

Computational Intelligence in Astronomy – A Win-Win Situation

Peter Tiño¹ and Somak Raychaudhury²

¹ School of Computer Science,

² School of Physics and Astronomy

University of Birmingham, Birmingham B15 2TT, UK

{P.Tino@cs,somak@star.sr}.bham.ac.uk

<http://www.birmingham.ac.uk>

Abstract. Large archives of astronomical data (images, spectra and catalogues of derived parameters) are being assembled worldwide as part of the Virtual Observatory project. In order for such massive heterogeneous data collections to be of use to astronomers, development of Computational Intelligence techniques that would combine modern machine learning with deep domain knowledge is crucial. Both fields - Computer Science and Astronomy - can hugely benefit from such a research program. Astronomers can gain new insights into structures buried deeply in the data collections that would, without the help of Computational Intelligence, stay masked. On the other hand, computer scientists can get inspiration and motivation for development of new techniques driven by the specific characteristics of astronomical data and the need to include domain knowledge in a fundamental way. In this review we present three diverse examples of such successful symbiosis.

1 Introduction

The field of Computational Intelligence is not only concerned with the theoretical basis of intelligent data processing and machine learning, but it has also made substantial contributions to a wide variety of disciplines, with new subject areas emerging from these interactions (e.g. bioinformatics, computational finance), with their own dedicated journals, conferences etc. In the field of Astrophysics, some of the fundamental research problems involve the discovery of unusual sources or patterns in multivariate data, where dimensionality is high and, due to the heterogeneous ways of acquisition, significant portions of the data might be in the form of limits (“censored data”) or missing altogether.

Large archives of astronomical data (images, spectra and catalogues of derived parameters) are being assembled worldwide as part of the Virtual Observatory (www.ivoa.net), to be made available to the wider community over the coming decade. This necessitates the development of fast automated techniques for feature extraction, classification and outlier detection for very large datasets. Computational Intelligence will enable the visualisation of the structure of high dimensional and structured data, as well as flexible extraction and study of relevant patterns and substructures. As with experimental data in other branches

of the physical sciences, the quantifiable systematic and random errors of measurement inherent in the data have to be taken into account in most problems [33].

One of the earliest areas of Astrophysics in which machine learning methods were used was in the morphological classification of galaxies, where the norm has been that of visual classification by experts, till the number of galaxies required to be classified for the purposes of the study of galaxy evolution exceeded 10^4 . Artificial neural networks [19,1] or support vector machines [36] have been popular in studying the nature of galaxy evolution from photometric images. In recent years, with the availability of optical spectra of $> 10^6$ galaxies, the study of individual stellar populations within galaxies have been helped by incorporating independent component analysis and other data-driven techniques [21,22].

In this review, we give examples from three areas of astronomical problems where our computational intelligence research has provided unique solutions, which, informed by the knowledge of physics, have helped understand the underlying astrophysical phenomena.

2 Automated Calibration of Galaxy Disruption Simulations [38]

In the hierarchical growth of structure of the Universe, galaxies grow by merging. A giant galaxy, like the Milky Way, grows predominantly by minor mergers (i.e. involving smaller dwarf galaxies or satellites). When a satellite merges with a giant galaxy, it gets tidally disrupted in the process and eventually is completely assimilated in the giant galaxy [20]. Deep observations of the nearby Universe have revealed many examples of such tidal streams that resulting from tidal disruptions. Detailed studies of these tidal streams and debris can provide valuable insight into the detailed mechanism of galaxy formation and evolution.

Other than the satellite system of our own galaxy (Milky way), the most-studied system of streams involves the closest (spiral) galaxy M31, where structures such as shells and streams of stars have been discovered in abundance in its vicinity. These exciting discoveries have led astronomers to investigate the possibility that such structures are in fact remnants of disrupted smaller satellite galaxies [8]. Models involving dark matter and stars are used to simulate the process of satellite galaxy disruption in the vicinity of a large galaxy [8]. There are models with a large number of particles, representing groups of stars. The simulation space is 6-dimensional (“phase space”), three describing the spatial position of each particle, the other three describing the velocity along the spatial coordinates. The particle evolution is governed by gravitational dynamics and hydrodynamics. To track the evolution process starting from a particular initial condition, the state of the simulation (values of all six phase space parameters) are recorded successive evolution stages. Hence, the disruption process is captured in a series of simulated datasets.

