number of measurements was fewer than 100.
– an estimated period was spurious, related to the mean sidereal day ($\sim$0.997 days), the mean sidereal month ($\sim$27.322 days), the mean synodic month ($\sim$29.531 days), and integer multiples thereof. We determined additional spurious periods by examining a period versus period signal-to-noise ratio (SNR) plot and a period histogram as done by

---

[1] https://goo.gl/xmF06Q

[2] Using the website http://ogledb.astrouw.edu.pl/~ogle/photdb/

Table 2. The number of sources in each class in the training set.

Superclass	Subclass	Number	Note
DSCT		3209	
RRL			
	ab	19 921	
	c	4974	
	d	1077	
	e	1327	
CEPH			
	F	2981	fundamental
	1O	2043	first overtone
	Other	443	
EB			
	EC	1398	contact
	ED	19 075	detached
	ESD	7339	semi-detached
LPV			
	Mira AGB C	1361	carbon-rich
	Mira AGB O	783	oxygen-rich
	OSARG AGB	25 284	
	OSARG RGB	29 516	
	SRV AGB C	6062	carbon-rich
	SRV AGB O	8780	oxygen-rich
T2CEPH		300	
NonVar		8050	
Total		143 923	

Kim et al. (2014). Most of the additional spurious periods are close to the sidereal day.

Table 2 shows the number of each variable type in the final training set. "CEPH Other" class contains subclasses of Cepheid variables such as second overtone (2O), double-mode in fundamental and first overtone (F/1O), double-mode in first and second overtone (1O/2O), etc. The total number of sources in the final training set is 143 923. As the table shows, the training set contains the major and strongly varying periodic variables. The goal of UPSILoN is not to classify other variability types such as rotational variability, eruptive variability, etc. Sect. 4 presents the possible contamination caused by such variability types using a set of HIPPARCOS periodic variable stars from Dubath et al. (2011).

The average number of data points in either an OGLE or an EROS-2 light curve is about 500. The observation duration is about seven years for the EROS-2 light curves, and about eight years for the OGLE light curves.

2.2. Survey-independent variability features

We used a total of 16 variability features (Table 3) to quantify variability of a light curve. Note that we only chose features that are relatively survey-independent[3] and excluded features that are only available in certain surveys (e.g., ones based on colors). This is important because our goal is to develop a general purpose classifier that can be applied to an essentially any long-duration time-series survey done in an optical band. Among these 16 features, 11 are taken from our previous work (Kim et al. 2011, 2012, 2014), and two are from Long et al. (2012). These publications show that these features are useful

---

[3] The minimum requirements for a light curve are given in Sect. 1.

and robust to classify either periodic or non-periodic variables. We added three more features, A, H_1, and W, defined as follows.

- A
  The ratio of the squared sum of residuals of magnitudes that are either brighter than or fainter than the mean magnitude. This is defined as

$$A = \frac{\sigma_{\rm f}}{\sigma_{\rm b}} \quad \text{where} \tag{1}$$

$$\sigma_{\rm f}^2 = \frac{1}{N} \sum_i^N (m_i - \overline{m})^2, \tag{2}$$

$$\sigma_{\rm b}^2 = \frac{1}{M} \sum_j^M (m_j - \overline{m})^2, \tag{3}$$

$\overline{m}$ is the mean magnitude of a light curve, m_i are measurements fainter than $\overline{m}$, and m_j are measurements brighter than $\overline{m}$. If a light curve has a sinusoidal-like variability, A is close to 1. For EB-like variability, having sharp flux gradients around its eclipses, A is larger than 1.

- H_1
  The rms amplitude of the first Fourier component of the Fourier decomposition of the light curve (Petersen 1986). This is

$$H_1 = \sqrt{a_1^2 + b_1^2} \tag{4}$$

with the Fourier series defined as

$$m_0 + \sum_{k=1}^M [a_k \sin(2\pi k f) + b_k \cos(2\pi k f)] \tag{5}$$

where M is the order of the Fourier series and f is the phase: $0 \le f < 1$; $j = 1, 2, \ldots, N$, where N is the number of data points. m_0, a_k and b_k are parameters to be fitted. In this work, the value of M is 5.

- W
  The Shapiro-Wilk statistic, which tests the null hypothesis that the samples (e.g., measurements in the light curve) are from a normal distribution (Shapiro & Wilk 1965). The Shapiro-Wilk test is generally more powerful than other normality tests (e.g., Kolmogorov-Smirnov test, Anderson-Darling test, etc.) for different sample sizes (see Saculinggan & Balase 2013 and references therein).
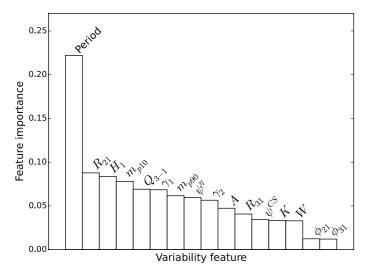
Figure 1 shows the importance of each feature in the trained subclass classifier (see Sect. 2.3) as estimated by the random forest algorithm on the training data. A random forest ranks the importance of ith feature by using *out-of-bag* (oob) samples, which is about one-third of the training data and not used for constructing a tree. A random forest predicts classes of the oob samples and counts the votes for the correct classes. A random forest then randomly permutes the values of ith feature in the oob samples, predicts classes again with the permuted oob samples, and counts the correct votes. The average difference of the count over all trees is the importance of ith feature. In addition, a random forest can also estimate the feature importance based on the mean decrease in node impurity (e.g., Gini impurity), which is the total sum of the decrease in node impurity averaged over all trees. This method gives generally consistent results with the method using permutation (see Breiman et al. 1984; Breiman 2001 and references therein). In this work, we used the Gini impurity measure to deduce the feature importance.

Table 3. 16 variability features.

