Automating Discovery and Classification of Transients and Variable Stars in the Synoptic Survey Era

Author(s): J. S. Bloom, J. W. Richards, P. E. Nugent, R. M. Quimby, M. M. Kasliwal, D. L. Starr, D. Poznanski, E. O. Ofek, S. B. Cenko, N. R. Butler, S. R. Kulkarni, A. Gal-Yam and N. Law

Reviewed work(s):
Source: *Publications of the Astronomical Society of the Pacific*, Vol. 124, No. 921 (November 2012), pp. 1175-1196
Published by: The University of Chicago Press on behalf of the Astronomical Society of the Pacific
Stable URL: http://www.jstor.org/stable/10.1086/668468
Accessed: 06/12/2012 10:51

Publications of the Astronomical Society of the Pacific, **124**:1175–1196, 2012 November

Automating Discovery and Classification of Transients and Variable Stars in the Synoptic Survey Era

J. S. Bloom,[1,2] J. W. Richards,[1,3] P. E. Nugent,[1,2] R. M. Quimby,[4,5] M. M. Kasliwal,[4,6] D. L. Starr,[1] D. Poznanski,[1,2,7] E. O. Ofek,[4,8] S. B. Cenko,[1] N. R. Butler,[1,9] S. R. Kulkarni,[4] A. Gal-Yam,[8] and N. Law[10]

Received 2011 June 27; accepted 2012 September 12; published 2012 October 24

ABSTRACT. The rate of image acquisition in modern synoptic imaging surveys has already begun to outpace the feasibility of keeping astronomers in the real-time discovery and classification loop. Here we present the inner workings of a framework, based on machine-learning algorithms, that captures expert training and ground-truth knowledge about the variable and transient sky to automate (1) the process of discovery on image differences, and (2) the generation of preliminary science-type classifications of discovered sources. Since follow-up resources for extracting novel science from fast-changing transients are precious, self-calibrating classification probabilities must be couched in terms of efficiencies for discovery and purity of the samples generated. We estimate the purity and efficiency in identifying real sources with a two-epoch image-difference discovery algorithm for the Palomar Transient Factory (PTF) survey. Once given a source discovery, using machine-learned classification trained on PTF data, we distinguish between transients and variable stars with a 3.8% overall error rate (with 1.7% errors for imaging within the Sloan Digital Sky Survey footprint). At >96% classification efficiency, the samples achieve 90% purity. Initial classifications are shown to rely primarily on context-based features, determined from the data itself and external archival databases. In the first year of autonomous operations of PTF, this discovery and classification framework led to several significant science results, from outbursting young stars to subluminous Type IIP supernovae to candidate tidal disruption events. We discuss future directions of this approach, including the possible roles of crowdsourcing and the scalability of machine learning to future surveys such as the Large Synoptic Survey Telescope (LSST).

Online material: color figures

1. INTRODUCTION

The arrival of the era of synoptic imaging surveys heralds the start of a new chapter in time-domain astrophysics, where the real-time processing of images taxes the capacity to transport the data from remote sites and pushes to the limit the computational capabilities at processing centers (e.g., Jurić & Ivezić 2011). More profoundly novel, however, is that the data volumes have begun to surpass what is possible to visually inspect by even large teams of astronomers and volunteer "citizen scientists." This necessitates an increasingly more central role for software and hardware frameworks to supplant the traditional roles of humans in the real-time loop.

This abstraction of people away from the logistics of the scientific process has been progressing rapidly, starting with the acquisition process itself. Indeed, robotic telescopes,[11] capable of taking data autonomously at remote sites, have become an increasingly common form of operation at the sub-meter- and meter-class level (cf. Castro-Tirado 2010). Many robotic systems use queuing algorithms that optimize nightly observing over several scientific programs and many are capable of being interrupted by external alerts to observe high-priority transients (e.g., Filippenko et al. 2001; Vestrand et al. 2002; Akerlof et al. 2003;

[1] Department of Astronomy, University of California, Berkeley, CA 94720-3411.
[2] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720.
[3] Department of Statistics, University of California, Berkeley, CA 94720-7450.
[4] Cahill Center for Astrophysics, California Institute of Technology, Pasadena, CA 91125.
[5] IPMU, University of Tokyo, Kashiwanoha 5-1-5, Kashiwa 277-8583, Japan.
[6] Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101.
[7] School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel.
[8] Department of Particle Physics and Astrophysics, Faculty of Physics, The Weizmann Institute of Science, Rehovot 76100, Israel.
[9] School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287-1404.
[10] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada.

[11] For a list of robotic telescopes currently operating, see http://www.uni-sw.gwdg.de/-hessman/MONET/links.html.

Cenko et al. 2006; Bloom et al. 2006; Saunders et al. 2008; Kubanek 2010). Data from such facilities can be automatically transported, processed, photometered, and ingested into databases without human intervention.

Since imaging data has spurious sources of noise and artifacts that can mimic real astrophysical sources, in the absence of watchful trained eyes on the images themselves, autonomous discovery of transients and variable stars on synoptic imaging surveys is a significant challenge. Threshold cuts on photometric quality, changes in apparent magnitudes, etc., are effective in discovering bona fide astrophysics sources (Drake et al. 2009; Sokołowski et al. 2010). However, multiparameter thresholding tends to be suboptimal because it treats each parameter derived from a given candidate as an independent variable when clearly there can be correlations between parameters. Matched filtering—looking for light curve trends that fit the scientific expectation from a certain class of variables (e.g., microlensing; Tomaney & Crotts 1996; Belokurov et al. 2003)—can be a very effective tool to discover new events, but other sorts of variables and transients are not easily recovered from that view of the dataset. Likewise, previous machine-learning based discovery (e.g., supernova discovery with the Supernova Factory; Bailey et al. 2007) have been optimized on domain-specific discovery, leaving aside the multitude of other variables not of direct interest to that particular project.

Discovery that a varying source is truly astrophysical does not mean that the origin of that variability is understood. Indeed, while it is tempting to conflate the process of discovery with classification, by making sequential the two decisions, different machineries can be brought to bear on each. The literature on autonomous classification, by various computational techniques, has been growing rapidly; indeed a wide range of machine-learning techniques have been applied to classification of large astronomical datasets (see Mahabal et al. 2008 and Bloom & Richards 2011 for review). Aside from domain-specific classification (microlensing and supernovae), most work concerns classification of variables stars on historical datasets *in retrospect*, when analysis is performed after most of the data have been collected and cleaned (e.g., Sarro et al. 2006; Debosscher et al. 2007; Richards et al. 2011; Willemsen & Eyer 2007; Sarro et al. 2009; Butler & Bloom 2011; Richards et al. 2012a).

We are interested in a related, but more urgent challenge: classification on streaming data, in which analysis is performed while the data are still being accumulated. At a logistical level, keeping up with classification (and discovery) assures that the survey producing the data can be continually informed of the progress, allowing the survey to change course midstream if scientifically warranted. But at a more fundamental level, the reason for real-time classification is that the vast majority of science conducted with time-variable objects, especially one-off transients, comes when more data are accumulated about the objects of interest. Enabling intelligent follow-up, then, becomes a main driver for rapid classification. Ultimately, one can view classification as a means to maximizing scientific return in a resource-limited environment.

Given this view of real-time classification, the advantages of a computational (rather than human-centric) approach become clear:

1. Machines, properly trained, are faster than humans at discovery and classification of individual candidates/events, allowing for operations at arbitrarily high data rates (limited only by computational resources).
2. The turn-around for well-informed follow-up can be almost instantaneous for computationally based discovery and classification. This allows for more efficient use of the suite of follow-up facilities. For example, observations on a small-aperture telescope can obtain the same signal-to-noise ratio of a fading transient as is obtainable on a large-aperture telescope observed after a longer delay.
3. Experimentation with new discovery and classification schema requires little more than rerunning new codes on existing data, whereas a change to human-based approaches requires additional labor-intensive work with people on a massive scale.
4. Machine-learned classification is reproducible and deterministic, whereas human-based classification is not.
5. The reproducibility allows for calibration of the uncertainties of classification probability statements, based on "ground-truth" results from the survey itself, with assurances that those classifications are sound as the survey proceeds.

Robust statements about the demographics of variability of different types requires well-calibrated discovery and classification. And this, in turn, suggests that a machine-based approach is also preferred. Ultimately, there may still be a vital role for humans in the real-time loop, such as serving as "tie-breakers" on ambiguous classifications or uncertain follow-up paths for a particular source (Gal-Yam et al. 2011), but our long-term view is that if a body of human-produced classification statements can be reproduced by machine-learned frameworks, those sorts of statements (during the full-scale production mode of a real synoptic survey) should not ultimately come from humans.

In this article, we describe a methodology and formalism for producing discoveries of astrophysical transients and variable stars using a machine-learned framework based on human expert-trained input (§ 2). We show how false-negative and false-positive rates can be calibrated with data from the survey itself. In § 3, we discuss a machine-based approach to autonomous classification based on feature sets derived from context and time-series data on individually discovered sources. In § 3.4 we show how a machine-learned model on Palomar Transient Factory (Rau et al. 2009; Law et al. 2010) data produces highly

reliable initial classifications.[12] We end with a discussion about the outstanding challenges and look to future incarnations that may be used on upcoming synoptic surveys.

2. DISCOVERY ON IMAGES

To identify new sources or brightness changes of known sources in synoptic imaging there are primarily two computational paths: catalog-based searches and imaging-differencing analysis. With the former, sources in each image are found and extracted into a database consisting of flux and position (and associated uncertainty) as well as ancillary metrics on individual detections (such as shape parameters and photometric quality flags). Time-variable sources are then found by cross matching detections on the sky and computing changes in brightness with time. With the latter, a deep reference image is constructed from several images of a portion of the sky, astrometrically aligned with and flux-scaled to an individual image, and subtracted from each individual image. The result is a "difference image" (e.g., Bond et al. 2001), in which objects are then found and extracted into a database. Since image differencing usually involves the expensive cross-convolution of two images, catalog-based searches are considered computationally faster than image-differencing. Catalog-based searches do well in the regime of large brightness changes and do not suffer from color-correlated misalignment effects due to differential chromatic refraction (Drake et al. 2009). However, in crowded fields (where the typical separation between objects is of order a few PSF distances) or in the presence of high-frequency spatial variations in the background (i.e., near galaxy positions), image-difference searches for variable sources excels.[13] For well-constructed reference images, photometric uncertainties of sources found in image differences can approach the statistical photon limit of an individual image (Wozniak 2000). Given the particular interest in finding variable stars in crowded fields and events in and around galaxies (supernovae, novae, and circumnuclear sources), especially while the sources are still faint and on the rise, the PTF collaboration chose to perform discovery on image differences. This is also the intended discovery path for most of the new upcoming synoptic surveys: Skymapper (Keller et al. 2007), Dark Energy Survey (Flaugher 2005), and the Large Synoptic Survey Telescope (LSST; Becker et al. 2005; Ivezić et al. 2008). The Catalina Real-Time Sky Survey (Drake et al. 2009) and the 3π survey of Pan-STARRS (Kaiser et al. 2002) conduct catalog-based searches for transients (cf. Gal-Yam & Mazzali 2011).

[12] To be sure, until mid-2012, an active group of citizen scientists enabled by the "Supernova Zoo" also offered an important discovery channel of supernovae (Smith et al. 2011) within the PTF collaboration that was largely separate from the autonomous discovery and classification framework described herein.

[13] If all detected objects are to be saved in each epoch, databases derived from image-differencing can be made vastly smaller, since only those sources which change are saved.

2.1. Identification

Frameworks for identifying and characterizing significantly detected objects (e.g., SExtractor; Bertin & Arnouts 1996) in images can be applied to image differences. One of the major drawbacks of discovery on image differences, however, is the number of spurious "candidate" objects that can arise from improperly reduced new images, edge effects on the reference or new image, misalignment of the images, improper flux scalings, incorrect PSF convolution, CCD array defects, and cosmic rays.[14] Even with signal-to-noise thresholds and some requirements on metrics related to the candidate shape (e.g., candidate FWHM compared with the image seeing), we have found that the vast majority of SExtracted objects on a given difference image are spurious: in PTF, only about 1 in 1000 (Brink et al. 2012) extracted candidate objects (considered to be at least as significant as a 5-σ detection) in a typical field are what we would deem to be astrophysically "real" (i.e., an origin owning to a change beyond the Earth's atmosphere). Nugent et al. (2011) provide details on the SExtractor extraction requirements and which candidate/subtraction parameters are saved into the real-time PTF database.

2.1.1. Real or Bogus?

Beyond the subtraction and source extraction steps, our first significant challenge is in determining which of the candidates are worth pursuing as real astrophysical events and which are "bogus." With training, many astronomers can identify when subtractions are poor or if a candidate is dubious to reasonable accuracy. But given the rate of candidate extractions, about 1–1.5 million per night for PTF, it is clearly not feasible to present candidates to human scanners to determine the reality of every candidate. To keep data volumes small enough to be human-scanned, several options are available. First, restrict the candidates to a certain domain-specific set. For example, scanning only those candidates that are near but offset from extended galaxies will generally succeed at finding some supernovae (and ignore most variable stars), but will fail to find supernovae far from their host galaxies, supernovae associated with low-luminosity hosts, and supernovae near the centers of galaxies (cf. Sullivan et al. 2011). There are active areas of research in all three of these cases (e.g., Miller et al. 2010). Second, require several candidates to appear at or near the same location on several epochs. This is indeed good at mitigating against cosmic rays and other transient artifacts, but missubtractions tend to correlate at the same locations even at different epochs (that is, when a subtraction is bad at some position on the sky at some epoch, there is an increased tendency for it to be bad at other epochs). This approach also runs the risk of waiting until too late to identify a (short-lived) astrophysical transient. Third,

[14] Of course, some of these effects are also present in catalog-based searches.

impose restrictive threshold cuts on the derived parameters of the candidate and the subtraction, such as requiring a 30-σ detection with a shape that is well-fit by the inferred PSF of the image. But, since most (real) candidates occur near the detection threshold and there is no guarantee that highly significant flux differences are all due to real astrophysical events, this approach will systematically exclude the lion's share of real events.