In these simulations, the observed low-dimensional structures evolve along with the satellite galaxy, differing slightly from one simulation to another according to the different initial conditions. It makes that the low-dimensional structure in these simulated datasets looks very similar, but not identical. The observational astronomer, however, observed only one snapshot in time, i.e., the current state of the evolution. The ultimate goal of astronomers is to identify the most plausible set of initial conditions leading to the distribution of stars currently observed by the astronomers.

One possible way of learning the most plausible set of initial conditions is to identify the simulated stars having the most similar distribution to that of the currently observed stars. Obviously, it is impossible to compare the particles in simulation to the real observations on a point by point bases, but the observed *density* of stars can be compared with that of the simulated particles. In [38] we reported first results from an ongoing work concentrating on measuring the ‘similarity’ between the simulated datasets and the observation dataset through non-parametric probability density estimation.

The approach stands and falls on the quality of density models on the simulated data. Unfortunately, parametric density estimation cannot be used, as the disrupted satellites can produce complex low-dimensional manifolds along which the stars are organized (tidal streams, shells). Semi-parametric approaches such as Gaussian mixture modelling, fitted in the maximum likelihood framework via EM algorithm, are not capable of capturing such complex density structure because of the high sensitivity to initialization. Unless we provide good estimates of the positions and shapes of low dimensional structures floating in the cloud of high dimensional points, already before the mixture fitting, there is no hope of finding a satisfactory density estimate.

Non-parametric density estimation methods seem like a plausible approach - unfortunately they cannot be directly used since they typically put a local density kernel on each observation and the number of simulated particles makes such an approach computationally prohibitive. To reduce the computational cost, several algorithms have been devised to either reduce the sample size, or to reduce the amount of components (kernels) in the original complex Parzen Windows model. The latter approaches, introduced recently, have proved successful in several applications [11,40,39]. The idea is to simplify a complex model (Parzen Windows in density estimation) while minimizing the distance between the new model and the target function. However, this process usually has complexity of $\mathcal{O}(N^2)$ or larger, where N is the size of the dataset. Compared with considering fewer points in the estimation, optimizing the simplified density model by using all of the available observations could avoid sacrificing useful information, but it also loses the simplicity of the non-parametric density model. On the other hand, the limitation of Parzen Windows could also be noticed when the data points distribute along, or partially along, low dimensional structures [35]. In such cases, spherical smoothing kernels are not optimal.

In [38], we proposed a simple new algorithm to reduce the computational cost of PW, but also to keep the simplicity of the nonparametric model by avoiding complex model construction procedures. The main idea was to cover the entire data space by a set of hyper-balls of fixed (carefully chosen) radii. For each hyper-ball, the local density was captured by a full covariance Gaussian kernel. Our model is formed by a mixture of such locally fitted Gaussians with appropriately set mixing priors.

The densities fitted in the 6-dimensional simulation can be projected into the observation space (typically two spatial coordinates + the line of sight velocity) and the likelihood given the real observations calculated. We tested the approach on 22 simulation sets representing 22 stages of galaxy disruption in a single simulation run. Each set had 32,768 data points (particles).

In the first set of experiments we used 10-fold cross-validation in each simulation data set to measure the quality of the estimated density models. The average log-likelihoods (ALL) on each individual set estimated by Parzen Window (PW), our approach - Fast PW (FPW) and Simplified Mixture Model (SMM) [40] are plotted in figure 1. We actually used two versions of FPW: **(1)** a simple ‘hard’ FPW (FPW-H) which estimated local means and covariances solely on the points within the covering hyperballs; **(2)** a ‘soft’ version (FPW-S) in which local means and covariances are estimated in a soft manner by positioning Gaussian weighting kernels in the hyperball centers. The hyper parameters were estimated through 10-fold cross validation within the training folds. The X-axes indicates the simulation set index (numbered from 0 to 21), the Y-axes shows the ALLs estimated by the models. Both FPW versions show a superior performance relative to both PW and SMM estimators.