Feature	Description	Reference[a]
Period	Period derived by the Lomb-Scargle algorithm	
ψ^η	η (von Neumann 1941) of a phase-folded light curve	
ψ^{CS}	Cumulative sum index of a phase-folded light curve	
R_{21}	2nd to 1st amplitude ratio derived using the Fourier decomposition	
R_{31}	3rd to 1st amplitude ratio derived using the Fourier decomposition	
ϕ_{21}	Difference between 2nd and 1st phase derived using the Fourier decomposition	
ϕ_{31}	Difference between 3rd and 1st phase derived using the Fourier decomposition	
γ_1	Skewness	
γ_2	Kurtosis	
K	Stetson K index	
Q_{3-1}	3rd quartile (75%) – 1st quartile (25%)	
A	A ratio of magnitudes brighter or fainter than the average	In this work
H_1	Amplitude derived using the Fourier decomposition	In this work
W	Shapiro-Wilk normality test statistics	In this work
m_{p10}	10% percentile of slopes of a phase-folded light curve	Long et al. (2012)
m_{p90}	90% percentile of slopes of a phase-folded light curve	Long et al. (2012)

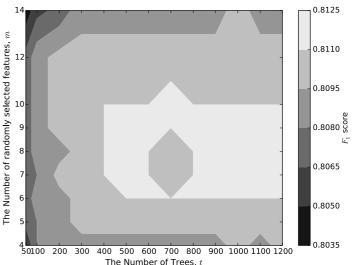**Notes.** [a] From Kim et al. (2014) if not specified.

Fig. 1. Feature importance as estimated using the random forest algorithm. As expected, period is the most powerful feature for separating periodic variables, although the other 15 features are non unimportant.

Fig. 2. F_1 scores of the trained random forest models with different t and m. Note that the F_1 scores are quite uniform over the parameter space, showing differences of less than 0.01.

To extract variability features from the training-set light curves, we used EROS-2 blue-band, B_E, light curves. For OGLE, we used I-band light curves. EROS-2 B_E-band and OGLE I-band light curves generally have more data points and/or better photometric accuracy than EROS-2 red-band, R_E, and OGLE V-band light curves, respectively.

2.3. Training a classification model

A random forest classifier (Hastie et al. 2009; Breiman 2001) is based on an ensemble of decision trees (Quinlan 1993). Each decision tree is trained using a randomly selected subset of the features at each node. The random forest uses a majority voting strategy to assign probabilities to all labels (here the variability classes) for a given sample (here the light curve). We accept the variability class corresponding to the highest probability among those assigned probabilities. See Kim et al. (2014) for a more detailed description of the random forest algorithm.

To train a random forest model, two hyperparameters must be set. These are the number of trees, t, and the number of

randomly selected features at each node of trees, m. Each tree is fully grown without pruning. Before tuning the parameters, we split the training set into two subsets, T_1 and T_2, each of which contains 50% of each class in the training set. We used T_1 to optimize the hyperparameters by performing a brute-force search over a two-dimensional grid in t and m, using ten-fold cross-validation. The metric for model performance is the F_1 score, which is the harmonic mean of the precision and recall, i.e.,

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad \text{where} \qquad (6)$$

$$\text{precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}, \qquad (7)$$

$$\text{recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}, \qquad (8)$$

and F_1 varies between 0 and 1. This is computed for each class separately, and the individual scores then averaged weighted by the number of objects in that class. Figure 2 shows the results of the grid search. The difference between the best and worst

Table 4. Classification quality of the trained model with subclasses.

Superclass	Subclass	Precision	Recall	F_1
DSCT		0.93	0.89	0.91
RRL				
	ab	0.99	0.99	0.99
	c	0.92	0.96	0.94
	d	0.94	0.83	0.88
	e	0.90	0.91	0.91
CEPH				
	F	0.98	0.98	0.98
	1O	0.93	0.92	0.92
	Other	0.88	0.76	0.81
EB				
	EC	0.76	0.68	0.72
	ED	0.90	0.91	0.90
	ESD	0.80	0.74	0.77
LPV				
	Mira AGB C	0.83	0.91	0.87
	Mira AGB O	0.76	0.74	0.75
	OSARG AGB	0.66	0.65	0.65
	OSARG RGB	0.74	0.80	0.77
	SRV AGB C	0.70	0.60	0.65
	SRV AGB O	0.70	0.67	0.69
T2CEPH		0.91	0.49	0.64
NonVar		0.87	0.89	0.88
Average[a]		0.81	0.81	0.81

Notes. [a] Weighted average by the number of source in a class.

models is 0.009, so the exact choice of t and m is unimportant. The classification uncertainty derived during the ten-fold cross-validation is ±0.005, which is comparable to this difference.

With $t = 700$ and $m = 10$ (the best values) we trained a classification model using the subset T_1 and applied it to T_2 in order to assess the classification quality. This is summarized in Table 4. Recall and precision for DSCT, RRL, CEPH, and their subclasses are generally higher than 0.90, sometimes as high as 0.99. The majority of misclassification is caused by confusion within a superclass, especially within EB and LPV, as can be seen in Fig. 3. In addition, if we train[4] a classification model without the subclasses, the F_1 score increases to 0.98 as shown in Table 5.

Finally, we trained a classifier using all 16 variability features, the entire training set (i.e., $T_1 + T_2$), and the subclasses (again with $t = 700$ and $m = 10$). Note that this model is the UPSILoN classifier that we use to classify the MACHO, LINEAR, ASAS, and HIPPARCOS periodic variables presented in the following sections.

3. Validation of the classifier on the MACHO, LINEAR, and ASAS datasets

We applied the UPSILoN classifier to light curves found as periodic variables from three time-series surveys: MACHO (Alcock et al. 1996a); ASAS (Pojmanski 1997); LINEAR (Palaversa et al. 2013). These each have different survey characteristics, such as the number of data points, observation duration, limiting magnitude, etc., so represent a suitable datasets to assess the UPSILoN classifier's generalization performance. Before we applied the classifier to these three datasets, we refined each light curve by 3σ clipping, and by excluding light curves having spurious periods, as described in Sect. 2. We then extracted the

---

[4] The hyperparameter optimization is done separately.

16 variability features from each light curve and used UPSILoN to predict a variability class.

3.1. MACHO

We obtained MACHO B-band light curves of periodic variables from Kim et al. (2011). The total number of light curves is 870 (Table 6). The number of data points in each light curve is between 200 and 1500, while 50% of the light curves have more than 1000 data points. The duration of the light curves (i.e., observation duration) is about 7 yr. The limiting magnitude of the MACHO survey is $V \sim 21$.