Our approach—to remove the human element in any real-time decision processes—is to use machine learning to provide a statistical statement about whether a given candidate should be considered astrophysically real or spuriously bogus. Such statements can then be combined over several epochs, if required, to determine if that identified candidate should be considered a *discovery* of an astrophysical source. To arrive at deterministic statements about each candidate, there are three broad classes of inputs that can be used to create a "labelled" set of candidates for use in the machine training: use trained/expert human scanners to opine on the real/bogus nature of a subset of the candidates, add a set of artificial sources to the raw data, or construct a ground-truth labelled set by using knowledge of which candidates turned out to real based on follow-up observations (e.g., using spectroscopically identified supernovae) of earlier incarnations of the survey.

Each labeling approach inheres advantages and drawbacks:

1. **Human-scanned**: Having humans provide the labels can ensure, by construction, that the machine-trained statements closely mimic what someone looking at a certain candidate might say about it. To fully capture the broad range of astrophysically real or spuriously bogus candidates, however, many (perhaps thousands) of candidates must be tediously labelled by hand. Moreover, there is no guarantee that a real source (especially near the detection threshold) will be labelled as such; and the converse is also true: Bogus candidates might be spuriously labeled as real even by experts.

2. **Artificial-source constructed**: Though computationally intensive, "fake" events can be placed at a variety of locations on the sky: in regions of high stellar density, near CCD chip edges, and at a variety of locations around a large diversity of galaxies. The main difficultly is in ensuring that the artificial candidates inserted into raw data are a close-enough representation of what a real source would look like in each image. That is, if all relevant effects (of the atmosphere, camera optics, telescope shake, etc.) are not properly modeled then there is a risk of a mismatch between what the derived parameters of the fake sources are and how real events are manifest in that parameter space.

3. **Ground-truth derived**: A ground-truth construction benefits from explicitly removing the vagueness and non-repeatability of human scanning but, in some cases, there remains an implicit reliance on human labels. For example, if spectroscopically identified supernovae are used to construct the "real" label set then there is a built-in bias towards spatial configurations that led previous observers to decide to follow-up such events. Further, if a catalog of known variable stars is used then there is bound to be a mismatch in survey characteristics; only bright variable stars, for instance, might be labelled as real. Determining bogus labels directly is difficult.

As there is no pure labeling process, we initially chose to use the human-scanned approach for the PTF data. (Brink et al. [2012], describes new efforts centered around the ground-truth approach). To facilitate the human labeling, we built a web-based system called "Group/think" based on the Python computing language[15] and the Google App Engine framework (Ciurana 2009; Fig. 1). During the commissioning phase of the project, several of the PTF collaboration members who had been hand-scanning each candidate every night were presented a series of images (each showing the reference image, the new image, and the subtraction) and asked to determine if the subtraction was "bogus" or "real," allowing them to assign a confidence level to their choice ranging from 0 (definitely bogus) to 1 (definitely real). The initial set of subtractions presented for human labeling were all made using R-band filter data. This set was constructed to include a mix of both supposed real candidates (drawn from confirmed transients) and bogus candidates. In particular, we used 74 real candidates associated with the first 11 spectroscopically confirmed supernovae discovered in the PTF commissioning. The 296 bogus candidates in the set were chosen to be the 4 nearest (but spatially distinct) candidates to the 74 supernova candidates. This initial set thus consisted of candidates that tended to be either obviously real or obviously bogus. A "realbogus" classifier was trained on the labels given by humans (see § 2.1.1) and applied to the first month of commissioning data. From that data, we created a new set of 574 candidates which spanned the range bogus to real, with a concentration of candidates intermediate to the two extremes.

So as not to bias the labeling to any one scanner, we determined the bias of each scanner relative to the group of scanners. Figure 2 shows the percentile distribution for each scanner relative to the other scanners for each candidate that scanner marked up. If all scanners for a given candidate gave the same realbogus value, then we assigned 50 percentile to every scanner. While most candidates show broad agreement, it is clear that some scanners were more or less optimistic in the aggregate than the group. Scanners 5–7 appear to believe fewer candidates are real and scanner #2 was more optimistic. For given scanners, their bias is determined from a mean of the percentile ranks of all their scanned candidates and an estimate of their 68% confidence scatter is determined using a Bayesian estimate, assuming a Jeffrys prior (Jeffreys 1946) for the standard deviation. Larger scatter indicates that the scanner agrees less often with the group. For every candidate, we create a realization of the debiased score for each scanner, adding it to a temporary list

[15] For more information please see http://python.org.

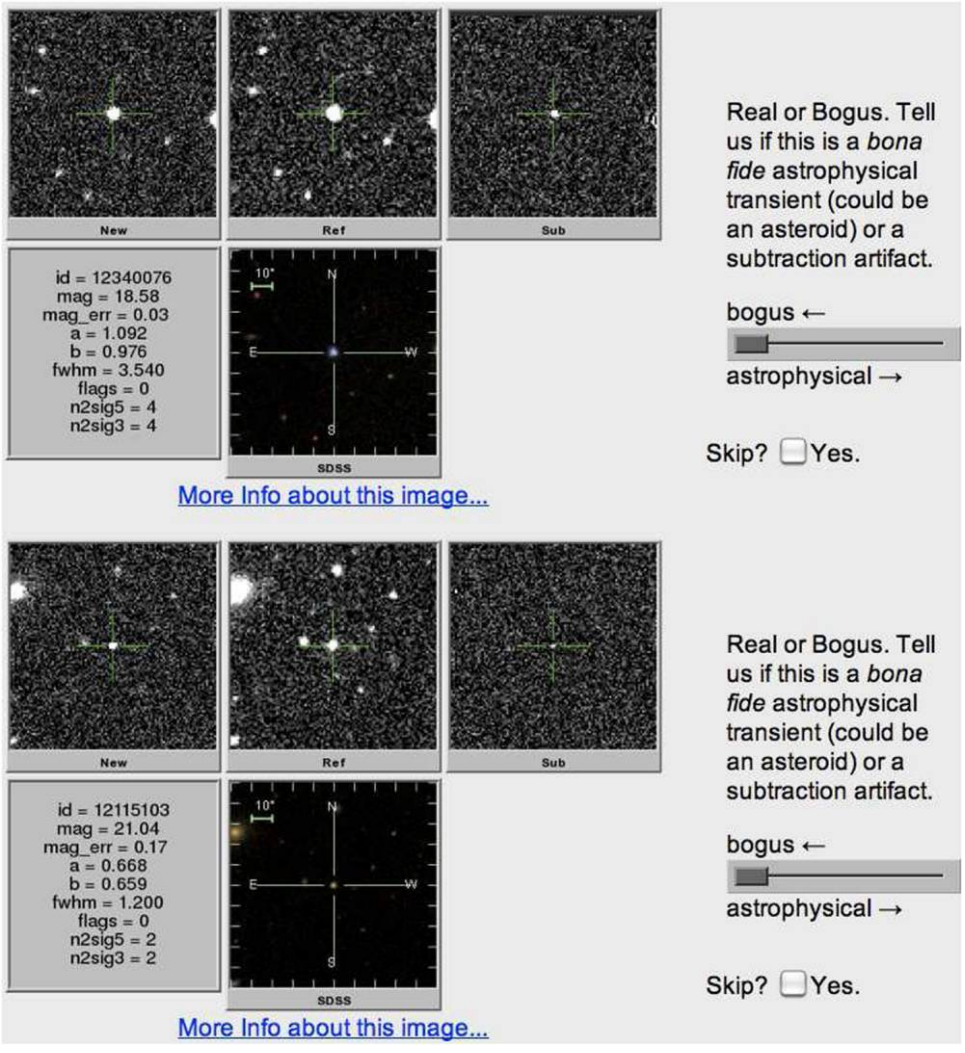
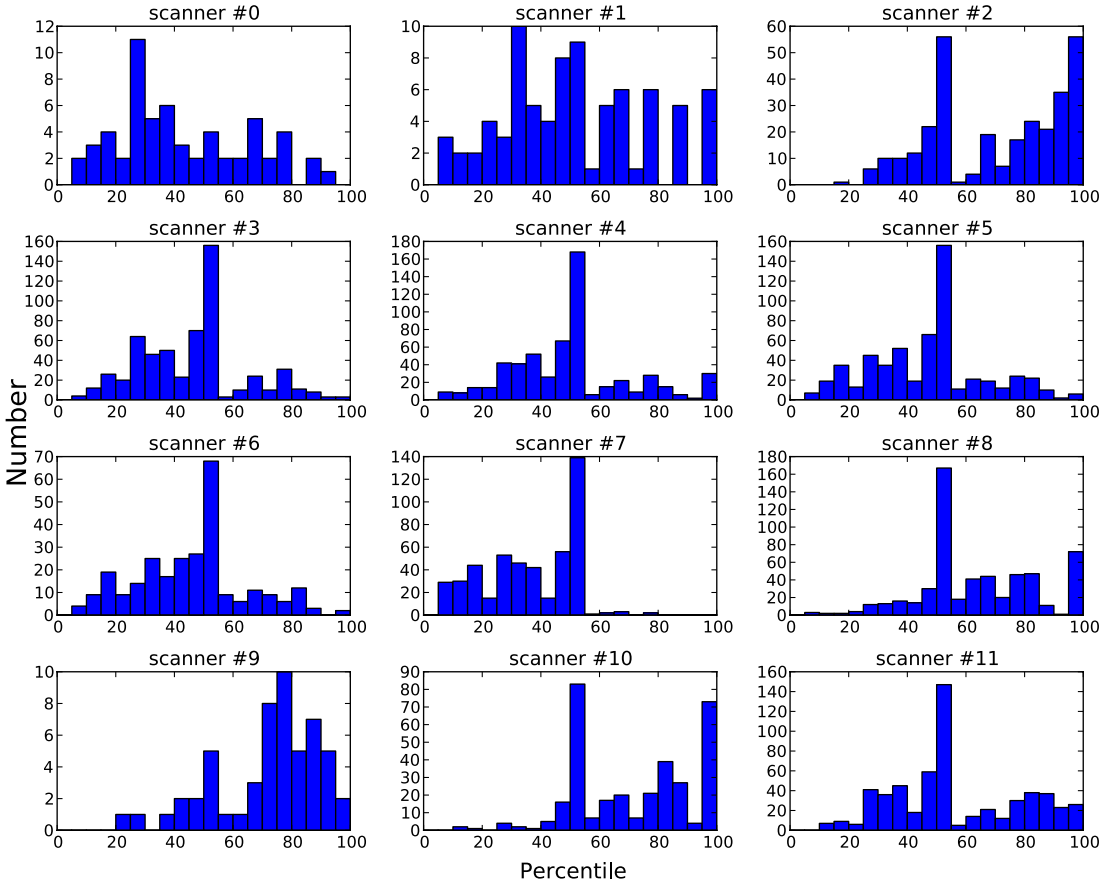


FIG. 1.—Example webpage showing two subtractions presented to human scanners: $1 \times 1'$ thumbnails of (*left to right*) the deep reference image, the new image, and the subtraction image. The *bottom panel* shows the SDSS image. Some metrics of the subtraction (such as the FWHM of the candidate source) are shown to the user to help them make a decision with more than just visual information. The responses were generated by a slidebar indicating the scanner's thoughts on whether the subtraction was bogus or astrophysical. See the electronic edition of the *PASP* for a color version of this figure.

if that scanner's standard deviation (s.d.) of percentile is less than a number chosen randomly from 0 to 100. Since the typical values of s.d. range from 15 to 25, approximately 80% of a scanner's biased score is used in a given realization. We take the median of 50 realizations of such lists. In this way, we create a scanner-weighted realbogus score for our labelled training sets. Figure 2 shows the distribution of the scanner-weighted realbogus score for the second labeling run of 574 candidates.

We wish to construct a parameter—generated rapidly at the time that the image differencing is completed—which reasonably mimics the human scanning decision of real or bogus. This necessitates the use of readily available metrics from our subtraction database on the candidates themselves used as input to train a machine-learned (ML) classifier (as opposed to some metrics which might be gleaned from external databases). For each candidate in the training sample, we derived 28 metrics (called "features" in ML parlance) from the SExtractor output (Table 1). Ill-derived (e.g., division by zero) or absent features were considered "missing" data for the purposes of the learning process. In this case, there are 255 missing values for `ellipticity_ref` (1.5% of all features) occurring if no reference source was detected. The scanner-weighted realbogus score for the training set was used as the ground-truth label for each candidate.

We found that the ML-regression techniques (e.g., M5P, Kohavi & Quinlan 2002) exposed in the Weka framework (Hall et al. 2009) were ill-suited to handle missing data and data with a mixture of numeric and nominal features. Instead, we created five nominal classes based upon the numeric scanner-weighted realbogus label: `bogus` (<0.10), `suspect` ([0.10,0.40)), `unclear` ([0.40,0.70)), `maybe` ([0.70,0.95)), `realish` (≥ 0.95).