In the second set of experiments, we investigated how reliably a stage in the galaxy disruption process can be detected, based on ‘observations’ not used in the model building process. We run a rolling window of size 3 over the series

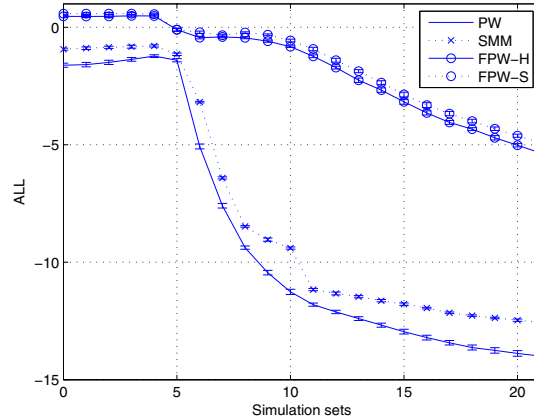


Fig. 1. Average log-likelihood on 22 simulation sets

of 22 simulation data sets, obtaining a set of 20 simulation set triplets (f, g, h) , starting with simulation sets (f, g, h) at stages (0,1,2) and ending with the triplet (f, g, h) containing simulation set for stages (20,21,22). For each triplet of consecutive simulation sets (f, g, h) , we estimated 3 models, one on 90% of data from f , one on 90% of data from g , and one on 90% of data from h . We also compared with Gaussian Mixture Model initialized randomly from the data (GMM) or initialized with K-means (GMM^k). To make experiments computationally feasible only 10% of the data was used to estimate the GMM models. All the models were then tested on the 10% hold-out set from g . In this way we could determine how well could the ‘true’ source density g be distinguished from the densities at the nearby stages f and h . This process was repeated 10 times.

We stress that the task is quite complicated as the densities corresponding to the nearby stages of galaxy disruption can be quite ‘similar’ and obtaining an accurate density model is essential for further investigations by the astronomers. As an illustration, we present in figure 2 two sets of 3-dimensional projections of the simulation data for the triplet $(f, g, h) = (20, 21, 22)$. The first projection is onto the spatial coordinates, the second is onto the leading 3 eigenvectors found by the Principal Component Analysis of the merged f, g and h sets.

All the models constructed on the set g have the highest hold-out ALL in each of the 20 triplets (f, g, h) . However, the margin with which the set g was proclaimed the ‘winner’ was quite different across the models. The margins of the two FPW versions were almost the same and the variations in the margins due to different experimental runs were negligible. The FPW methods outperformed the classical PW estimation and showed performance levels very similar to those of GMM, but with much less computational effort. The number of components in FPW-H and FPW-S varied from 150 to 850 and 700 to 4000, respectively. Note that the number of components in PW estimates was $\approx 30,000$.

3 Time Delay Estimation in Gravitationally Lensed Signals [4,5]

Time delay estimation between arrival times of two signals that originate from the same source but travel along different paths to the observer is a real-world problem in Astronomy. Signals to be analysed can represent repeated measurement, over many months or years, of the flux of radiation (optical light or radio waves) from a very distant quasar - a very bright source of light usually a few trillion light-years away. Some of these quasars appear as a set of multiple nearby images on the sky, due to the fact that the trajectory of light coming from the source gets bent as it passes a massive galaxy on the way (the “gravitational lens”). As a result, the observer receives the light from various directions, resulting in the detection of several images [16,32]. This phenomenon is called *gravitational lensing*. It is a natural consequence of a prediction of the General theory of Relativity, which postulates that massive objects distort space-time and thus cause the bending of trajectories of light rays passing near them.