In the left panel of Table 7 we show the classification performance achieved by UPSILoN. MACHO does not provide subclass classifications so thus we merged predicted subclasses into corresponding superclasses. As the table shows, the average F_1 score is 0.94, which confirms that the UPSILoN classifier is successful at classifying the MACHO periodic variables. Nevertheless, the recall of CEPH is only 0.75, meaning that 25% of CEPH were classified as another variability type. Almost all of these were classified as T2CEPH, which is not surprising since CEPH and T2CEPH are similar. Note that MACHO did not use the class T2CEPH, so it is quite possible that some of these sources are in fact of class T2CEPH. The recall for RRL is 0.85, i.e., 15% of true RRLs were misclassified. Many of these were classified as EB, and we visually confirmed that they indeed show EB-like variability (Fig. 4).
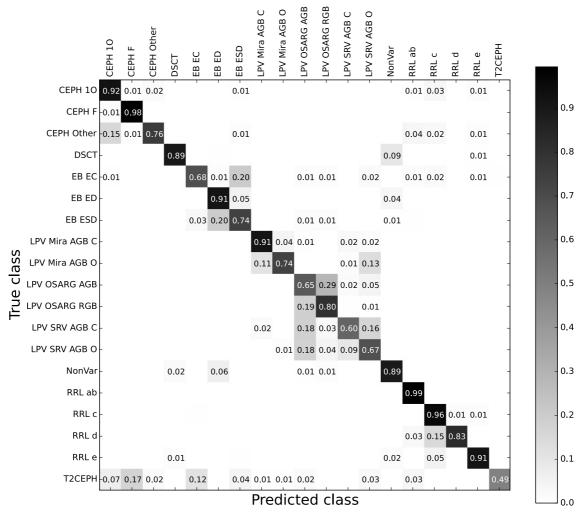
Table 7 also shows the classification performance using a random classifier with a training-set prior (middle panel), and an uniform prior (right panel). The term "random classifier" means that we randomly assigned a class to each MACHO light curve. The probability of assignment to a given class is proportional to the frequency of the class in the training set in the middle panel, and is equal for the classes in the right panel. We do these tests to get a measure of the minimum expected performance (which is not zero), and to quantify to what degree beyond "random" the data actually inform about the class. As expected, the precision and recall of both random classifiers are always considerably lower than that achieved by UPSILoN.

3.2. LINEAR

The light curves and catalog of LINEAR periodic variables are from Palaversa et al. (2013). We selected sources with the classification uncertainty flags of 1 and excluded others on the ground that their class assignment was ambiguous. The total number of sources after the selection is 5279 (Table 8). The number of data points of each light curve lies between 100 and 600, although most of the light curves have fewer than 300 data points. The duration of the light curves is about 5 yr and the faintest magnitude among these periodic variables is $r \sim 17$.

The left panel of Table 9 shows the UPSILoN classification performance. All classes except DSCT are successfully classified by UPSILoN. The low precision for DSCT occurs because 199 of the 1965 EBs were misclassified as DSCT, of which there are only 104 true ones. The reason for this is because 96% of the estimated periods of the LINEAR EBs are shorter than 0.25 days (Fig. 5), which is associated with the problem that the Lomb-Scargle algorithm typically returns half of the true period for EB-type variables in which the shapes of the primary and secondary eclipses are similar (see Graham et al. 2013 and references therein). Consequently, almost all of the LINEAR EB periods are within the period range for the DSCTs in the training

Fig. 3. Confusion matrix of the subclass model. Each cell shows the fraction of objects of that true class (rows) assigned to the predicted classes (columns) on a gray scale. Thus the values on the leading diagonal are the recall rate (Eq. (8)). We show the number only if it is larger than or equal to 0.01. See Table 4 for the numerical values of the recall and precision for each class.

Table 5. Classification quality of the trained model without subclasses.

Superclass	Precision	Recall	F_1
DSCT	0.94	0.87	0.90
RRL	0.99	0.99	0.99
CEPH	0.98	0.96	0.97
EB	0.97	0.96	0.96
LPV	0.99	1.00	1.00
T2CEPH	0.90	0.48	0.62
NonVar	0.87	0.87	0.87
Average[a]	0.98	0.98	0.98

Notes. [a] Weighted average by the number of source in a class.

Table 6. Number of objects per true class in the MACHO dataset.

Class	Number
RRL	270
CEPH	75
EB	189
LPV	336
Total	870

set (i.e., ≤0.25 days), which could cause the misclassification. Despite the overlapping ranges of the estimated periods between LINEAR EB and DSCT, 88% of EBs (i.e., 1 734 among 1 965) were correctly classified.

3.3. ASAS

We downloaded ASAS V-band light curves and the accompanying catalog (ASAS Catalog of Variable Stars version 1.1;

hereafter ACVS) from the ASAS website[5] (Pojmanski 1997). We selected only sources with a unique classification (e.g., DSCT), thereby excluding sources with multiple classifications (e.g., DSCT/RRC/EC). We also discarded MISC-type sources, which have unclear variability class. After this filtering, the number of remaining sources is 10 779 (Table 10). The number of data points in the light curves varies between 200 and 2000, while 60% of the light curves have more than 400 data points. The duration of the light curves is about 9 yr. The limiting magnitude of the ASAS survey is $V \sim 14$, which is substantially brighter than the limits for OGLE, EROS-2, MACHO, and LINEAR.

---

5 http://www.astrouw.edu.pl/asas/?page=download

Table 7. Classification quality of the MACHO dataset.

Superclass	Precision	Recall	F_1	Superclass	Precision	Recall	F_1	Superclass	Precision	Recall	F_1
RRL	1.00	0.85	0.92	RRL	0.27	0.16	0.20	RRL	0.34	0.24	0.28
CEPH	1.00	0.75	0.85	CEPH	0.03	0.01	0.02	CEPH	0.13	0.24	0.17
EB	0.90	0.97	0.94	EB	0.21	0.20	0.20	EB	0.24	0.16	0.20
LPV	0.99	0.98	0.99	LPV	0.39	0.49	0.44	LPV	0.35	0.29	0.32
Average	0.98	0.92	0.94	Average	0.28	0.28	0.28	Average	0.31	0.24	0.27

Notes. *Left*: the UPSILoN classifier. *Center*: random classifier with a training set prior. *Right*: random classifier with a uniform prior.
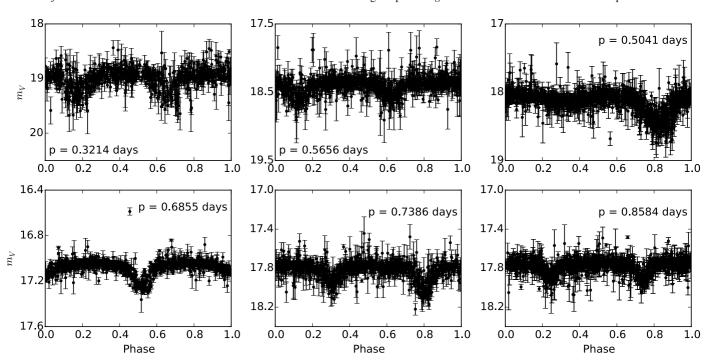

Fig. 4. Phased-folded light curves of the MACHO RRLs classified as EBs by the UPSILoN classifier. They show EB-like variability, implying that the MACHO classification might be incorrect. m_V is the Johnson *V*-band magnitude derived from the MACHO magnitude using the conversion equation from Kunder & Chaboyer (2008). In each panel we show the periods of the light curves obtained from the Lomb-Scargle algorithm. We used twice these periods to plot the phase-folded light curves, because Lomb-Scargle occasionally returns half of the true period for EB-like variability.

