Advanced Tools for Exploring Large EB Datasets

E. J. Devinney, A. Prša, E. F. Guinan and M. Degeorge

Villanova University, Dept. of Astronomy & Astrophysics, 800 Lancaster Ave,
Villanova PA 19085, USA; email: `edward.devinney@villanova.edu`

Abstract. Thanks to OGLE, Kepler, CoRoT and planned new ambitious survey projects, the eclipsing binary (EB) community is beginning to experience a long-predicted data deluge. Beyond the analysis of the many fascinating individual objects yielded by these programs, these complete datasets themselves should yield further insights. Because objects in such datasets are characterized by many parameters, tools that assist in understanding high-dimensional data are acquiring increasing relevance. Chiefly among these are new Advanced Visualization (AV) tools and various methods of clustering data, both approaches complementing each other naturally. We illustrate the use of these tools as applied to OGLE II LMC EB data and respective EBAI light curve solutions.

Keywords. methods: data analysis, methods: numerical, binaries: eclipsing, stars: fundamental parameters, stars: statistics

1. Introduction

GGOBI (www.ggobi.org) and Mondrian (www.rosuda.org/Mondrian) are widely-used AV toolsets, both supporting simultaneous views of multiple, linked two dimensional scatterplots and histograms/barcharts. Objects selected in one plot are highlighted immediately in all the others (brushing), thus helping to identify groups of physically-related objects and offering clues to the nature of outliers. GGOBI also provides for a grand tour of a high-dimensional dataset. This is implemented as a rotating 3D plot into which is projected all the (higher dimensional) data in round-robin fashion. Clusters can then be identified as relatively isolated groupings of points during various phases of the data presentation. While not discounting the hands-on familiarity with a dataset offered by the grand tour process, in recent years the machine-learning community has developed a panoply of automatic clustering algorithms, able to identify clusters of data points in high-dimensional space and thus candidate physical groups. For example, the Waikato University WEKA clustering package (www.cs.waikato/ml/weka) is fast and easy to use, providing a variety of clustering algorithms and the ability to plot the results.

2. Example Application: OGLE II LMC EBs

The two three-part panels, Figs. 1 and 2, illustrate the use respectively of the GGOBI and WEKA packages on OGLE II data from the LMC with light curve solutions using the EBAI neural network system (Prša *et al.* 2008). In each panel, from left to right are: V-I vs I-magnitude; logP[d] vs I-magnitude; and V-I vs logP[d]. In the top GGOBI panel, the middle figure reveals a group of objects showing a period vs I-mag behavior. The left panel indicates these are giants, while the right panel dotted box shows that they are virtually exclusively defined by period and color, as found through interactive exploratory plotting. The WEKA panel below shows the same result, but the cluster of
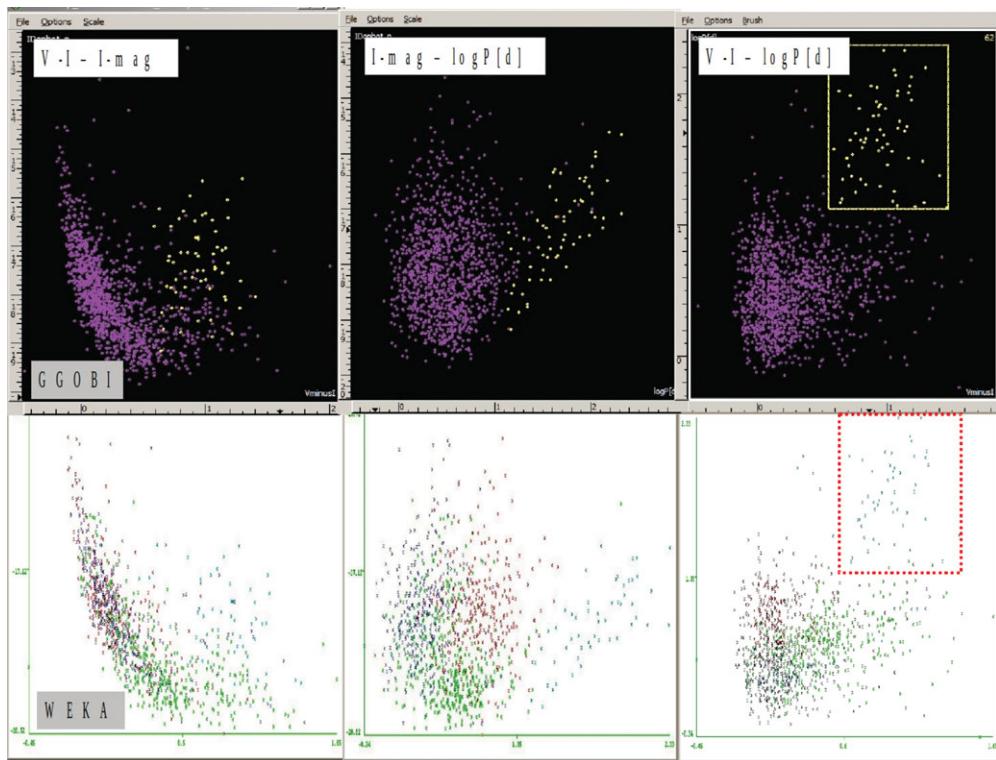
Figure 1. GGOBI (top) and WEKA (bottom) plot panels. Top panel, right, interactive discovery of objects responsible for the period vs I-magnitude trend in the middle plot. Bottom panel, right, the same objects automatically found by the clusterer.

points in the right plot (dotted box) was found by automatic clustering applied to the combined OGLE II LMC and EBAI dataset. This same clustering also automatically delineated main-sequence objects from evolved ones (not shown).

3. Conclusions and New Work

Our experience suggests that AV and clustering tools are destined to become ever more valuable with the increasing data deluge, and that wider effort to exploit them will yield scientific payoffs. We also found the existing AV tools have some deficiencies. Neither GGOBI nor Mondrian are convenient for extended analysis and both have issues in axis scaling. The most serious drawback is that on quitting these applications, all plots and the state of the application are lost. As a result, one of us (MD) is well-advanced on a Python-based AV package PyVU that overcomes these limitations. PyVU allows brushing in multiple plots, has rational axis scaling and labeling, supports display of light curves associated with one or more selected data points and, on quitting, saves the state of the application along with user comments/session summary. With this latter feature, branching explorations of datasets are supported. PyVU will be made available to the user community.

References

Prša, A., *et al.*, 2008, *ApJ*, 687, 542