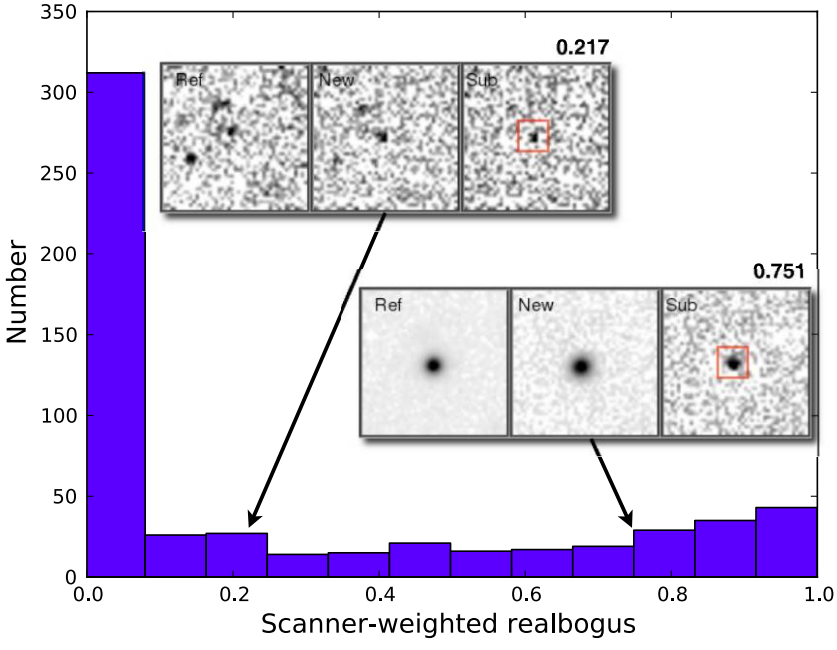


FIG. 2.—(*Top*) Distribution of training set scoring for 12 human scanners over 574 subtraction candidates. These distributions were used to compute the weights and biases for each scanner. (*Bottom*) Final scanner-weighted distribution of the training realbogus set. Examples of probable (*0.22*) and likely (*0.75*) subtraction candidates are shown. See the electronic edition of the *PASP* for a color version of this figure.

TABLE 1
REALBOGUS FEATURES

Feature name	Type	Description
mag	Numeric	USNO-B1.0 derived magnitude of the candidate on the difference image
mag_err	Numeric	Estimated uncertainty on mag
a_image	Numeric	Semi-major axis of the candidate[a]
b_image	Numeric	Semi-minor axis of the candidate[a]
fwhm	Numeric	Full-width at half maximum of the candidate
flag	Numeric	Numerical representation of the SExtractor extraction flags[a]
mag_ref	Numeric	Magnitude of the nearest object in the reference image if less than 5″ from the candidate
mag_ref_err	Numeric	Estimated uncertainty on mag_ref
a_ref	Numeric	Semi-major axis of the reference source[a]
b_ref	Numeric	Semi-minor axis of the reference source[a]
n2sig3	Numeric	Number of at least negative 2 σ pixels in a 5×5 box centered on the candidate
n3sig3	Numeric	Number of at least negative 3 σ pixels in a 5×5 box centered on the candidate
n2sig5	Numeric	Number of at least negative 2 σ pixels in a 7×7 box centered on the candidate
n3sig5	Numeric	Number of at least negative 3 σ pixels in a 7×7 box centered on the candidate
nmask	Numeric	Number of masked (suspect) pixels within a 5×5 box centered on the candidate
flux_ratio	Numeric	Ratio of the aperture flux of the candidate relative to the aperture flux of the reference source
ellipticity	Numeric	Ellipticity of the candidate using a_image and b_image
ellipticity_ref	Numeric	Ellipticity of the reference source using a_ref and b_ref
nn_dist_renorm	Numeric	Distance in arcseconds from the candidate to reference source
magdiff	Numeric	When a reference source is found nearby, the difference between the candidate magnitude and the reference source. Else, the difference between the candidate magnitude and the limiting magnitude of the image
maglim	Nominal	True if there is no nearby reference source, False otherwise.
sigflux	Numeric	Significance of the detection, the PSF flux divided by the estimated uncertainty in the PSF flux
seeing_ratio	Numeric	Ratio of the FWHM of the seeing on the new image to the FWHM of the seeing on the reference image
mag_from_limit	Numeric	Limiting magnitude minus the candidate magnitude
normalized_fwhm	Numeric	Ratio of the FWHM of the candidate to the seeing in the new image
normalized_fwhm_ref	Numeric	Ratio of the FWHM of the reference source to the seeing in the reference image
good_cand_density	Numeric	Ratio of the number of candidates in that subtraction to the total usable area on that array
min_distance_to_edge_in_new	Numeric	Distance in pixels to the nearest edge of the array on the new image

[a] Bertin & Arnouts (1996).

Using Weka, we trained a random forest classifier (Breiman 2001), using 10-fold cross validation, on the labelled data and developed a "cost" matrix to penalize gross misclassifications and to mitigate the effect of having many more bogus candidates than reals in the training sample. The random forest classifier operates by constructing an ensemble of classification decision trees, and subsequently averaging the result. The key to the good performance of random forest is that its component trees are *decorrelated* by sub-selecting a small random number of features as splitting candidates in each non-terminal node of the tree. As a result, the average of the decorrelated trees has highly decreased variance over each single tree. Missing data are replaced by an imputation step in the Weka RandomForest package with the median (mode) of numerical values (categorical values). The classifier produces a probability $P_i(C_j)$ of the i-th candidate belonging to each of the $j = 5$ classes. The sum of $P_i(C_j)$ over all i is unity. The ML-trained realbogus value for the i-th candidate is constructed using:

$$\mathrm{RB}_i = \sum_j P_i(C_j) \times w_j, \tag{1}$$

where the class weights for C_j = [bogus, suspect, unclear, maybe, realish] were set, ad hoc, to be $w_j = [0.0, 0.15, 0.25, 0.50, 1.0]$. The maximum value of $\mathrm{RB}_i = 1$ and the minimum is 0. A comparison between the scanner-weighted scores and the ML-trained realbogus score on those candidates is shown in Figure 3.

To evaluate the effectiveness of the classifier we constructed a "receiver operating characteristic" (ROC) curve (Fig. 4) showing the false-negative rate (FNR; real candidates set as bogus) versus the false-positive rate (FPR; bogus candidates selected as real) for a variety of different real/bogus cuts on the training data and the learned results. Since we do not know a priori the cutoff value for real vs. bogus candidates among scanner weighted scores, in Figure 4 we assume that every candidate with scanner-weighted scores greater than X (where X ranges between 0.05 and 0.5) is real. Then, we sequentially step through each ML-realbogus score ranging from $Y = 0$ to 1 to be the machine-learned cut off for making the real vs. bogus decision. Real candidates (as determined by X) that have a score $<Y$ are said to a

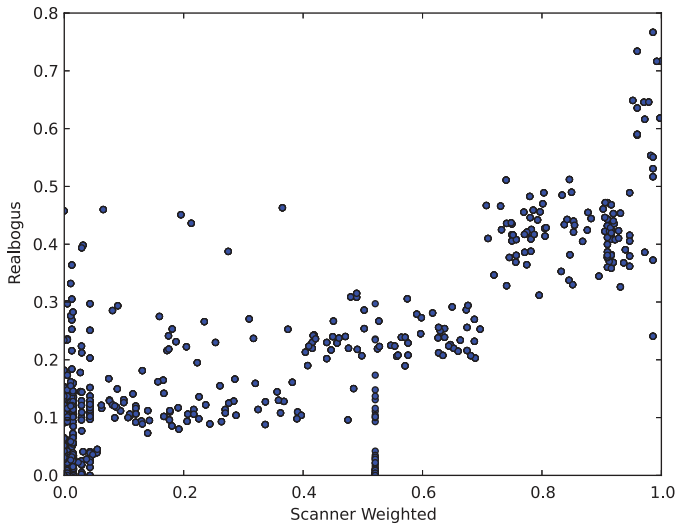


FIG. 3.—Comparison of the scanner-weighted realbogus score of the 574 candidates to the ML-trained/constructed realbogus score. There is some clear scatter and non-linearly in the relation; this is due to both an imperfect classifier and "label noise." However, the Pearson correlation coefficient is strong (*0.870*).

false negative. Likewise, bogus candidates with a score $>Y$ are said to be a false positive. At an ML-determined realbogus cut of 0.2 we expect an FPR between 0.08 and 0.12 and an FNR between ~0.02–0.2 (the range of uncertainty comes from an uncertainty in where the true cut should be for the scanner-weighted realbogus). At an ML-determined realbogus cut of 0.4 we expect an FPR between 0.01 and 0.02 and an FNR between 0.18 and 0.45. In § 2.2, we discuss how these ROC curves are used in the discovery process.

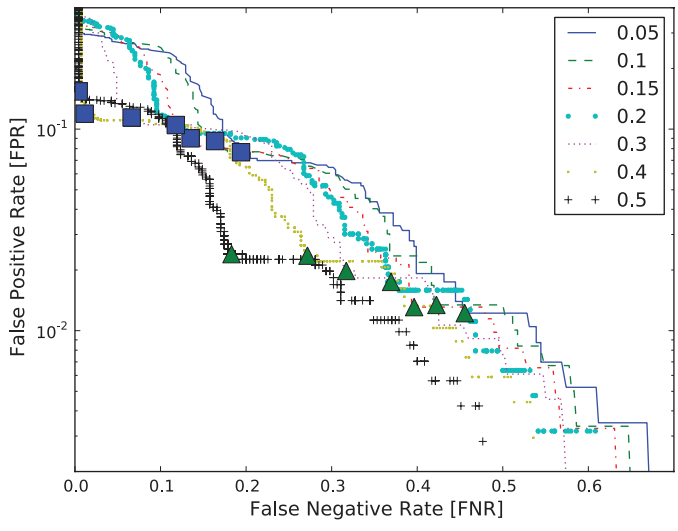


FIG. 4.—ROC curve for the trained realbogus sample as implemented for the Palomar Transient Factory. The seven curves were generated using a cut on the scanner-weighted RB scores (value shown in legend) in which all candidates with that cut value or larger were assumed to be definitely real and those below definitely bogus. The higher the value, the more conservative the human discovery threshold would be. Those candidates that were real but below the ML-determined realbogus cut value (for several cuts) were considered false negative (Type II error). Those candidates that were bogus but above the ML-determined realbogus cut were considered false positive (Type I error). The *blue squares* (*green triangles*) show the results for each curve assuming an 0.2 (*0.4*) ML-determined realbogus cut.

To validate the ML-classification we created a list of known asteroids passing through 4150 subtractions ($=2615$ deg^2) over three nights of data in fall 2009 (starting at JD $= 2455045.6648$). These data should be fairly typical, representing the diversity of fields and image quality in the survey: observations during these nights were not biased towards or away from the stagnant asteroid zone nor were they especially focused on imaging in the Galactic plane. The catalog positions and calculated magnitude of each asteroid were found for each subtraction, using a custom parallelized Python code (PyMPChecker) that made use of the Minor Planet Center asteroid data tables.[16] This code, which typically runs 10 times to 100 times faster than queries to the Minor Planet Center site, is made available by us for the community as an open webservice.[17] We identified 19,954 asteroids within the subtraction footprint. We created a subsample of those with good ($<$10–15″) a priori location accuracy from the catalog, bright enough to have been detected (i.e., the catalog magnitude at least as bright as the limiting magnitude of the image), and which were not close to the edges of the arrays (position $>$30″ from the nearest edge). Further, so as not to identify candidates associated with elongated asteroid observations, we restricted the sample to asteroids calculated to have a proper motion at the time of observation of less than 50 arcsec/hr, resulting in less than 0.83″ of total motion during the 1 minute exposure. There were 9034 asteroid-associated candidates in this subsample. Figure 5 shows the distribution of asteroids relative to the nearest candidates on the sky.

Figure 6 shows a validation of the ML-classified realbogus on these candidates. Nominally all these candidates are taken to be bona fide "sources," providing a ground-truth set for us to test the ML-classifier. In practice, however, near the faint end of the distribution there will be some pollution of this set with bad-subtraction candidates: if a known (faint) asteroid happens to be near a poorly-subtracted region, that candidate will be incorrectly included in the sample. There is a clear trend for brighter candidates to receive a higher realbogus value. There are many sources with realbogus around 0.35–0.50 that show no trend with magnitude; this locus reflects the distribution of the classifier output convolved with the weighting scheme (eq. [1]). The line near realbogus = 0.2 (FNR $=$ 0.18) is in rough agreement with, but higher than, the FNR predicted (~0.11) from the training set shown in Figure 4 (*blue squares*). This difference might be in part explained by

[16] For more information please see http://minorplanetcenter.net/iau/mpc.html.
[17] For more information please see http://dotastro.org/PyMPC/webservice_readme.html.

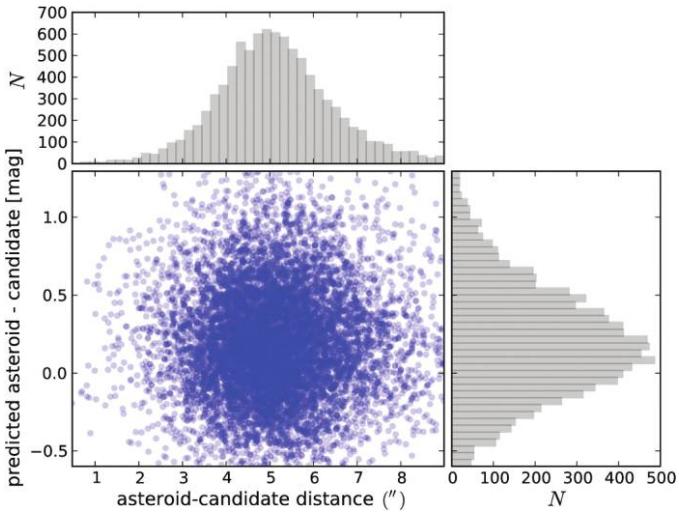


FIG. 5.—Distribution of the offset and magnitude differences of asteroids in the validation sample from the nearest-detected candidate in the PTF subtraction database. There is a clear locus of candidates from 2 to 8″ of the predicted position and within ~1.5 mag of the predicted brightness at the time of observation. The overall magnitude difference and scatter is expected given the PTF filter plus zeropointing uncertainties coupled with the approximate nature of Minor planet magnitude predictions. The positional offset is likely due to a combination of imprecise absolute astrometry on PTF images (improved since 2009) as well as the approximate nature of the orbit calculations in PyMPC: the code makes use of the orbital parameters downloaded from the Minor Planet Center that are updated only monthly, and do not include the most precise small-body gravitational perturbations. For the purposes of the creation of this validation set, the positional offsets are not important.