Table 8. Number of objects per true class in the LINEAR dataset.

Superclass	Subclass	Number
DSCT		104
RRL		
	ab	2415
	c	740
EB		1965
LPV		55
Total		5279

Table 11 (left panel) shows the UPSILoN classification performance. The average F_1 score is 0.85, demonstrating that most ASAS periodic variables are successfully classified. However, F_1 scores are relatively low for some classes such as RRL c and CEPH. This could be associated with the poor quality of the ACVS classification, which has been noted by previous studies also (Schmidt et al. 2009; Richards et al. 2012b). For instance, Schmidt et al. (2009) found that there exist serious biases in the CEPH classification of ACVS. Richards et al. (2012b) found the significant discrepancies between their classification and ACVS classification for RRL c and DSCT. We also find that our low

precision for DSCT is mainly caused by the UPSILoN's misclassification of EB EC (contact binaries) as DSCT. We visually checked the variability pattern of EB EC and DSCT classified by ACVS, and found that they are generally similar, as shown in Fig. 6. The low precision of EB ESD (semi-detached binary), 0.57, is likewise caused primarily by the confusion with EB EC, again due to their similar variability patterns. In contrast, the precision and recall of EB ED (detached binaries) are both fairly high: 0.93 and 0.96 respectively. This is because the variability pattern of EB ED – typically sharp flux gradients near its eclipses – is substantially different from that of EB EC or DSCT.

4. Investigation of classification performance in the presence of other variability types: the HIPPARCOS periodic variable stars
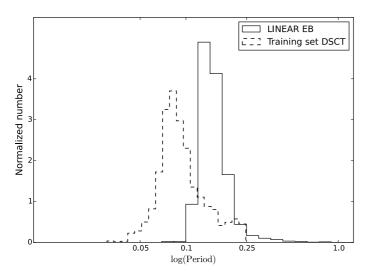
We obtained a catalog of HIPPARCOS periodic variable stars and their corresponding light curves[6] from Dubath et al. (2011). This dataset contains many variability types (e.g., rotational variability, eruptive variability) that are not in the UPSILoN training set, as well as types which are. Thus this dataset is suitable for

---
[6] Periodic variables in Table 1 from Dubath et al. (2011).

Table 9. Classification quality of the LINEAR dataset.

| Superclass	Subclass	Precision	Recall	F_1	Superclass	Subclass	Precision	Recall	F_1	Superclass	Subclass	Precision	Recall	F_1
DSCT		0.33	0.99	0.50	DSCT		0.03	0.03	0.03	DSCT		0.00	0.01	0.01
RRL					RRL					RRL				
	ab	0.99	0.91	0.95		ab	0.48	0.14	0.21		ab	0.48	0.06	0.10
	c	0.92	0.81	0.86		c	0.15	0.04	0.07		c	0.15	0.05	0.08
EB		0.97	0.88	0.93	EB		0.38	0.20	0.26	EB		0.37	0.16	0.22
LPV		0.95	1.00	0.97	LPV		0.01	0.49	0.02	LPV		0.01	0.29	0.02
Average		0.96	0.89	0.92	Average		0.38	0.15	0.20	Average		0.38	0.10	0.14
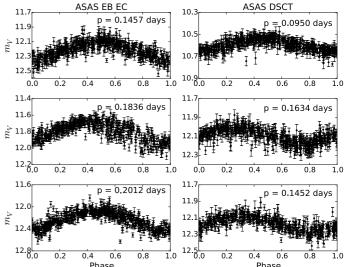
Notes. *Left*: the UPSILoN classifier. *Center*: random classifier with a training set prior. *Right*: random classifier with a uniform prior.

Fig. 5. Period histograms of EB (solid line) from the LINEAR dataset. LINEAR EB have relatively shorter periods, which could cause misclassification of them into the short-period variables such as DSCT. The dashed line is a period histogram of the training set DSCT.

Fig. 6. Phase-folded ASAS light curves of EB EC (*left*) and DSCT (*right*) from the ACVS classification. As the figure shows, there are no clear variability differences between them, which could yield misclassification. Periods (in days) of the light curves are given in each panel.

examining the potential contamination to the UPSILoN classification by unaccounted for variability types. We refined each HIPPARCOS light curve by 3σ clipping, and by excluding light curves with spurious periods (see Sect. 2) as we did for the MACHO, LINEAR, and ASAS datasets (see Sect. 3). Table 12 gives the number of each HIPPARCOS variability type in the dataset after this refinement. The table also shows the corresponding UPSILoN classes. We then extracted the 16 features to predict a variability class. The average number of data points in the light curves is about 100, and the duration of the light curves varies from one to three years, which is much fewer and shorter than the other three datasets.

In Fig. 7, we show the confusion matrix of the classification results. This is more informative than a table of recall/precision since many HIPPARCOS classes do not exist in the training set. As the matrix shows, the HIPPARCOS classes from DSCT to LPV (from top to middle row), which are those which correspond directly to UPSILoN classes, are relatively classified well by the UPSILoN classifier. The recall rates for those classes are generally high, up to 0.96, which indicates that UPSILoN successfully classifies the major types of periodic variables even though the number of data points is relatively small and the duration is relatively short. Nevertheless, we see the confusion between subclasses of eclipsing binaries, which is expected following what we showed ealrier for the UPSILoN classifier performance (see Fig. 3). About 40% of HIPPARCOS DCEPS (i.e., first overtone Cepheid) are misclassified as CEPH F (i.e., fundamental

Table 10. Number of objects per true class in the ASAS dataset.