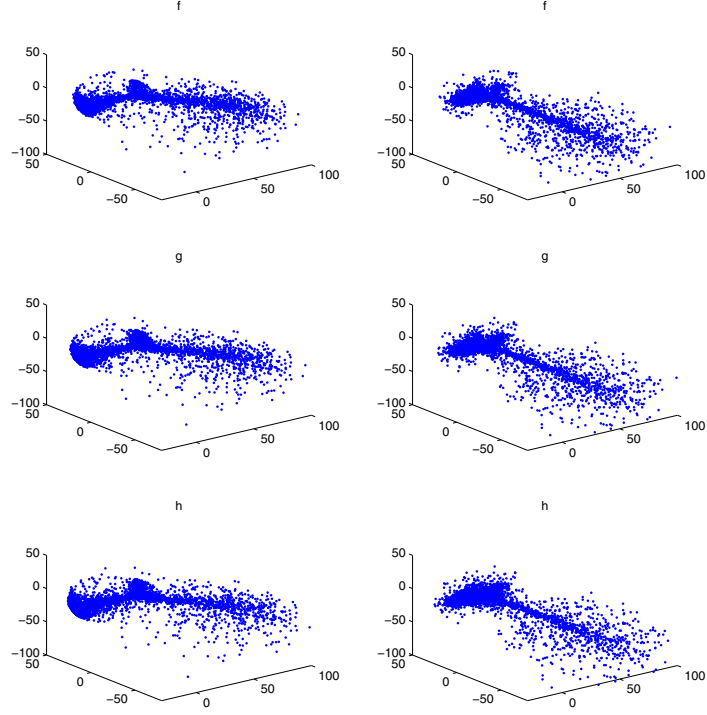


Fig. 2. 3-dimensional projections of the simulation data for the triplet $(f, g, h) = (20, 21, 22)$. The first projection (1st column) is onto the 3 spatial coordinates, the second (2nd column) is onto the leading 3 eigenvectors found by the Principal Component Analysis of the merged f, g and h sets.

Quasars are variable sources, and the same sequence of variations is detected at different times in the different images, according to the travel time along the various paths. Crucially, the time delay between the signals depends on the mass of the lens, and thus it is the most direct method to measure the distribution of matter in the Universe, which is often dark [31,16].

In this context, the underlying temporal pattern of emitted radiation fluxes from a quasar gets delayed, and the process of observation adds all kinds of noise. Moreover, like all astronomical time series measurements, they are also typically irregularly sampled with possibly large observational gaps (missing data) [24,29,23,15]. This is due to practical limitations of observation such as equipment availability, weather conditions, the brightness of the moon, among many other factors [7]. Currently, over a hundred systems of lensed quasars are currently known and about a dozen of these have been monitored for long periods. In some of these cases, the measurement of a time delay has been claimed.

In [5,4] we introduced a model based technique for estimating time delays in fluxes from gravitationally lensed objects (such as quasars). The main idea of the method was to impose an internal model on the quasar variability in time and then expect that the multiple images will follow that model, up to observational noise and time delays. We formulated the internal model of quasar variability in the framework of kernel regression (with free parameters determined from available data). Our approach was compared with currently popular methods of estimating time delays from real astronomical data: **(1)** *Dispersion spectra* method [26,27,25], and **(2)** the structure-function-based method (*PRH*, [30]). Two versions of Dispersion spectra were used; D_1^2 is free of parameters [26,27] and $D_{4,2}^2$ has a decorrelation length parameter δ involving only nearby points in the weighted correlation [27,25]. In the case of PRH method, we used the image A from the data to estimate the structure function [30].