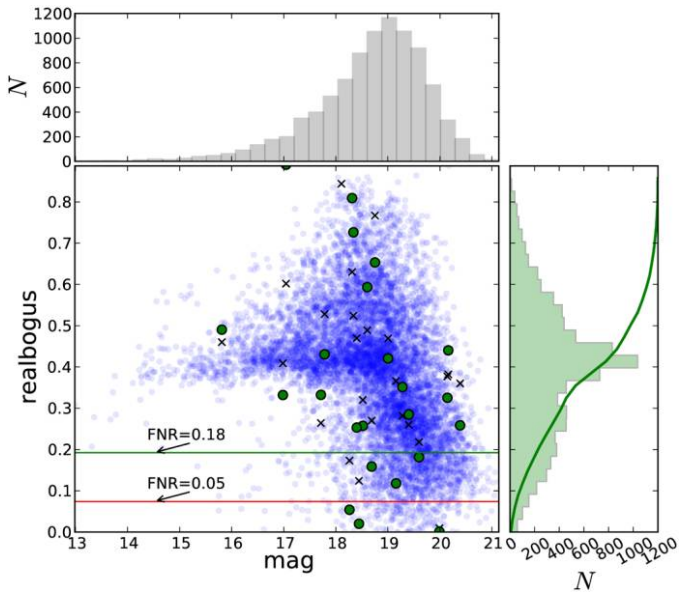


FIG. 6.—Distribution of asteroid-associated candidate ML-classified realbogus values. The cumulative distribution of realbogus values is shown as a green curve in the outset histogram at *right*. The *horizontal lines* show two effective false-negative rates (misidentified real candidates) for two different realbogus cuts. Crosses note the 24 asteroids in the subsample that are within 1″ of a source detected in the reference image. *Green circles* show the contextualized realbogus "score" for those candidates (§ 2.1.2). See the electronic edition of the *PASP* for a color version of this figure.

the inclusion of bad-subtraction candidates in the asteroid set. Since most asteroids are found far from stars and galaxies on the sky, there is a legitimate concern that this introspection is only validating the ability of the ML-classifier to identify spatially-isolated transients. However, by selecting the two dozen candidate asteroids that happen to be near (<1″) detected objects in the reference images (× symbols in Fig. 6) we find no clear trend of those sources to be preferentially different in their realbogus values. Figure 7 shows examples of asteroid-associated candidates near and far from reference objects.

2.1.2. Contextualized Statements

The metrics used in automatically classifying individual subtractions (Table 1) relate entirely to the candidate itself and not its surroundings (save the `good_cand_density` parameter). Candidates generated from poor subtractions—often owing to misregistration and/or to a poorly characterized convolution kernel—tend to cluster spatially. A high realbogus value on one candidate might be considered suspect if neighboring realbogus values are also high (under the reasonable assumption that significant variability is not common and should not be spatially correlated); the most egregious example would be when the misalignment of the new and reference images are more than a few times the scale of the seeing, leading most candidates on that subtraction to have high realbogus values.

This consideration calls for a contextualized statement of realbogus that takes into account what has happened both locally and globally on the subtraction. A simple scaled realbogus value is determined in the PTF pipeline by taking the ratio of the candidate realbogus to the mean of the nearest two candidates' realbogus values on that subtraction. A more complex scaled realbogus value takes into account all candidates in the subtraction frame, weighting more heavily those other candidates nearby to (and with similar magnitudes of) the candidate and the reference source themselves. We create a contextualized score with an ad hoc formula that takes into account the realbogus value itself and the two scaled versions. The formula itself is the multiplication of a set of logistic cumulative distribution functions (smoothly variable from 0 to 1 depending on each input parameter, such as realbogus). The score serves to downweight the candidates whose realbogus is not much higher than neighboring realbogus values. The score is also downweighted for sources very near diffraction spikes or bleeding trails near very bright stars (mag < 13). In PTF, scores are used to rank-order discoveries from most promising to least likely.

2.2. Discovery

If the unit of discovery—the moment of identifying an event as a true astrophysical source—was only a realbogus

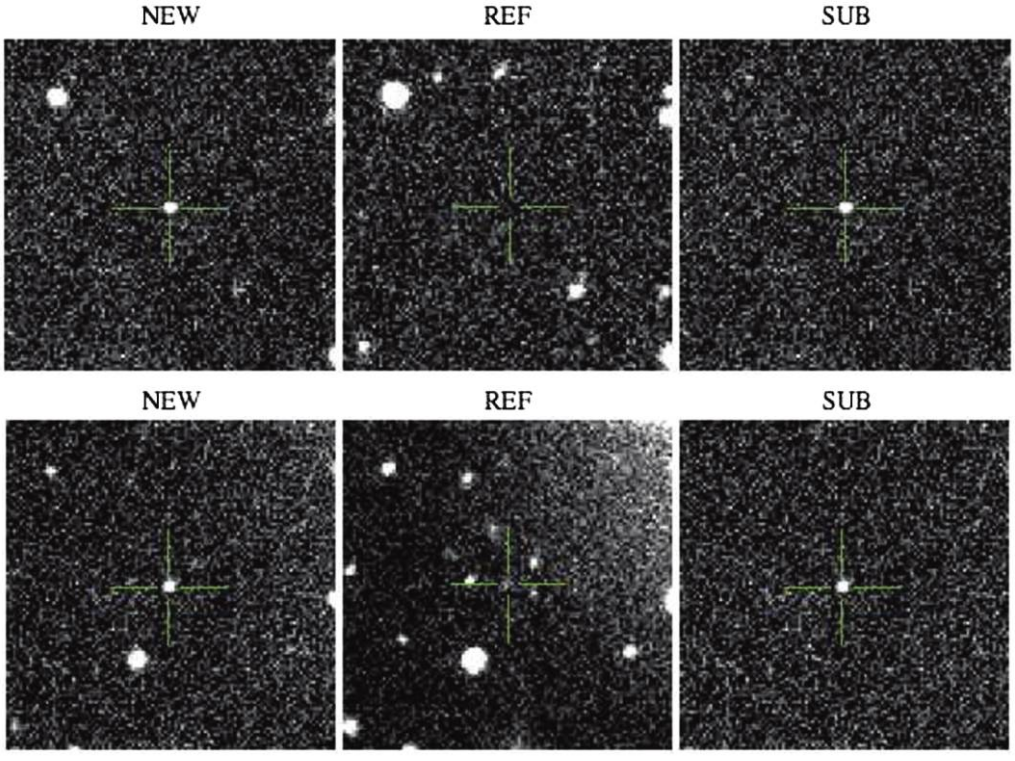


FIG. 7.—Subtractions showing examples of asteroids far from reference objects (*top*) and close (*bottom*; crosses in Fig. 6). At *top*, the candidate (#41844041) realbogus score is 0.43 and the minor planet is identified as (56152) 1999 CK106. At *bottom*, the candidate (#39899371) realbogus score is 0.63 and the asteroid is identified as (55162) 2001 QT238. A faint reference source near the asteroid centroid (*green crosshairs*) is seen in the *middle panel*.

statement (and associated score) about a single candidate, there would be enormous inefficiencies and impurities in PTF. Roughly 10% of all bogus candidates would be "discovered" and ~20% of all real sources would be missed (Figs. 4 and 6). Moreover, at the single-epoch sensitivity limit of PTF, there are at least as many as 10 times the number of slow-moving asteroids as stationary transients and variable stars, meaning most discoveries would be of asteroids and not the events and variables of interest. To mitigate against asteroid detection, PTF is generally scheduled to observe away from the stagnant asteroid zone and, more importantly, places a high priority in getting more than one image of the same field in a given night separated by at least 45 minutes–1 hr (Law et al. 2009; Rau et al. 2009). By requiring two reasonably good candidates to be coincident in space but separated by at least 45 minutes in time, we largely avoid asteroid "discovery" and can also build a higher degree of confidence in the astrophysical nature of the variability.

Since multiple candidates are required for discovery, the ROC curves for a single candidate are not the appropriate measure of efficiency and purity ($\mathcal{P}$) of discovery. We define purity as

$$\mathcal{P} = \frac{\mathcal{R}_{\rm dis}({\rm real})}{\mathcal{R}_{\rm dis}({\rm real}) + \mathcal{R}_{\rm dis}({\rm bogus})}, \qquad (2)$$

where the rate of discovery of real sources is:

$$\mathcal{R}_{\rm dis}({\rm real}) = \mathcal{R}({\rm real}) P({\rm discovery}|{\rm real}), \qquad (3)$$

and the rate of discovery of bogus sources is:

$$\mathcal{R}_{\rm dis}({\rm bogus}) = \mathcal{R}({\rm bogus}) P({\rm discovery}|{\rm bogus}). \qquad (4)$$

Note that $P({\rm discovery}|{\rm real})$ is just the *efficiency* of discovery. We expect in PTF (and other imaging surveys where detections are made on subtractions) that in any single subtraction $R({\rm bogus}) \gg R({\rm real})$. Roughly, in PTF, $R({\rm bogus}) \approx 1000\times R({\rm real})$. If discovery were done on just a single epoch then, $P({\rm discovery}|{\rm real}) = (1 - {\rm FNR})$ and $P({\rm discovery}|{\rm bogus}) = {\rm FPR}$. Following § 2.1.1, with ${\rm FNR} \approx 0.2$ and ${\rm FPR} \approx 0.1$ this implies $\mathcal{P} = 0.008$; this is unacceptably low. If we adopt a more conservative cut (Fig. 4), with ${\rm FNR} \approx 0.4$ and ${\rm FPR} \approx 0.01$, then $\mathcal{P} = 0.06$.

To keep $\mathcal{P}$ near unity (a high purity of discoveries to maximize followup resources), equation (2) requires that we create a detection classification scheme that satisfies $P({\rm discovery}|{\rm real}) \gg P({\rm discovery}|{\rm bogus})$. When multiple detections are required to cross a threshold for a discovery, then $P({\rm discovery})$ changes, and importantly, this probability changes differently for bogus events than real events. In the

simple case where two observations are made and two good detections (i.e., high realbogus) required, then

$$P(\text{discovery}|\text{real}) = P(\text{good detection}|\text{real}) \wedge P(\text{good detection}|\text{real}) = (1 - \text{FNR})^2. \quad (5)$$

This assumes that the probability of getting the same classification value is the same for both epochs, which might be nominally expected in the case of a source with approximately constant flux and similar observing conditions. For a bogus source to be called a discovery, however, two bogus subtraction candidates must be both incorrectly identified as real and occur close on the sky. In PTF, we have found that the existence of a bogus candidate is (unfortunately) highly correlated with the existence of another bogus candidate near the same place on the sky at different times: that is, certain places on the sky will preferentially yield bad subtractions (due to a combination of poor astrometry, imperfections in the reference image, and proximity to bright stars or chip edges). Ignoring correlations of realbogus values,[18] we expect with the two-candidate requirement $\mathcal{P} = 0.06$ and $\mathcal{P} = 0.78$ for $\text{FNR} = 0.2$ ($\text{FPR} = 0.1$) and $\text{FNR} = 0.4$ ($\text{FPR} = 0.01$), respectively. This means for the 2-candidate discovery process that at 78% purity, we are 36% efficient in finding real sources.

In practice, the source-discovery process in PTF is complicated by the fact that real source brightnesses are changing in time (and so too the respective realbogus values). We were also wary of missing faint (and low realbogus-valued) events occurring in nearby galaxies and so decided to err on the side of lower purity and higher efficiency. Since much of the science of PTF is focused on fast variables and short-lived transients (Rau et al. 2009), we also search for sources that are changing on relatively short timescales. Indeed, in the current incarnation of the framework, our initial query of the candidate database returns all candidates in a certain date range with realbogus greater than 0.17 (and with contextual realbogus greater than 3.3 times nearby sources; see § 2.1.2). The positions of these candidates are then cross-matched with other candidates with realbogus ≥ 0.07 within 2.0″ on the sky that were imaged at least 45 minutes (and no more than 6 days) before or after the candidate. PTF tries to obtain at least two epochs on the same part of the sky per night and repeat visits that part of the sky every 3–5 days. Given this we are reasonably assured that, if the source is real and still detected, at least one other candidate will be matched given the temporal criteria.[19] As a fail-safe against missing bright nearby supernovae, human-scanners were (for most of the PTF survey) presented candidates near large resolved galaxies with a much lower realbogus threshold (Smith et al. 2011).

Once a set of subtractions have finished loading (typically every 45 minutes for the 10^5 candidates in 100–200 square degrees of imaging) into the real-time subtractions database housed at Lawrence Berkeley National Laboratory (LBNL), an email with the date range of the subtractions is sent to an account which is then parsed automatically by a script running on the University of California, Berkeley campus. Depending on the density of stars in the field and the prior cadences in that part of the sky, typically 30–150 sources are identified. These sources, and associated candidates, are then saved as preliminary discoveries in an internal database of the automated system. The $\sim 10^5$ candidates generated per reduction run are typically vetted in 5 minutes via remote database queries.

3. CLASSIFICATION

Discovery inheres no more insight than the identification of a set of candidate events as belonging to a changing astrophysical source. The physical origin of the emission—the classification of the source into an established hierarchy of known variable and transient types—requires a different set of questions and another round of inspection now abstracted from the two-dimensional images. Indeed, once a source is preliminarily discovered, classification is done using only the data derivable from the LBNL databases and other (remote) webservice queries.