Superclass	Subclass	Number
DSCT		520
RRL		
	ab	1218
	c	305
CEPH		
	F	567
	1O	179
EB		
	EC	2550
	ED	2167
	ESD	823
LPV		2450
Total		10 779

mode Cepheid). Dubath et al. (2011) also found a similar degree of confusion between these two classes and suggested that an overlap between the two classes in the period-amplitude space might be the reason. In addition, about half of the CEP(B) (i.e., multi-mode Cepheid) are classified as CEPH F rather than CEPH Other. The rest are classified as CEPH 1O (first overtone). In other words, all of the HIPPARCOS multi-mode Cepheid variables are misclassified. This confusion could be associated with their multi-period characteristics, because the UPSILoN classifier is

Table 11. Classification quality of the ASAS dataset.

Superclass	Subclass	Precision	Recall	F_1	Superclass	Subclass	Precision	Recall	F_1	Superclass	Subclass	Precision	Recall	F_1
DSCT		0.70	0.90	0.79	DSCT		0.04	0.02	0.02	DSCT		0.06	0.07	0.06
RRL					RRL					RRL				
	ab	0.92	0.94	0.93		ab	0.11	0.13	0.12		ab	0.11	0.05	0.07
	c	0.50	0.80	0.61		c	0.02	0.03	0.03		c	0.04	0.07	0.05
CEPH					CEPH					CEPH				
	F	0.94	0.54	0.69		F	0.04	0.02	0.02		F	0.06	0.06	0.06
	1O	0.78	0.53	0.63		1O	0.01	0.01	0.01		1O	0.01	0.04	0.02
EB					EB					EB				
	EC	0.91	0.67	0.77		EC	0.26	0.01	0.02		EC	0.22	0.06	0.09
	ED	0.93	0.96	0.94		ED	0.21	0.14	0.17		ED	0.18	0.05	0.08
	ESD	0.57	0.72	0.64		ESD	0.07	0.04	0.05		ESD	0.09	0.05	0.07
LPV		0.96	0.96	0.96	LPV		0.23	0.50	0.32	LPV		0.23	0.32	0.26
Average		0.88	0.84	0.85	Average		0.18	0.17	0.13	Average		0.17	0.11	0.12

Notes. *Left*: the UPSILoN classifier. *Center*: random classifier with a training set prior. *Right*: random classifier with a uniform prior.

Table 12. Number of objects per true class in the HIPPARCOS dataset.

Type	HIPPARCOS Class	Number	UPSILoN class
δ Scuti	DSCT	45	DSCT
low amplitude	DSCTC	80	DSCT
RR Lyrae	RRAB	71	RRL ab
	RRC	19	RRL c
δ Cepheid			
fundamental mode	DCEP	170	CEPH F
first overtone	DCEPS	28	CEPH 1O
multi-mode	CEP(B)	11	CEPH Other
Eclipsing binary			
detached	EA	222	EB ED
semi-detached	EB	246	EB ESD
contact	EW	99	EB EC
Long-period variable	LPV	254	LPV
Ellipsoidal	ELL	27	
Slowly pulsating B star	SPB	78	
α^2 Canum Venaticorum	ACV	75	
RS Canum Venaticorum	RS+BY	35	
+ BY Draconis			
γ Doradus	GDOR	25	
β Cephei	BCEP	29	
α Cygni	ACYG	17	
B emmission line star	BE+GCAS	12	
+ γ Cassiopeiae			
SX Arietis	SXARI	7	
W Virginis			
long-period (>8 days)	CWA	9	
short period (<8 days)	CWB	6	
RV Tauri	RV	5	
Total		1570	

trained using only a single period but not using multiple periods, as also suggested by Dubath et al. (2011). However this explanation might not clarify the confusion of this case because the recall/precision of CEPH Other is relatively high, as shown in Fig. 3. Thus we suspect that the training set might have an insufficient number of sources of the same variability type CEP(B). Most of the training sources in the CEPH Other class are double-mode Cepheids F/1O and 1O/2O, whereas we do not know

which pulsation modes these 11 CEP(B) stars are in. Further investigation is therefore not possible.

The other HIPPARCOS classes, from ELL to RV, contaminate the classification results, which is not very surprising because the UPSILoN training set that defines its known class space does not contain those variability types. The most contaminated UPSILoN classes are DSCT, CEPH F, CEPH 1O, and EB ESD whereas other UPSILoN classes are less contaminated
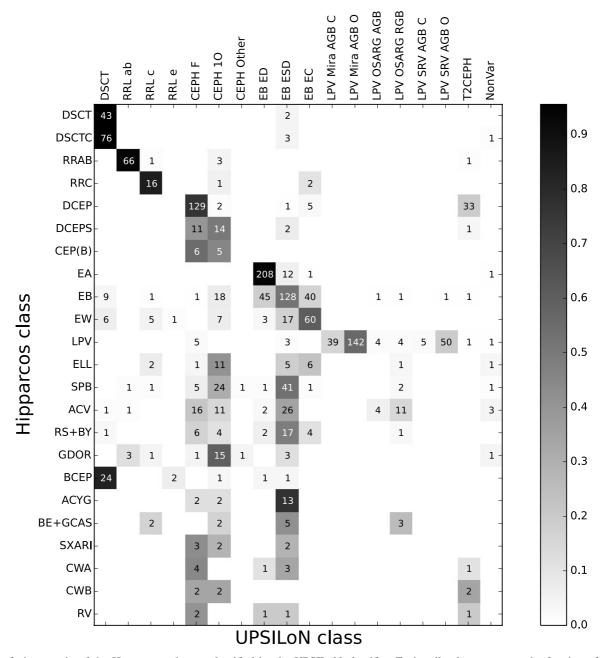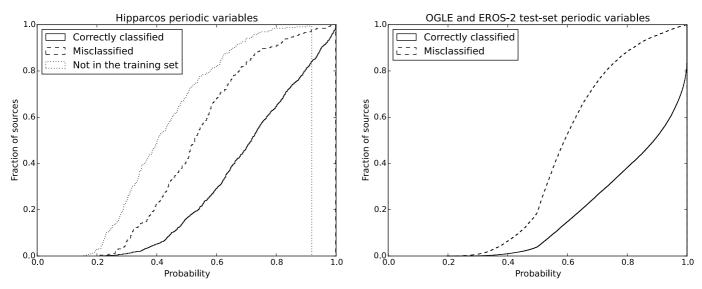
Fig. 7. Confusion matrix of the HIPPARCOS dataset classified by the UPSILoN classifier. Each cell-color represents the fraction of objects of that HIPPARCOS class (rows) assigned to the predicted UPSILoN classes (columns) on a gray scale. The value in each cell is the number of the HIPPARCOS sources.

