Mixing Bayesian Techniques for Effective Real-time Classification of Astronomical Transients

Ashish Mahabal, S. G. Djorgovski, Ciro Donalek, Andrew Drake, Matthew Graham, and Roy Williams
California Institute of Technology, Pasadena, CA, USA
Baback Moghaddam and Michael Turmon
Jet Propulsion Laboratory, Pasadena, CA, USA

Abstract. With the recent advent of time domain astronomy through various surveys several approaches at classification of transients are being tried. Choosing relatively interesting and rarer transients for follow-up is important since following all transients being detected per night is not possible given the limited resources available. In addition, the classification needs to be carried out using minimal number of observations available in order to catch some of the more interesting objects. We present details on two such classification methods: (1) using Bayesian networks with colors and contextual information, and (2) using Gaussian Process Regression and light-curves. Both can be carried out in real-time and from a very small number of epochs. In order to improve classification i.e. narrow down number of competing classes, it is important to combine as many different classifiers as possible. We mention how this can be accomplished using a higher order fusion network.

1 Introduction

In the last several years interest in time domain astronomy has been on the rise since the immense potential to find hitherto unknown classes and subclasses is increasingly being realized (e.g. AM CVn; SN before they reach their peak magnitude). The increasing number of surveys at varied wavelengths bear a witness to this. While obtaining a spectrum remains the definitive standard for determining class, just the magnitude of objects to follow makes this an unlikely option except for a small fraction. The other standard techniques to look for transients and classify objects include light-curves i.e. change in flux as a function of time, colors i.e. change in flux as a function of wavelength, as well as discriminating contextual information like (a) distance from a galaxy - smaller value more suggestive of an extragalactic origin and of type supernova, and (b) galactic latitude - lower absolute values more suggestive of galactic nature. The nearly coming of age of the Virtual Observatory and the use of its tools to mine the data from past surveys is another big asset for classification. However, most of these techniques need several data for reliable classification much of which is collected from follow-up observations especially for the rarer/newer sub/classes where archival information is not available for similar objects in sufficience. Follow-up itself is tricky since the resources are scarce and the number of objects needing follow-up is continually growing. To deal with this we
Figure 1. (a) CRTS transients are followed-up at the Palomar 60-inch telescope in griz filters. Stellar locus shown here is from SDSS (filled circles) with g-r and r-i colors overplotted (asterisks). There are other transients with more extreme colors that have not been included here e.g. g- or z-band dropouts. (b) One way we use these colors is to classify the transients using a conceptually simple Bayesian network. Besides the three colors we currently use two context parameters viz. Galactic latitude and proximity to a galaxy. The output classes are four as described in the text. However, the network is trivially extensible to more output classes as well as to include additional inputs e.g. x-ray flux. Since missing data are elegantly handled, adding such new parameters does not affect the outcome for objects where those data are missing. The final aim is to link the classifications to the phenomenology i.e. the physical processes which manifest as the different classes of objects.

have been developing Bayesian techniques that provide consistently improving classification based on increasing amount of information starting from just two datapoints. These include a Bayesian Network and Gaussian Process Regression (GPR) which we have described earlier (Mahabal, A., et al., 2008a; Mahabal, A., et al., 2008b; Donalek, C., et al., 2008). Here we provide recent updates and enhancements. We then briefly describe an evolving scheme of combining such classifiers to provide a more unified picture using a fusion network. Optimized feedback can then further improve the classification.