The PTF collaboration maintains a database of source discoveries, each assigned a unique name (such as "PTF 09dov"). During commissioning and during the start of the science operations of PTF, sources were discovered by human scanners who looked at individual candidates and associated candidates at other epochs. "Discovering," in that context, required that a button be clicked on a candidate scanning webpage. At the time of discovery, the scanner is also asked to suggest a crude classification choice, between variable star (`VarStar`), transient (`Transient`), and asteroid (`rock`). To mimic this interaction, removing the need for human scanning, one of the main roles of the automation is to provide the same set of initial classifications based on available data. As we now describe, the classification routines also try to provide more refined statements about the nature of the variability.

3.1. Features Based on Available Data

At a given place in the sky, there are broadly two categories of information available (in principle): the changes of brightness

[18] The positive correlation between bogus detections means that $P(\text{2nd detection}|\text{bogus, 1 detection}) > P(\text{detection}|\text{bogus})$, implying that $P(\text{discovery}|\text{bogus}) = P(\text{1st detection}|\text{bogus}) \times P(\text{2nd detection}|\text{bogus, 1 detection}) > \text{FPR}^2$.

[19] Clearly, when weather adversely affects observing over several nights real sources may go undiscovered because of this temporal windowing.

TABLE 2
TIME-DOMAIN FEATURES USED FOR OARICAL CLASSIFICATION

Feature name	Description
negatives	Number of candidates found in negative image differences associated with the source, That is, the number of epochs where the source was fainter than its reference brightness
positives	Number of candidates found in the image differences associated with the source
neg_pos_sub_ratio	Ratio of the number of negatives to all candidates (negatives+positives)
mag_scatter	RMS of the image difference magnitudes of positive candidates
mag_tot_scatter	RMS of the total aperture photometry of all candidates
max_cand_totalmag_diff	Maximum of the total-aperture magnitude minus the reference image source magnitude
diff_last_first_data	Difference in time (units of days) between the first and the last observation associated with the source
pm1	Apparent proper motion (arcsecond/hour) between the first and second epoch associated with the source
pm2	Apparent proper motion (arcsecond/hour) between the second and second-to-last observation of the source

in time as a function of wavelength and the context of where a source is located in relation to known objects (e.g., stars and galaxies) and coordinates (Supergalactic plane, ecliptic, etc.). Context information also includes the metrics on those nearby objects, such as color, apparent size, redshift, and spectroscopic type. To condense and homogenize all of the available information on a given transient or variable, like with image classification, we compute both context and time-domain features which may be used in decision rules or in a machine-learned classifier.

Since one the primary goals of the PTF collaboration is to rapidly identify new transient sources or extreme variable stars (e.g., Gal-Yam et al. 2011), we wanted to build a classification engine that was capable of making decisions with only a few epochs of imaging. To this end, we generated time-domain features that could have meaning in the limit of even a small number of epochs.[20] Those features are described in Table 2.

3.1.1. Context

With limited time-domain data available, it is clear that strong classification statements can be made based on context alone. A variable point source with quiescent colors in the SDSS bands of $0.7 < u - g < 1.35$ mag and $-0.15 < g - r < 0.4$ mag is very likely an RR Lyrae star (Sesar et al. 2010). A transient source near the outskirts of an intrinsically red galaxy is very likely a type Ia supernova. When a new discovery is made, in addition to computing the time-domain features, we make separate HTTP/GET external database queries to SDSS (DR7), USNO-B1.0, and SIMBAD. We also search a database of galaxies within 200 Mpc and record the projected offset of the source to the nearest galaxy. For all queries, information about nearby sources (and the distances to them) is saved in a database and associated with the newly discovered source. A subset of that information is converted into features for that source and becomes available to the classifier. Table 3 describes our context features. Some of the features are determined ad hoc (such as usno_host_type) based on experience with these catalogs. In a few cases, where the position is nearest (but not consistent with) the position of a star and consistent with a large SDSS galaxy, we will assign that galaxy as the host. In addition to usno_host_type, we also make a complex decision about the best "host" type using the SDSS and the local galaxy catalog. In particular, if SpecObjAll.specClass is "galaxy" or near_local_gal is "yes" or apparently_circumnuclear is "yes" then we set best_host_galaxy to "galaxy." If SpecObjAll.specClass is "qso" and the sdss_spec_warning does not contain "NOT_QSO" then we set best_host_galaxy to "qso." We set best_host_galaxy to "star" otherwise.

3.2. Oarical

The main purpose of the classifier, which we call Oarical, is to quickly label a newly-discovered source with as much specificity as possible and with as little time-series data as available. In particular, since the main science of the PTF collaboration focuses on transient/explosive events on short timescales, a particular premium was placed on the ability to recognize such events (i.e., supernovae, extragalactic "gap transients", novae, and Galactic outbursts). The workflow and major interfaces are diagrammed in Figure 8. The heavy reliance on context features (3.1.1) reflects the immediacy of the transient classification. The initial classification (Fig. 9) is separated into four groups: VarStar (variable star), SN/Nova (supernova or nova), AGN-cnSN-TDE (circumnuclear event, such as a tidal disruption flare, AGN/QSO activity or a circumnuclear supernova), and rock (asteroid). We produce an ordering of confidence of each classification for all discovered sources (what the most likely class is) and an overall scale of the confidence in the most likely class. If the discovery score of the source itself is low (near the realbogus discovery threshold) that scale will be low as well.

Oarical started routine operations on 2010 April 6 with the first robotic discovery and classification of PTF 10fhb (α(J2000): $10^h17^m00^s.30$, δ(J2000): $+45°30'48''.2$). It was classified by Oarical as "Transient." Spectroscopic followup

[20] There is a rich and growing literature that makes use of many epochs of high-quality photometry to produce robust classifications on variable stars, quasars, and supernova. See Bloom & Richards (2011) for review.

TABLE 3
CONTEXT FEATURES USED FOR OARICAL CLASSIFICATION

Feature name	Type	Description
		USNO-B1.0 based
usno_b	Numeric	B-band magnitude of the nearest source within 5″
usno_i	Numeric	I-band magnitude of the nearest source within 5″
usno_r	Numeric	R-band magnitude of the nearest source within 5″
usno_b_minus_r	Numeric	B-band minus R-band magnitude of the nearest source within 5″
usno_r_minus_i	Numeric	R-band minus I-band magnitude of the nearest source within 5″
usno_host_type	Nominal	Based on the average of the star/galaxy index ("s/g") USNO-B1.0[a]. Set to "galaxy" if $s/g < 3.8$, "star" if $s/g > 6.7$ and, otherwise, "uncertain"
		SDSS DR7 based
in_footprint	Nominal	Position is in the SDSS DR7 footprint ("yes" or "no")
dist_in_arcmin	Nominal	Distance in arcminutes of the source from the SDSS catalog position
dered_u_minus_g	Numeric	Dereddened u minus g magnitude of the nearest source
dered_g_minus_r	Numeric	Dereddened g minus r magnitude of the nearest source
dered_r_minus_i	Numeric	Dereddened r minus i magnitude of the nearest source
dered_i_minus_z	Numeric	Dereddened i minus z magnitude of the nearest source
chicago_class	Numeric	Galaxy principal component classification[b]
best_z	Numeric	Best redshift available: spectroscopic when SpecObjAll.zConf flag is >0.5 photoz2.photozcc2 when the r magnitude of the reference source >20, photoz2.photozd1 when the r magnitude of the reference source ≤20, photoz.z otherwise
best_z_err	Numeric	Uncertainty in the best_z
best_dm	Numeric	Distance modulus (mag) associated with the best_z
best_offset_in_kpc	Numeric	Projected physical offset in kpc from dist_in_arcmin and best_z
first_flux_in_mJy	Numeric	21 cm flux in mJy based on a cross-match with the FIRST survey
rosat_cps	Numeric	Counts per second of the cross-matched source in the ROSAT All-Sky Survey
sdss_spectral_stellar_type	Nominal	Spectroscopic classification (sppParam.sptypea)[c]
sdss_spec_warning	List of nominal	Spectroscopic flags related to classification[d]
		PTF and local galaxy catalog based
nn_dist	Numeric	Distance of the nearest source in the reference image in arcseconds (if <10″), and unknown otherwise
nn_kpc	Numeric	Distance of the nearest source in the reference image in kpc (if nn_dist <10″ and bestz >0.0001), and unknown otherwise
near_local_gal	Nominal	Is within 10 kpc or 3 Petrosian radii of a galaxy in the 200 Mpc sample?
apparently_circumnuclear	Nominal	Is the source consistent with occurring at the center of a local universe galaxy?

[a] See http://www.usno.navy.mil/USNO/astrometry/optical-IR-prod/icas/icas-usno-b1-format.
[b] From the sppParams table of SDSS. See Yip et al. (2004).
[c] See http://www.sdss.org/dr7/products/spectra/spectroparameters.html.
[d] See http://cas.sdss.org/astrodr7/en/help/browser/enum.asp?n=SpeczWarning.

of PTF 10fhb with the Double Spectrograph on the Palomar 200 inch Telescope on 12 April 2010 revealed it to indeed be Transient: a Type Ia supernova near maximum light at redshift $z = 0.1329$. During each night, after each subtraction run has completed (usually every 30–45 minutes), Oarical operates on the candidates from that subtraction run with discoveries noted in an internal database (following § 2.2); high-scoring sources are saved automatically as PTF-named events in the "PTF Marshal." The PTF Marshal is a database housed at the California Institute of Technology (Caltech), which serves as the official central repository for discoveries, followup, and collaboration interaction over PTF sources. Initial classification, following the prescription below, are also saved into the PTF marshal. Given the complex decision process used by the PTF collaboration in determining which discovered sources are followed-up spectroscopically (Gal-Yam et al. 2011), we do not have an unbiased view of the success rate and error rates in the Oarical classification (see below).

During the first year of the PTF survey, one of the main challenges in getting Oarical to produce reliable classifications was the lack of sufficient PTF data and ground-truth sources to train a machine-based classifier (we discuss this further in § 4). As such we built and refined a tree-based classifier to match our own *expectations* of classification based on a series of decisions

TABLE 4

OARICAL DISCOVERY AND CLASSIFICATION STATISTICS

PTF Type...robotclass	Oarical[a] discovery	Human[b] discovery	Oarical-only[c] discovery	Human[d] rediscovery	Oarical[e] rediscovery	Human different[f] type
VarStar	8322	2806	5516	13	2793	184
… CV	271					
… Periodic	3081					
Transient	6246	1938	4308	269	1669	852
… AGN-cnSN-TDE	2295					
… QSO	1059					
… SN/Nova	2427					

[a] Total number of autonomous discoveries and identification of PTF type.
[b] Total number of human-scanned discoveries and identification of PTF type.
[c] Total number of sources where Oarical was the only discoverer.
[d] Number of sources for which human-scanned discovery occurred after autonomous Oarical discovery.
[e] Number of sources for which autonomous Oarical discovery occurred after human-scanned discovery.
[f] Number of sources for which human-scanned PTF type differs from Oarical-determined type.

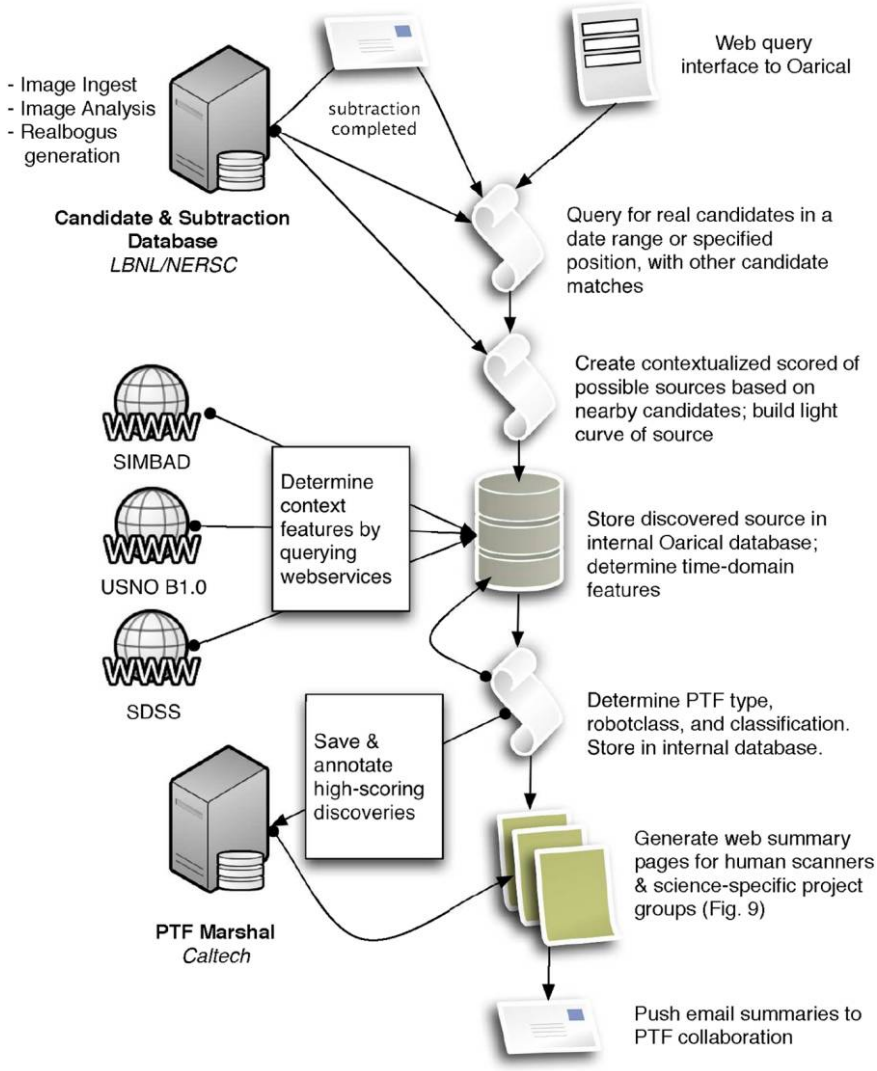


FIG. 8.—Flow diagram of Oarical, showing the major input and output components of the classification framework for PTF.