or almost entirely uncontaminated by them (i.e., all RRL subclasses, CEPH Other, EB ED, EB EC, all LPV subclasses, and T2CEPH). As the table shows, 24 of the 29 BCEP are misclassified as DSCT due to BCEP's short periods and low amplitudes, which are also the characteristics of DSCT. In addition, contamination caused by SPB, ACV, RS+BY, GDOR, and ACYG to CEPH F, CEPH 1O and/or EB ESD is also high. For instance, 90% of SPB and 70% of ACV are classified as one of these UPSILoN classes. This might be because we did not use colors as a variability feature to train the UPSILoN classifier, which is one of the most useful features to separate different variability types. Training and validating an another model with survey-specific features (e.g., colors) is out of the scope of this paper. Nevertheless, most variable star surveys, being area and magnitude limited, will generally have a small fraction of variable stars of the rare "contaminating" classes considered here. Thus
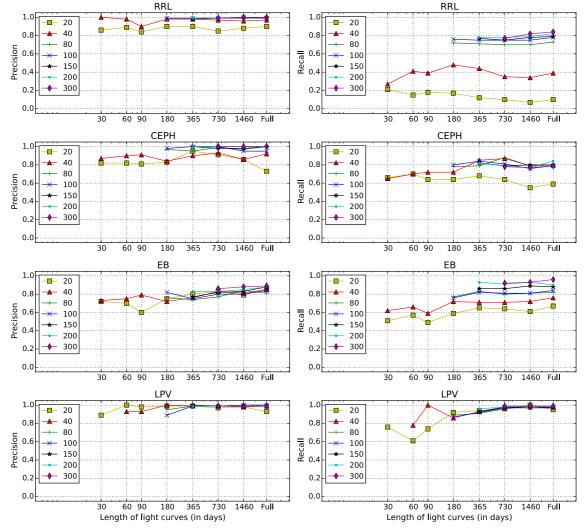
they would not be a significant source of contamination for the UPSILoN classes. Of course, UPSILoN will be unable to classify these rare classes, because it is not designed to do that.

Figure 8 shows the cumulative distribution of the assigned class probabilities for both the HIPPARCOS periodic variables (left panel) and the OGLE/EROS-2 test-set periodic variables (T_2 from Sect. 2.3, right panel)[7]. The solid line shows the sources whose variable types are in the training set and which are correctly classified by UPSILoN. The dashed line shows the sources of the same variability types but which are misclassified. As both panels show, the correctly classified sources tend to have higher probabilities than the misclassified sources. More importantly,

---

[7] Note that the probabilities for the OGLE/EROS-2 test-set were derived during the validation phase using the model trained on only T_1 (Sect. 2.3).

Fig. 8. Fraction of sources (*y*-axis) whose probabilities of the accepted classes are lower than a certain probability (*x*-axis) for the HIPPARCOS periodic variables (*left panel*) and the OGLE/EROS-2 validation test-set, T_2 (see Sect. 2.3), (*right panel*). The correctly classified sources (solid line) have generally higher probability than other two groups, each of which is 1) the misclassified sources (dashed line), and 2) the sources whose variability types are not in the training set (dotted line).

Fig. 9. Precision (*left column*) and recall (*right column*) of the resampled MACHO dataset as a function of length of light curves, *l* (horizontal axis), and the number of data points, *n* (different lines), for different variability classes (rows).
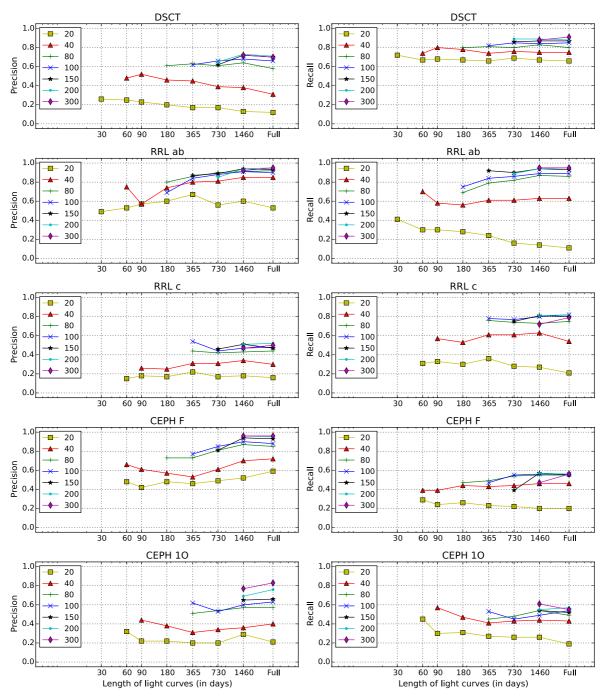
Fig. 10. As Fig. 9, but for the resampled ASAS light curves.

the sources whose variability types are not in the training set (dotted line) generally have lower probabilities than other two groups (solid line and dashed line). This is not surprising, because UPSILoN does not know the data space which these variability types occupy. The data are poorly explained by several classes, no class dominates, so UPSILoN assigns all of them low probabilities (as the probabilities must sum to one). Thus, in order to reduce the contamination, one can set a probability cut. For instance, excluding sources with probabilities lower than 0.5 removes about 70% of the contaminating sources and also ~40% of the misclassified sources, as can be seen in the left panel of Fig. 8. Such cuts, however, will reduce the recall rate at the same time (16% loss of the correctly classified sources in this case).