2 The Methods and the Preliminary Results

2.1 Bayesian Network

Our current results are based on colors obtained from the Palomar 60-inch telescope from follow-up of objects detected with the Catalina Real-time Transient survey (CRTS) (http://crts.caltech.edu; Drake, A., et al., 2009), a subset of which is shown in Figure 1a. CRTS is based on detections from the Catalina Sky Survey (CSS) which is operated in an open mode (no filter). The current transient detection threshold for $dm$ is high and a function of magnitude. As a result a majority of the objects tend to be supernovae or CVs. Of the remaining those that have a radio counterpart are currently loosely called Blazars. The remaining ones can not correctly be classified from the sparse data. As we start lowering the $dm$ threshold we expect to see more variety in types of transients.
Real-time Classification

Figure 1b provides a schematic of the Bayesian Network based on colors and contextual information. Even from the current set-up it is clear that more context information will help (the use of the two context parameters improves the classification by $\sim 10\%$). The issues of uniformity are however thornier. The priors for different classes typically originate in different sets of observations and hence lead to different completenesses. Owing to the selection based on large $dm$ several types get lumped into the ‘Rest’ class resulting in a mixed prior and hence lower classification efficiency ($\sim 20 - 30\%$). Blazars too have a smaller representation compared to CVs and SNs and that too is reflected in the numbers ($\sim 35 - 40\%$ compared to $\sim 70 - 80\%$ efficiency). If all three colors are demanded, the available prior size decreases, and the efficiency falls by a few percent. Large uniform priors are extremely important for reliable classification especially when it is based on just a couple of points. With large number of possible classes, if evaluated simultaneously, even the cleanest classes may have part of their probability hijacked by observed values of certain variables leading to inaccurate classification. Another interesting issue is of deciding which type of winner to choose. The winning class needs to have a higher probability than the other classes. We have currently been using three criteria viz. (a) the winner is based on the highest probability, however small; (b) the winner needs to have at least 50% probability; and (c) the ‘40%-10%’ case where the winner needs to have a probability of at least 40% and the next class should at least be 10% lower. With the existing prior sizes, these criteria affect the classification efficiency by not more than a few percent. But what if the probability of the winning class is only marginally higher than the others? What if all classes have small probabilities? Such an object would in fact be quiet interesting since it matches none of the existing/input priors. A few such objects have indeed vanished before a spectrum could be obtained.

2.2 Gaussian Process Regression (GPR)

We have described the GPR framework in Mahabal, A., et al. (2008a). We are now starting to experiment with adding different types of covariance functions to cater to different types of variables. Transients like supernovae, for instance, are best dealt with by a covariance function with scale factors (with $t$ and $t'$ being two epochs) $k(t, t') = \sigma_f^2 \exp\left[-\frac{(t-t')^2}{2 l^2}\right] + \sigma_n^2 \delta(t, t')$ whereas for periodic variables, functions with one (or more) sinusoidal components are needed $k(t, t') = \sigma_f^2 \exp\left[-\frac{(t-t')^2}{2 l^2}\right] + \exp\left\{-2 \sin^2\left[\nu \pi (t - t')\right]\right\} + \sigma_n^2 \delta(t, t')$.

Models are built for different classes and the couple of observational points available for a newly detected transient are compared against each of these. Owing to the possible range of variation for different parameters internal to a class (e.g. period, amplitude, maximum brightness) it is easier to eliminate certain classes and differentiate between groups of classes rather than an accurate classification. But once more data start accumulating based on iterative follow-up, zooming in on the correct class is possible. Incorporation of upper limits and error-bars is however not trivial and is an active area of research.
Figure 2. Distinct classification information from varied methods is combined in a meaningful manner to take advantage of the ‘expertise’ of these methods. This piece forms a part of the larger picture which includes semantically connected transient portfolios and real-time alerts.

2.3 Fusion Network

Besides the two methods mentioned above, there are others like neural networks (which operate best when there are no missing data). The probabilities that each of these produce can vary based on the inputs and the applicability of the method. Results from the different classifiers can be combined through a fusion module using a sleeping experts framework, where each specialist makes a prediction only when the instance to be predicted falls within its area of expertise. External and contextual knowledge can be used to awaken an expert or to put it to sleep and to modify online the weights associated with each classifier. The fusion module can help improve the classification, narrowing down the number of competing classes and leading to optimal follow-ups and new discoveries.

Acknowledgments. This work has been supported in part by the NSF grants AST-0407448, AST-0909182, NASA AISRP grant 08-AISR08-0085, and by the Fishbein family foundation.

References