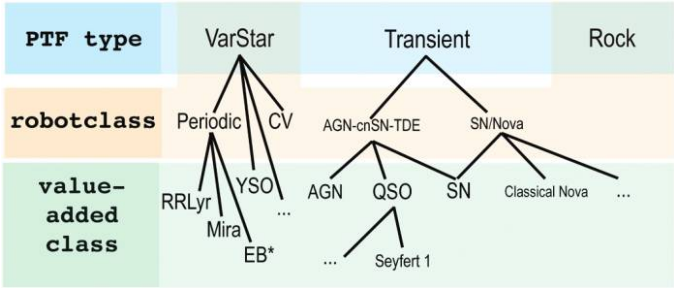


FIG. 9.—Taxonomy of classification used by Oarical. The *top bar* shows the PTF type, the initial classification used when saving candidates as sources. The *second tier*, "robotclass," shows the four classifications determined by Oarical for a new source. The *bottom tier* shows example classifications determined from SIMBAD identifications and SDSS spectroscopic analysis.

using the context and time-domain features. In this classifier, we rely on a hierarchy of input authorities, from most reliable to less reliable:

1. *Minor-Planet Center:* After the context and time-domain features are assembled, Oarical queries our parallelized minor-planet webservice to determine if the source is consistent in time and position with a known asteroid. If so, the source is classified as class `Rock` with high confidence and all other confidences are set to zero. If there is non-negligible proper motion (typically >0.1 arcsec per hr) and the ecliptic latitude is small ($b < 15°$), then the source is classified as likely class `Rock` (ad hoc, we ascribe a 90% confidence to this).

2. *SIMBAD:* About 8.6% of PTF-discovered sources cross-match with SIMBAD. Some of those types[21] are definitive statements about the class of variability (such as "EB*" for eclipsing binary star, "Mira," "BLLac," and "YSO"). Other SIMBAD types are useful in determining whether the source is galactic or extragalactic in nature ("GinGroup" for galaxy in group, "V*" for variable star). Some SIMBAD types are ambiguous (e.g., "Pec*" for peculiar star, "Radio," and "Blue"). For a source near (but not consistent with the center of) a SIMBAD-designated galaxy, we label the source `SN/Nova`.

3. *SDSS:* Spectroscopic redshifts (found in `best_z`) and galaxy/star separation (based on the PSF of the host) were used as reliable sources of the extragalactic/galactic nature of the PTF host source. We use the spectral typing (`sdss_spectral_stellar_type`) to determine the nature of extragalactic events (i.e., labels as `QSO` were taken as definitive). Hosts labelled as "star" but with X-ray or radio matches (`rosat_cps` and `first_flux_in_mJy`) were taken as likely QSOs.

4. *USNO-B1.0:* Host color, offset, and star/galaxy classification are used to make decisions about the extragalactic or galactic nature of the source. Astrometric coincidence with the centers of putative host galaxies are labelled as `AGN-cnSN-TDE`.

[21] See http://simbad.u-strasbg.fr/simbad/sim-display?data=otypes.

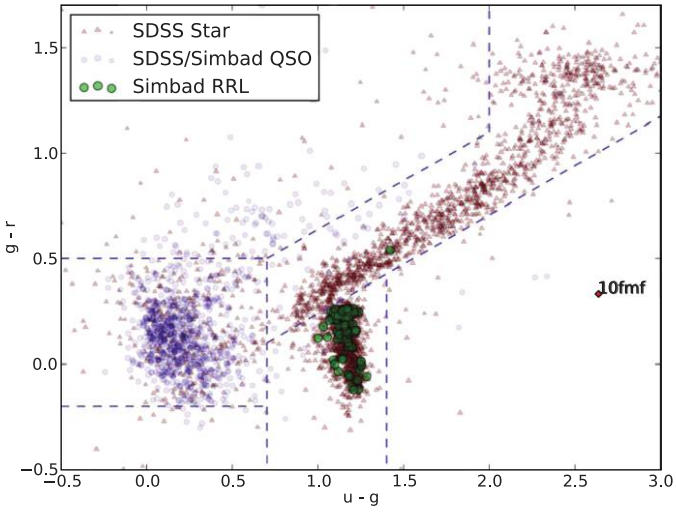


FIG. 10.—SDSS color-color diagram of the hosts (labelled in the SDSS photometric table as "star") of 3979 PTF sources observed until 2010 June. Oarical was used to type and classify (Fig. 9) these sources using SDSS and SIMBAD. The *dashed lines* show the regions traditionally used to classify sources (see Sesar et al. 2010). Known QSOs are shown in *blue circles*. Cataloged "stars" are shown in *red triangles*. Most of the "stars" in the QSOs locus are likely quasars without spectroscopy. Known RR Lyrae stars from SIMBAD are shown with *green circles*. PTF 10fmf is discussed in the text.

We used a hand-tuned aggregate weighting of all available authorities to produce a single set of confidence statements about the nature of the variables. Internal Oarical discoveries with high real-bogus (>0.3) and high classification confidence are saved automatically through the web interface of PTF Marshal, thus assigning an official PTF name and an initial type to the source. When more refined classifications are available (e.g., from SIMBAD or SDSS spectroscopy) that class is also annotated to the PTF databases as a value-added classification (Fig. 9).

Figure 10 shows the subset of the Oarical-classified PTF sources with a putative quiescent counterpart in SDSS. The stellar-, QSO-, and RR Lyrae-loci are seen and the density of variables is qualitatively similar to that seen in the Stripe 82 survey of variable sources (Ivezić et al. 2003)—that is, the relatively rare blue sources tend to be more significantly variable than red stars. There are 78 known RR Lyrae in this sample (from SIMBAD) with an additional 1502 sources matching the color locus of RR Lyrae suggested in Sesar et al. (2010)—of these, there are 8 known QSOs matching the RR Lyrae colors.[22] Since the locus of high-redshift quasars cuts across the RR Lyrae locus (Sesar et al. 2010) to larger $u - g$ color at roughly constant $g - r$, we decided to obtain a spectrum of one variable "star" (PTF 10fmf = SDSS J173630.59+642308.5; $u - g = 2.6$ and

[22] Two of these are misclassified spectroscopically and are indeed likely RR Lyrae.

$g - r = 0.33$) to ascertain whether PTF was indeed discovering high-redshift QSOs redward of the RRL locus. The spectrum taken with the Keck1+LRIS on 2010 June 14 UT of PTF 10fmf revealed a broad-line QSO at $z = 3.2$, making this source the highest redshift PTF-discovered transient.

As of 2011 June, there were roughly 40,000 sources discovered by Oarical and stored in the internal Oarical databases. There a total of 28,078 sources in the PTF Marshal database, with 20,355 discoveries or rediscoveries[23] since Oarical began running autonomously. Oarical accounted for 14,466 automatic discoveries or rediscoveries—that is, 29% of PTF sources were only discovered by human scanners while Oarical was running and about 36% of Oarical internal discoveries are saved automatically in the PTF Marshal without any humans in the loop. The other two-thirds of those Oarical sources that are not high-enough quality in score to warrant an automated discovery are presented to human scanners who decide on whether possible new events should be promoted to discovery (see § 3.3 and Fig. 11). We do not currently record when a human discovery was assisted directly by an Oarical-generated webpage—the majority of the 29% of human-scanner discovery likely originates from the Oarical-generated webpages of possible candidates. The Supernova Zoo accounts for the human-generated discovery of many nearby SNe (Smith et al. 2011).

Of the Oarical-discovered sources of PTF type VarStar, there were 79 spectroscopic observations recorded in the PTF Marshal (usually obtained after the Oarical discovery). Twenty two (28%) of these were spectroscopically typed to be supernovae—that is, incorrectly typed by Oarical. Interestingly, 14 of these had SDSS host identifications as "star" and almost all hosts appeared to be very large galaxies where the SDSS source classification broke up the large host into smaller subregions classified incorrectly as stars.[24]

Of the Oarical discovered sources of PTF type Transient, there were 645 sources with spectroscopic observations recorded in the PTF Marshal. Of these, there were 529 sources spectroscopically classified as supernovae, 43 were classified as variable stars, 23 identified as cataclysmic variables, 37 as some type of AGN, and the remaining 39 as unknown or uncertain (26 had more than one classification recorded that differed between categories). That is, about 7% of Oarical-discovered Transients were definitely incorrect. From the time that Oarical began until 2011 June, there were a total of 740 sources spectroscopically classified as supernovae of which 535 (72%) were discovered or rediscovered by Oarical. Table 4 provides a summary of Oarical classification and discovery.

[23] A rediscovery is when a scanner (human or robotic) saves a candidate into the PTF Marshal which is associated with a source already previously saved/discovered by the PTF collaboration.

[24] Improved sky-subtraction in SDSS may alleviate some of this problem in the future (Blanton et al. 2011).

3.3. Query Mechanisms

For all sources saved in the PTF Marshal databases, whether or not Oarical discovered the source, Oarical is run as an annotation service in near real-time. Information from SIMBAD and SDSS are saved as Comments in the PTF Marshal (Fig. 12) and available for users interested in a particular source to get a detailed set of metrics (if available) about that place in the sky. For instance, if a user saves a source as a VarStar but SDSS has a spectrum of that source, within about 15 minutes the Marshal will have annotations related to the SDSS spectrum (e.g., whether it is a quasar, what the spectroscopic redshift is determined to be, what the errors on redshift is, etc.). At the time that the source is annotated, any positional coincidence with known IAU Circular supernovae is also marked up in the PTF Marshal.

In addition to the automatic discovery of sources, Oarical provides webpage summaries of possible new sources from each reduction run. An email is sent to PTF subscribers about 30 minutes after the data are obtained allowing quick perusal of possible new sources; this allows humans to save sources which might not otherwise meet the thresholds for automatic discovery (see Fig. 11). A duty astronomer (primarily at the Weizmann Institute of Science) manually scans the Oarical discoveries and possible discoveries every day in near real-time and assigns followup priorities (Gal-Yam et al. 2011).

A webbased interface to Oarical is available to the PTF collaboration. This allows a PTF source, position, or candidate ID to be analyzed even if Oarical has not ingested that source into the databases. In addition, Oarical is automatically queried about once an hour for recently active sources that meet the criteria of certain science key projects. Fast transients (for example, changing by more than 0.5 mag in less than 3 hr) and tidal disruption candidates (circumnuclear events atop quiescent galaxies) have custom webviews autogenerated during the night based on these queries.

3.4. Machine-Learned Classification

With an eye to eventually replacing the manually tuned classification algorithm, we have explored the feasibility of using machine-learned classification for immediate PTF source classification. Using a sample of 1953 PTF sources with either spectroscopically-confirmed or SIMBAD-determined class, we train a random forest classifier (Breiman 2001) to predict class as a function of 43 different features. These features include 9 derived from the PTF light curves and 35 context features. To handle missing feature values—which arise due to incompleteness in the context features—we use the `missForest` imputation method of Stekhoven & Bühlmann (2011), which estimates the value of each missing feature via an iterative nonparametric approach to minimize imputation error. The `missForest` algorithm—available in the R package (R Development Core Team 2005)—was not available within Weka (the learning framework used for realbogus; § 2.1.1).

Candidate science class breakdown:

N(circumnuclear event)=14, N(varstar/galactic event)=56, N(SN/nova)=6

Click the header to sort by that column.
The best candidates should have high discovery score, high medscore, and high value (> 0.5) in one of the four major categories (ieg1, ieg2, irock, igal)
Candidate quality is color coded (green is a very high-quality candidate; red is a very questionable candidate)
If you're looking for only high quality candidates, look at the green and light green sources.

Transient/VarStar Candidates

Name	ID	Viz	RB	ieg1	ieg2	irock	igal	best class	oarical class (origin)	discovery score	medscore	mag	mag_ref	number of matches	LBL ID matches
PTF11bju	504167590 [jsb = 58861] Oarical...								sn (simbad)	0.517	0.640	18.53	16.83	67	503785019 502968286 502895695 502060910 501963955 500947714 500818735 500244003 500097846 499498023 and 57 more...
PTF09fga	504182281 [jsb = 6697] Oarical...							galactic	cv (sdss)	0.603	0.505	18.05	19.78	67	504069209 503929226 503205017 503105519 502367155 502228096 501223471 501042356 500626245 500372789 and 57 more...
PTF11hx	504133566 [jsb = 48094] Oarical...							galactic	cv (sdss)	0.367	0.189	18.07	16.60	78	502180021 500180395 495098400 492180297 489791594 489075158 488482252 487901902 479579686 479418256 and 68 more...
None	504173339 [jsb = 65545] Oarical...		0.186	0.536	0.116	-0.395	-0.484	SN/nova	sn (simbad)	0.160	0.003	18.44	13.65	23	502922522 496109430 494050807 488238060 488044811 483883139 483236374 482740452 477642237 474643822 and 13 more...
PTF11cei	504185367 [jsb = 61494] Oarical...		0.448	-1.289	-1.386	-2.173	1.642	varstar/galactic event*	rrl (sdss)	0.645	0.401	17.46	0.00	36	504073062 503937769 503243159 503132011 502379975 502239187 501341293 500623577 500368667 495375119 and 26 more...
PTF11bov	504133659 [jsb = 59870] Oarical...		0.406	2.224	0.445	-1.743	-1.962	SN/nova*	sn (simbad)	0.560	0.327	16.86	0.00	61	503911654 503813467 503115488 503009257 502178578 502043721 500943718 500816199 500180905 500003903 and 51 more...
None	504154106 [jsb = 65564] Oarical...		0.234	-0.766	-0.782	-0.759	0.896	varstar/galactic event	varstar (sdss)	0.195	0.001	19.29	16.92	51	500159847 498505222 495055971 493719403 491881648 489920029 489775317 489279340 489053216 488732027 and 41 more...
PTF11bka	504175076 [jsb = 11791] Oarical...		0.382	-0.954	-0.954	-1.575	1.148	varstar/galactic event*	rrl (sdss)	0.518	0.171	18.36	19.80	83	503882907 503829392 503155720 503033757 502231417 502076615 501065233 500885349 500252558 500108270 and 73 more...
[illegible]	504150767 [jsb =		[illegible]	-	-	-	[illegible]	varstar/galactic	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	500983547 500164544 500028984 490700361 484739273 484630610

new ref sub SDSS

FIG. 11.—Screenshot of a webpage generated for human-scanners to view possible candidates for discovery. Previously discovered sources are named following the PTF naming convention, all others are labelled as "None" (on this page there are two previously unknown sources). Color coding of each row shows the relative confidence in the source as a true astrophysical event. The sources with *blue asterisks* ("best class" column) are ones that Oarical has discovered. When the user mouses over the image thumbnail, a pop up of the subtraction is shown. About 20–30 such pages are generated nightly. Oarical-assisted human discoveries originate from these pages.