5. Investigation of classifier performance as a function of light curve length and sampling

In this section we present the results of experiments to assess the classification performance of the UPSILoN classifier for light curves having different numbers of data points and different observation durations. To achieve this we constructed a set of light curves by resampling the MACHO and ASAS light curves from the previous sections based on two parameters: the length of a light curve, l (i.e., observation duration); and the number of data points, n. We chose $l = 30$, 60, 90, 180, 365, 730, and 1460 days, and also the full observation duration (i.e., 7 yr for MACHO and 9 yr for ASAS). For the number of data points, we chose $n = 20$,
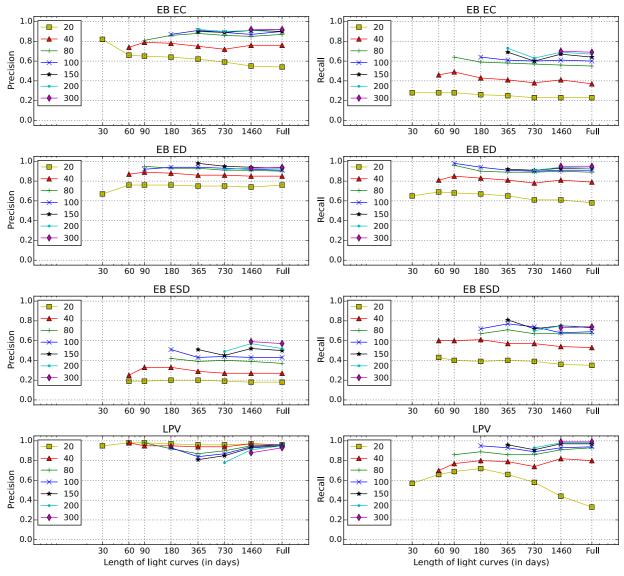
Fig. 10. continued.

40, 80, 100, 150, 200, and 300. The resampling procedures were as follows. For a given light curve we

1. extract measurements observed between the starting epoch of the light curve and l days after the starting epoch. If the number of the selected measurements is fewer than n, skip the current light curve and move to the next light curve.
2. randomly select n unique samples among these measurements.

These n unique samples give a resampled light curve of duration l. We did not resample the LINEAR and HIPPARCOS light curves because most of them have fewer than 300 data points.

We applied the UPSILoN classifier to these resampled MACHO and ASAS light curves. Figure 9 shows the precision and recall for each variability class for the MACHO data. Figure 10 show the same for ASAS. The horizontal-axis is the length of light curves, l, the vertical-axis is either precision or recall, and each line represents a different number of data points, n. In general, the more data points, the higher the precision and recall.

In the case of MACHO (Fig. 9), precision rapidly reaches its maximum value as the number of data points, n, reaches 80. Even when there are fewer data points (e.g., 20 and 40), precision is fairly high. Length, l, hardly affects precision. In contrast, recall is low for some classes (e.g., RRL) when $n = 20$ or 40. The reason for the low recall is because lots of light curves were classified as non-variables. This happens because the classification of variables with relatively short periods, such as RRL and CEPH, needs a sufficient number of data points over a short timescale. As the figure shows, the recall of RRL and CEPH slightly decreases as l increases when $n = 20$ or 40, for the same reason. In addition, to quantify EB-like variability, which has steep flux gradients around the eclipses, light curves have to be very well sampled to cover the area around the flux gradients. Therefore, recall for EB is also relatively low.

Figure 10 show the classification results of the resampled ASAS light curves. Precision and recall is generally lower than MACHO's because ACVS provides subclasses whereas MACHO does not. We saw earlier that recovering subclasses is a harder problem than recovering superclasses, as we would expect. The poor quality of the ACVS classification could also

contribute to the low recall and precision, as explained in the previous section. Nonetheless, when there are more than 40 data points, precision and recall are reasonably stable and rapidly reaches their maxima. We also found that recall and precision of most of the classes, namely DSCT, RRL ab, RRL c, CEPH, EB, and LPV, decreases as l increases when $n = 20$ or 40, as we saw from the resampled MACHO classification results.

From the experiments presented above, we see that if the number of data points, n, is larger than or equal to 80, precision and recall are not significantly lower than what we can achieve using more data points and longer duration light curves. More importantly, the duration of the light curves does not significantly alter the classification quality if $n \geq 80$.

6. Summary

The quality of any supervised classification model depends strongly on the quality of the training data, and yet constructing a clean and reasonably complete training set is quite difficult. This is particularly the case in the early stages of a survey, when only a few known objects will have been observed, making it hard to build a classifier.

In this work, we have developed a general purpose classifier for periodic variable stars to help alleviate this problem. First of all, it is known that periodic variables are relatively straightforward to separate from non-periodic variables or non-variables because of the clear periodic patterns of their light curves. Second, lots of well-sampled light curves from different surveys are available that can be used as a training set. As the characteristics of periodic variables can be captured by survey-independent features (e.g., periods, amplitudes, shape of phase-folded light curves), it is possible to build a universal training set.

We constructed a training set using the OGLE and EROS-2 periodic variables consisting of 143 923 well-sampled light curves for a number of different types of periodic variable stars. We mixed two different surveys having different characteristics (sampling rates, observation durations, optical bands, etc.) in order to train a classification model that is relatively tolerant of the diverse characteristics that could be encountered in the application datasets from other surveys. We extracted 16 variability features that are relatively survey-independent and used these features to train a random forest classifier, the UPSILoN classifier. Using periodic variables identified in the MACHO, LINEAR, and ASAS surveys, we have shown that UPSILoN can successfully classify periodic variable types in the training set with high precision and recall. Using the Hipparcos periodic variables, we found that the UPSILoN classes, CEPH F, CEPH 1O, and EB ESD, can be contaminated by other variability types that are not in the training set. We also tested the classifier using the resampled MACHO and ASAS light curves, where we found that precision and recall were fairly stable regardless of the time sampling and duration, as long as there was a sufficient number of data points (≥ 80). The UPSILoN classifier should prove useful for performing an initial classification into the major periodic variability classes of light curves coming from a broad set of optical monitoring surveys.

multi-core machine, Mekong, equipped with eight AMD Opteron 6276 CPUs and 264 GB memory, at the Max-Planck Institute for Astronomy.