For the PTF Type classification problem, we have 1573 TRANSIENT and 380 VARSTAR sources.[25] Using features derived at the time of discovery, we obtain a 3.8% overall error rate (all error rates stated are found using 10-fold cross validation). For the 1422 sources with SDSS coverage, the error rate is 1.7%, while for the other 531 sources with no SDSS coverage the error rate jumps to 9.4%. In Figure 13 we plot the ROC curves for both variable star and transient source classification. The ROC curves show that at 90% purity, the random forest classifier

[25] There is some ambiguity in the initial typing scheme in the boundary between VARSTAR and TRANSIENT: cataclysmic variables (CVs), for instance, could be considered in either category. However, for definiteness, we put CVs in the VARSTAR category.

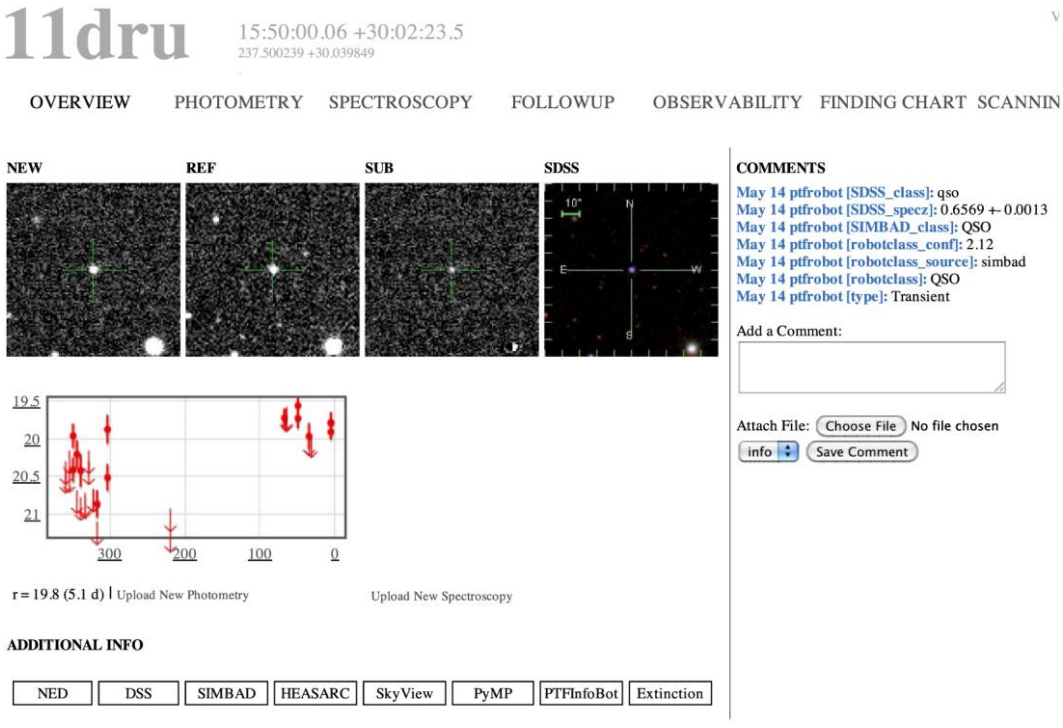


FIG. 12.—Screenshot of the PTF Marshal with automatic annotations from Oarical ("PTFROBOT"). See the electronic edition of the *PASP* for a color version of this figure.

attains 96.6% efficiency of variable star classification and 99.7% efficiency of transient classification. Notably, for SDSS sources, we achieve a 96.6% (100%) efficiency of VARSTAR (TRANSIENT) classification at a 90% purity level.

In robotclass classification, which divides the sources into five science classes, random forest obtains an error rate of 6.5%. Figure 14 shows that for the `AGN-cnSN-TDE`, `SN/Nova`, and `VarStar-Periodic` classes the classifier attains 97%, 93%, and 89% recovery, respectively. Due to a large class imbalance, performance of the classifier suffers for the smaller classes `VarStar-CV` and `VarStar-misc`. Again, our classifier performs significantly better for sources in SDSS, attaining a 3.7% error rate compared to 14.1% error for sources with no SDSS coverage. As more data are collected (post time of discovery), the robotclass random forest error rate decreases slightly: the error rate for objects without SDSS coverage drops to 13.2% after 30 days and 12.8% after 90 days, while the error rate for objects in SDSS does not change significantly with increased PTF observations. This implies that additional PTF data only helps in classification when no SDSS features are available.

Finally, the RF classification trees allow us to construct an estimate of the importance of each feature in the classifier. Using the prescription of Breiman (2001), we compute the importance of each feature as the increased number of sources that are correctly classified when using that feature instead of a replacement feature of random noise. In Figure 15 we plot the importance of each feature for each of the robotclass classes (`VarStar-misc` was omitted due to a scarce amount of data), and the average importance across all classes. Overall, the most important features are context based, while some light-curve-derived features (such as the ratio of the number of negative subtractions to positive subtractions) are important for distinguishing between certain classes. In the future, we may add more descriptive time-series features (such as those related to periodograms) which should also be useful in classification.

There are some biases in the sample generation that require a careful interpretation of these ML results. For a source to be included in the training sample via existing catalogs, it must have a SIMBAD label (e.g., "RRLyr*" or "QSO") that provides a "definitive" ground-truth statement about the nature of the variability. In some cases, that SIMBAD label comes from SDSS spectroscopy (particularly for quasars); since SDSS spectroscopy is used in the ML classification, the information in some of the training set is essentially known perfectly in the classifier (this is one explanation why classification is inferior in non-SDSS footprint fields). Also, SIMBAD sources tend to be brighter than many PTF sources and so the above analysis can be thought of as applying to the brighter end of the distribution. Spectroscopically confirmed SNe candidates found in PTF which are used in the training are obtained after humans in the PTF collaboration have vetted the PTF image-difference-based discoveries and decided to pursue spectroscopic followup.

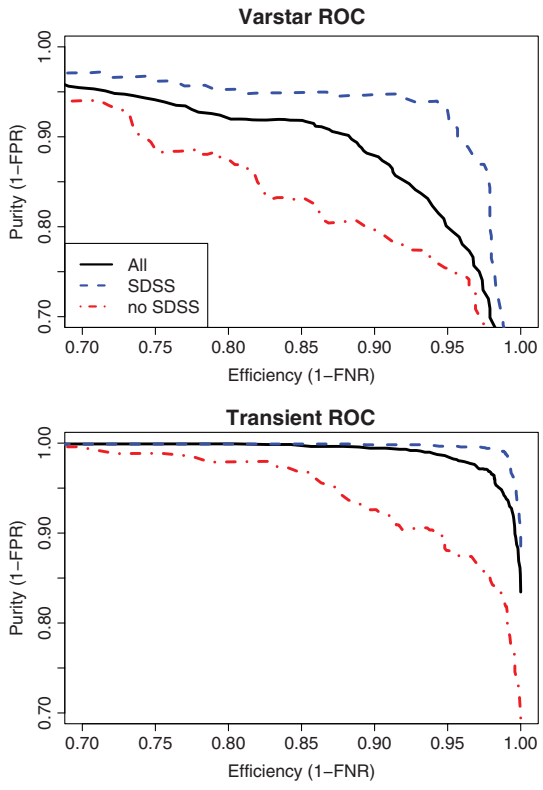


FIG. 13.—ROC curves for PTF Type classification. For each of variable star (*top*) and transient (*bottom*) classification, we plot the efficiency and purity of the random forest classifier as a function of the probability threshold. For the sample of objects used, we recover ∼80% of variable stars and ∼99% of the transient sources at a purity level of 90%. The ROC curves for SDSS objects (*blue dashed*) dominate those for non-SDSS objects (*red dot-dash*).

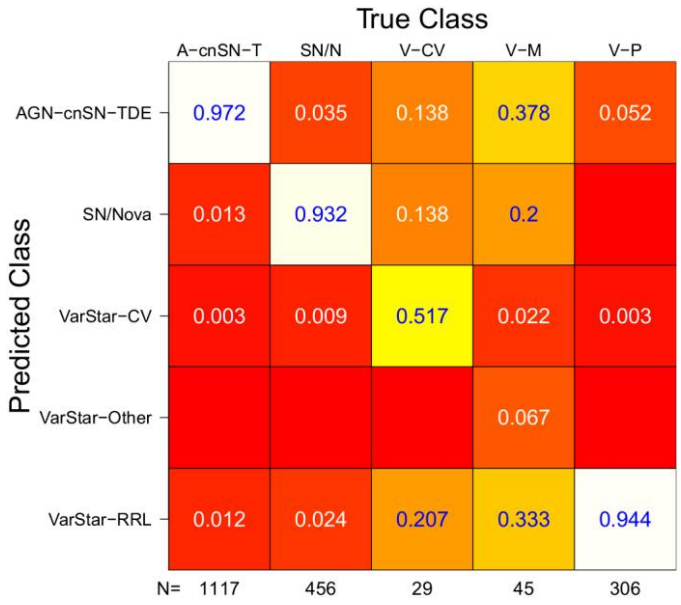


FIG. 14.—Confusion matrix for robotclass random forest classification. Classes are aligned so that entries along the diagonal correspond to correct classification. Probabilities are normalized to sum to unity for each column. Recovery rates are ≥90%, with very high purity, for the three dominant classes. Classification accuracy suffers for the two classes with small amounts of data (class size is written along the *bottom*). See the electronic edition of the *PASP* for a color version of this figure.

A bright supernova that Oarical (or humans) initially type as `VarStar` might not be inspected by humans and therefore not receive a spectroscopic classification. Likewise, if a source is initially labelled as an SN but a human decides not to pursue spectroscopic followup because the candidate is of poor or dubious quality then that source will not be included in the ML training sample. In this sense, the ML results should (conservatively) be viewed as classification results given that the source is (1) observed to vary significantly in the image differences and (2) is a bona fide astrophysical variable or transient.

4. DISCUSSION AND CONCLUSIONS

We have described a framework for building discovery and classification on astronomical synoptic survey streams without humans in the real-time loop. Some features of this framework have been employed previously but, to the best of our knowledge, this is the first example of such an end-to-end framework working in (near) real-time and with real-world data. The use of Oarical in PTF is part an even more expansive thrust of the project in that:

1. The data themselves are acquired on an autonomously operated telescope with a computer-generated observing schedule (Law et al. 2009).
2. Images are transported, reduced, and photometered in near-real time (Law et al. 2009).
3. Discovery and classification results are marked up in a central PTF-wide database.
4. Triggers are then generated for followup by autonomous robotic telescopes (namely P60 and PAIRITEL), which follow-up some high-priority TRANSIENT sources without humans in the loop (Cenko et al. 2012; Gal-Yam et al. 2011).

There is, in this sense, a recognition that follow-up of time-variable sources is crucial for the scientific impact in many domains of interest to the PTF community. Autonomous discovery and classification allows for the initial imaging follow-up to be conducted without astronomers in the real-time loop. Our collaboration also routinely conducts (human-intensive) spectroscopic followup on newly discovered Oarical sources with minimal turnaround times from PTF image to spectroscopy to inference. For instance, we obtained with Keck a spectrum on a newly discovered TRANSIENT 29 minutes after Oarical discovery. The source was a peculiar Type Ia supernova at a redshift $z = 0.18$, and analysis of the spectrum was published less than 18 hr after it was first observed with PTF (Nugent et al. 2010). Gal-Yam et al. (2011) gives a full description of rapid discovery, follow-up, and the scientific results with PTF.

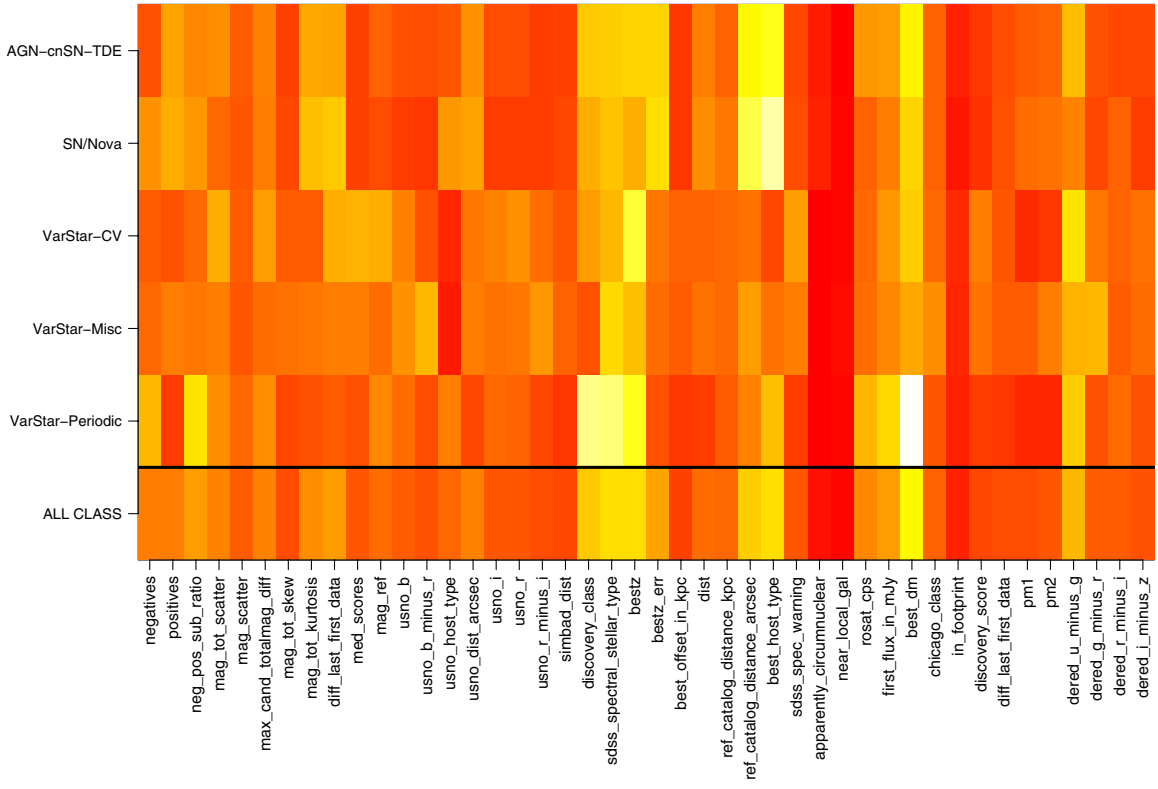


FIG. 15.—Importance of each feature, determined using a pair-wise decision tree algorithm, for classifying objects of each class. Importance ranges from *red* (*low*) to *yellow* (*high*). The average importance across classes shows that PTF light curve features have high importance in the classifier. See the electronic edition of the *PASP* for a color version of this figure.

The 529 spectroscopically-confirmed SNe discovered autonomously by Oarical since 2010 April represent more than half of the SNe discovered by the PTF collaboration over the lifetime of the project. Several key papers have been the result of Oarical discoveries, including discoveries and real-time classification of (1) PTF 10iya, a possible tidal disruption event (Cenko et al. 2012), (2) PTF 10vdl, a subluminous type IIP supernova (Gal-Yam et al. 2011), (3) PTF 10qpf, a TTauri star that appeared to be an FU Ori system in outburst (Miller et al. 2011), (4) PTF 10nvg, an outbursting Class I protostar (Covey et al. 2011), and (5) PTF 10hmv, a type Ia supernova found more than 10 days before maximum and observed with the Hubble Space telescope around maximum light (Cooke et al. 2011). After the original submission of this article, the realbogus score of PTF 11kly (SN 2011fe) was the highest in the list of candidates associated with nearby galaxies on 2011 August 24; this allowed quick discovery (from human-scanning of an automatically generated "local universe" page) and rapid followup of what turned out to be the nearest well studied Type Ia supernova in a generation (Nugent et al. 2011; Li et al. 2011; Bloom et al. 2012). Oarical also discovered PTF 11kx, a Ia SN with a symbiotic nova progenitor (Dilday et al. 2012).

The core discovery and classification codebase has been largely frozen since 2010 April, allowing us to study the results under the assumption of relative uniformity. However, there are several aspects of the framework that we have identified where improvements could be made in future versions (with PTF or otherwise). First, we now have a good deal more ground-truth events in the PTF database that we know are real astrophysical candidates. This larger training set, coupled with new shaped based metrics on the image differences, should much improve the Type I and Type II errors on the discovery front; a new incarnation of "realbogus" is being developed (Brink et al. 2012). Second, there has been much improvement in the astrometric tie of PTF to SDSS (as well as an expanding footprint of public SDSS imaging), which should continue to improve the reliability of distance-to-host features. Third, the database-based photometry used to calculate the time-series features is known to be suboptimal. New routines developed within the collaboration can now allow automated forced-aperture and PSF photometry at the candidate positions. Last, we have now approached a regime where there are enough known classes of sources (from SNe to variable star types) that reliable cross-validated classification can be employed to run machine-learned classifications instead of the manually-tuned classification algorithm (§ 3). It is clear from § 3.4 that ML-based classifications

are reasonably predictive, with TRANSIENT/VARSTAR classification errors at the 5% level.

It is clearly early days for large-scale discovery and classification frameworks for synoptic astronomical surveys. As we look to future implementations, there are several avenues and questions to explore:

1. How do we efficiently discover and classify anomalous sources, those that do not easily fit into the classification categories? Likewise, how can we implement something like a matched-filter discovery of certain classes of sources that have predicted optical light curves but have not been observed before?

2. What should be the unique roles for citizen scientists in the real-time discovery and classification loop; can some forms of citizen-science markups be adequately reproduced by machine-learned codes?

3. Is there a path to using context information immediately with new surveys without having to train with real-world data? That is, is a full prior three-dimensional model of the transient and variable universe needed to train a classifier on the expected contextual data of a survey just coming online?

4. How applicable is the framework detailed herein to other surveys (with different depths, cadences, etc.)? That is, are real-time classification algorithms and codebases more tuned to the PTF survey specifics and idiosyncrasies than we believe?

5. How can we use PTF-tuned classification models to predict classes of sources discovered in other surveys? That is, is there a formal ML-based workflow to bootstrap learning into new survey data? Active (expert) learning might be an appropriate path for exploration (Richards et al. 2012b).

6. Can classification statements be improved markedly as follow-up results are automatically flowed back into a central repository of photometry? We currently do not rerun classification on sources after new data is obtained by the survey.

7. What mechanisms can we use to build up a feedback loop into the classification models? If a source is labelled an SN/Nova but is spectroscopically identified as an RR Lyrae star, how do we automatically learn from our classification mistakes?

8. When, in the course of a survey, is it appropriate to relearn classification based on previous results from the survey? How can the discovery and classification biases from previous incarnations of the framework be controlled in new learning iterations while maintaining control of systematics that are crucial for determining event rates?

These are questions and areas of study we expect to explore in the coming years. With each iteration of the framework, we can hope to produce a more complete and robust framework for use in new surveys. We expect that automatic discovery workflows will need to be highly tuned for each survey but "classification as a service" should evolve as a more general framework that could be hosted and maintained by third parties. This appears to be the direction that the LSST collaboration is heading.

The authors acknowledge the generous support of a CDI grant (#0941742) from the National Science Foundation. J.S.B. and D.L.S. also thank the Las Cumbres Observatory for support during the early stages of this work. S.B.C. wishes to acknowledge generous support from Gary and Cynthia Bengier, the Richard and Rhoda Goldman Fund, National Aeronautics and Space Administration (NASA)/Swift grant NNX10AI21G, NASA/*Fermi* grant NNX1OA057G, and National Science Foundation (NSF) grant AST-0908886. The National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract #DE-AC02-05CH11231, provided staff, computational resources, and data storage for this project.

Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France (Ochsenbein et al. 2000).

Observatories and Facilities: Palomar: 48 inch (PTF) Palomar: 200 inch (DBSP) Keck (LRIS) Apache Point Observatory (2MASS).

REFERENCES

Akerlof, C. W., et al. 2003, PASP, 115, 132
Bailey, S., Aragon, C., Romano, R., Thomas, R. C., Weaver, B. A., & Wong, D. 2007, ApJ, 665, 1246
Becker, A., Axelrod, T., Ivezic, Z., Lupton, R., Silvestri, N., & Rest, A. 2005, Bulletin of the American Astronomical Society, 37, 1206, American Astronomical Society Meeting Abstracts
Belokurov, V., Evans, N. W., & Du, Y. L. 2003, MNRAS, 341, 1373
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, AJ, 142, 31
Bloom, J. S., et al. 2012, ApJ, 744, L17
Bloom, J. S., & Richards, J. W. 2011, Advances in Machine Learning and Data Mining for Astronomy, ed. M. J. Way, J. D. Scargle, K. M. Ali, & A. N., Srivastava (1st ed.; Boca Raton: Chapman and Hall/CRC), 89–112
Bloom, J. S., Starr, D. L., Blake, C. H., Skrutskie, M. F., & Falco, E. E. 2006, in ASP. Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique (San Francisco: ASP), 751
Bond, I. A., et al. 2001, MNRAS, 327, 868
Brink, H., Richards, J. W., Poznanski, D., Bloom, J. S., Rice, J., Negahban, S., & Wainwright, M. 2012, preprint (arXiv/1209.3775)

Butler, N. R., & Bloom, J. S. 2011, AJ, 141, 93
Castro-Tirado, A. J., Bloom, J. S., Hanlon, L., & Kotani, T. 2010, Adv. Astron., 2010 (Hindawi Publishing Corporation)
Cenko, S. B., et al. 2006, PASP, 118, 1396
Cenko, S. B., Bloom, J. S., Kulkarni, S. R., Strubbe, L. E., Miller, A. A., Butler, N. R., Quimby, R. M., Gal-Yam, A., et al. 2012, MNRAS, 420, 2684
Ciurana, E. 2009, Developing with Google App Engine (Berkeley: Apress)
Cooke, J., et al. 2011, ApJ, 727, L35
Covey, K. R., et al. 2011, AJ, 141, 40
Debosscher, J., Sarro, L. M., Aerts, C., Cuypers, J., Vandenbussche, B., Garrido, R., & Solano, E. 2007, A&A, 475, 1159
Dilday, B., Howell, D. A., Cenko, S. B., Silverman, J. M., Nugent, P. E., Sullivan, M., Ben-Ami, S., Bildsten, L., et al. 2012, Science, 337, 942
Drake, A. J., et al. 2009, ApJ , 696, 870
Filippenko, A. V., Li, W. D., Treffers, R. R., & Modjaz, M. 2001, in IAU Colloq. 183, Small Telescope Astronomy on Global Scales, ed. B. Paczynski, W.-P. Chen, & C. Lemme (ASP Conf. Ser. 246; San Francisco: ASP), 121
Flaugher, B. 2005, Int. J. Mod. Phys. A, 20, 3121
Gal-Yam, A., et al. 2011, ApJ, 736, 159
Gal-Yam, A., & Mazzali, P. 2011, preprint (arXiv/1103.5165)
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. 2009, SIGKDD Explorations, 11, 10
Ivezić, Ž., et al. 2003, Mem. Soc. Astron. Italiana, 74, 978
———. 2008, preprint (arXiv/0805.2366)
Jeffreys, H. 1946, Proc. R. Soc. London A Math. Phys. Sci., 186, 453
Jurić, M., & Ivezić, Ž. 2011, EAS Pub. Ser., 45, 281
Kaiser, N., et al. 2002, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. J. A. Tyson, & S. Wolff (Waikoloa, HI, August 22, 2002), 154
Keller, S. C., et al. 2007, PASA , 24, 1
Kohavi, R., & Quinlan, J. 2002, in Handbook of Data Mining and Knowledge Discovery (Oxford: Oxford University Press), 267
Kubanek, P. 2010, preprint (arXiv/1002.0108)
Law, N. M., et al. 2009, PASP, 121, 1395
———. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 7735 (San Diego, CA, June 27, 2010)
Li, W., et al. 2011, Nature, 480, 348
Mahabal, A., et al. 2008, in AIP Conf. Proc. 1082, American Institute of Physics Conference Series, ed. C. A. L. Bailer-Jones (New York: AIP), 287
Miller, A. A., Smith, N., Li, W., Bloom, J. S., Chornock, R., Filippenko, A. V., & Prochaska, J. X. 2010, AJ , 139, 2218
Miller, A. A., et al. 2011, ApJ, 730, 80
Nugent, P., et al. 2010, ATel, 2600, 1
Nugent, P. E., et al. 2011, Nature, 480, 344
Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23
R Development Core Team. 2005, R: A Language and Environment for Statistical Computing (Vienna: R Foundation for Statistical Computing)3-900051-07-0
Rau, A., et al. 2009, PASP, 121, 1334
Richards, J. W., et al. 2011, ApJ, 733, 10
Richards, J. W., Starr, D. L., Miller, A. A., Bloom, J. S., Butler, N. R., Brink, H., & Crellin-Quick, A. 2012a, ApJ Supp., in press
Richards, J., et al. 2012b, Astrophys. J., 744, 192
Sarro, L. M., Debosscher, J., López, M., & Aerts, C. 2009, A&A, 494, 739
Sarro, L. M., Sánchez-Fernández, C., & Giménez, Á. 2006, A&A, 446, 395
Saunders, E. S., Naylor, T., & Allan, A. 2008, Astron. Nachr., 329, 321
Sesar, B., et al. 2010, ApJ, 708, 717
Smith, A. M., et al. 2011, MNRAS, 412, 1309
Sokołowski, M., Małek, K., Piotrowski, L. W., & Wrochna, G. 2010, Adv. Astron., 2010, 11
Stekhoven, D. J., & Bühlmann, P. 2011, preprint (arXiv/1105.0828)
Sullivan, M., et al. 2011, ApJ, 732, 118
Tomaney, A. B., & Crotts, A. P. S. 1996, AJ, 112, 2872
Vestrand, W. T., et al. 2002, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 4845, ed. R. I. Kibrick, 126
Willemsen, P. G., & Eyer, L. 2007, preprint (arXiv/0712.2898)
Wozniak, P. R. 2000, Acta Astron., 50, 421
Yip, C. W., et al. 2004, AJ, 128, 585