References

Alcock, C., Allsman, R. A., Axelrod, T. S., et al. 1996a, ApJ, 461, 84
Alcock, C., Allsman, R. A., Axelrod, T. S., et al. 1996b, AJ, 111, 1146
Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411
Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, A&A, 559, A74
Blomme, J., Sarro, L. M., O'Donovan, F. T., et al. 2011, MNRAS, 418, 96
Breiman, L. 2001, Machine Learning, 45, 5
Breiman, L., Friedman, J., Stone, C., & Olshen, R. 1984, Classification and Regression Trees, The Wadsworth and Brooks-Cole statistics-probability series (Taylor & Francis)
Brown, T. M., & Gilliland, R. L. 1994, ARA&A, 32, 37
Carliles, S., Budavári, T., Heinis, S., Priebe, C., & Szalay, A. S. 2010, ApJ, 712, 511
Carretta, E., Gratton, R. G., Clementini, G., & Fusi Pecci, F. 2000, ApJ, 533, 215
Catalan, M. 2009, Astrophys. Space Sci., 320, 261
Debosscher, J., Sarro, L. M., López, M., et al. 2009, A&A, 506, 519
Dubath, P., Rimoldini, L., Süveges, M., et al. 2011, MNRAS, 414, 2602
Eyer, L., Evans, D. W., Mowlavi, N., et al. 2014, in EAS Pub. Ser., 67, 75
Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, ApJ, 553, 47
Graczyk, D., Soszyński, I., Poleski, R., et al. 2011, Acta Astron., 61, 103
Graham, M. J., Drake, A. J., Djorgovski, S. G., et al. 2013, MNRAS, 434, 3423
Hastie, T., Tibshirani, R., & Friedman, J. 2009, The elements of statistical learning: data mining, inference and prediction, 2nd edn. (Springer)
Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, ArXiv e-prints [arXiv:0805.2366]
Kim, D.-W., Protopapas, P., Byun, Y.-I., et al. 2011, ApJ, 735, 68
Kim, D.-W., Protopapas, P., Trichas, M., et al. 2012, ApJ, 747, 107
Kim, D.-W., Protopapas, P., Bailer-Jones, C. A. L., et al. 2014, A&A, 566, A43
Kunder, A., & Chaboyer, B. 2008, AJ, 136, 2441
Long, J. P., Karoui, N. E., Rice, J. A., Richards, J. W., & Bloom, J. S. 2012, PASP, 124, 280
Mackay, D. J. C. 2003, Information Theory, Inference and Learning Algorithms (Cambridge University Press)
Masci, F. J., Hoffman, D. I., Grillmair, C. J., & Cutri, R. M. 2014, AJ, 148, 21
Nun, I., Pichara, K., Protopapas, P., & Kim, D.-W. 2014, ApJ, 793, 23
O'Donovan, F. T., Charbonneau, D., Mandushev, G., et al. 2009, in NASA/IPAC/NExScI Star and Exoplanet Database, TrES Lyr1 Catalog, 6
Paegert, M., Stassun, K. G., & Burger, D. M. 2014, AJ, 148, 31
Palaversa, L., Ivezić, Ž., Eyer, L., et al. 2013, AJ, 146, 101
Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49
Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
Petersen, J. O. 1986, A&A, 170, 59
Pichara, K., Protopapas, P., Kim, D.-W., Marquette, J.-B., & Tisserand, P. 2012, MNRAS, 427, 1284
Pojmanski, G. 1997, Acta Astron., 47, 467
Poleski, R., Soszyński, I., Udalski, A., et al. 2010, Acta Astron., 60, 1
Quinlan, J. R. 1993, C4.5: programs for machine learning (San Francisco: Morgan Kaufmann Publishers Inc.)
Richards, J. W., Starr, D. L., Brink, H., et al. 2012a, ApJ, 744, 192
Richards, J. W., Starr, D. L., Miller, A. A., et al. 2012b, ApJS, 203, 32
Riess, A. G., Macri, L., Casertano, S., et al. 2011, ApJ, 730, 119
Saculinggan, M., & Balase, E. A. 2013, J. Phys.: Conf. Ser., 435, 012041
Schmidt, E. G., Hemen, B., Rogalla, D., & Thacker-Lynn, L. 2009, AJ, 137, 4598
Sesar, B. 2011, in RR Lyrae Stars, Metal-Poor Stars, and the Galaxy, ed. A. McWilliam, 135
Shapiro, S. S., & Wilk, M. B. 1965, Biometrika, 52, 591
Soszyński, I., Poleski, R., Udalski, A., et al. 2008a, Acta Astron., 58, 163
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2008b, Acta Astron., 58, 293
Soszyński, I., Udalski, A., Szymañski, M. K., et al. 2009a, Acta Astron., 59, 239
Soszyński, I., Udalski, A., Szymański, M. K., et al. 2009b, Acta Astron., 59, 1
Stokes, G. H., Evans, J. B., Viggh, H. E. M., Shelly, F. C., & Pearce, E. C. 2000, Icarus, 148, 21
Szymanski, M. K. 2005, Acta Astron., 55, 43
Tisserand, P., Le Guillou, L., Afonso, C., et al. 2007, A&A, 469, 387
Torres, G., Andersen, J., & Giménez, A. 2010, A&ARv, 18, 67
Udalski, A., Kubiak, M., & Szymanski, M. 1997, Acta Astron., 47, 319
Vivas, A. K., Zinn, R., Andrews, P., et al. 2001, ApJ, 554, L33
von Neumann, J. 1941, Ann. Math. Statist., 12, 367

Appendix A

A.1. How to use UPSILoN

The following pseudo-code shows how to use the UPSILoN library to extract features from light curves and then to classify them.

```python
# Import the UPSILoN library.
import upsilon

# Load a random forest model.
rf_model = upsilon.load_rf_model()

# Extract features and classify.
for light_curve in set_of_light_curves:

    # Read a light curve.
    date = np.array([:])
    mag = np.array([:])
    err = np.array([:])

    # Extract features.
    e_fts = upsilon.ExtractFeatures(date, mag, err)
    e_fts.run()
    features = e_fts.get_features()

    # Classify the light curve.
    label, probability, flag = upsilon.predict(rf_model, features)
```

`label` and `probability` are a predicted class and a class probability, respectively. The `flag` gives additional information of the classification result. For details of how to use the UPSILoN library, visit the GitHub repository[8]. UPSILoN is released under the MIT license, so it is free to use, copy, and modify.

If you use UPSILoN in scientific publication, we would appreciate citations to this paper.

A.2. UPSILoN runtime

We used the Macbook Pro 13″ equipped with 2.7 GHz Intel Core i5, 8 GB memory, and 256 GB SSD to measure the UPSILoN runtime[9] using the EROS-2 light curves, which contains about 500 data points. The average runtime for feature extraction per light curve is ~1.1 s. The most time-consuming part of the feature extraction is the period estimation. UPSILoN can reduce the period estimation runtime by utilizing multiple cores to speed up the Fourier transform part in the Lomb-Scargle algorithm. We observed that the period estimation runtime is inversely scaled with the number of cores. The runtime for class prediction per light curve is less than 0.01 s, which is negligible when compared to the feature extraction runtime. In addition, loading the UPSILoN random forest model file takes ~4 s. Note that the loading is not a recurring task.

---

[8] https://goo.gl/xmFO6Q
[9] Wall-clock time.