For experimental (observational) data, we focused on Q0957+561, the first multiply-imaged quasar to be discovered [37]. This source, which has a pair of images (that we refer to as A and B), has been monitored for over twenty years (in both radio and optical range). In figure 3 we show an example of real optical data measured from Q0957+561. The optical fluxes observed from images A and B are given in the astronomical unit of magnitude (mag m), defined as $m = -2.5 \log_{10} f$, where f is the flux measured when observed through a green filter [18] (g -band). The measurement errors are shown as error bars. The source was monitored nightly, but many observations were missed due to cloudy weather and telescope scheduling. The big gap in Fig. 3 is an intentional gap in the nightly monitoring, since a delay of about 400 days, the pattern, was known ‘a priori’ – monitoring programs on this quasar started in 1979. Therefore, the peak in the light curve of image A, between 700 and 800 days, corresponds to the peak in that of image B between 1,100 and 1,200 days.

Despite numerous claims, a universally agreed value for the time delay in the Q0957+561 system has not emerged [18,5]. Indeed, a major problem with time delay estimation in astrophysics literature has been that these estimates are routinely produced for individual quasars, for which we have no idea what the ‘true’ time delay is (e.g. [30,28,3,7,13]). The uncertainty bounds in the reported estimates are mostly due to assumed noise model on the observations – the estimation has been repeated in a series of Monte Carlo experimental data generated from the measured flux values, under the noise model. However, for an unbiased comparison of alternative methods, before they are employed to the analysis of real data, they should be subjected to a large data collection in a controlled experimental setting where the time delay is externally imposed and known.

In [4] we generated a large number of flux times series resembling fluxes from real quasars. Three data generation mechanisms were considered:

- We simulated optical observations as in [24]. The data sets were irregularly sampled with three levels of noise and observational gaps of different size. 50 data set realisations per noise level were generated, yielding 38,505 data sets

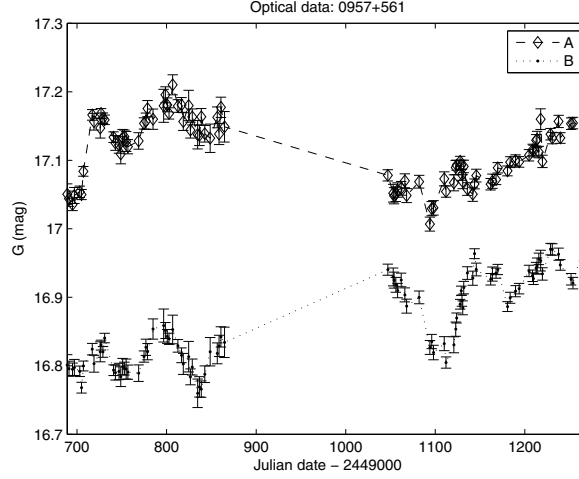


Fig. 3. Optical observations of the brightness of the doubly-imaged quasar Q0957+561, in the g -band, as a function of time (Top - Image A; Bottom - Image B). The time is measured in days (Julian days-2,449,000 days).

- Data streams were generated by Gaussian process, following [30], with a fixed covariance matrix given by a structure function according to [29]. The data was highly sampled with periodic gaps [7], simulating a monitoring campaign of eight months; yielding 61 samples per time series. We imposed seven ‘true’ delays and 100 realisations for each value of true delay [5].
- The data was generated from a Bayesian model [13], simulating three levels of noise with 225 data sets per level of noise (there were 3 levels of noise). The data were irregularly sampled with 100 observations per time series.

It is remarkable how the imposition of a unifying smooth internal model in our method stabilized the delay estimation. After tests for statistical significance, the machine learning based method employing kernel regression emerged as a clear overall winner in the large set of controlled experiments on synthetic data with known time delay. We stress that this method was able to sometimes outperform even methods that were used to generate the data sets themselves (Gaussian process or the Bayesian model of [13]).

In terms of real data from Q0957+561, the best (smallest estimated error) time delay quotes were 417 ± 3 [18] and 419.5 ± 0.8 [6]. Our results were consistent with these findings. However, we speculate that the estimate of 417 ± 3 days constitutes an underestimate. For the quasar Q0957+561, the latest reports also give estimates around 420 days by using other data sets [24]. The delay estimates should be robust across the wavelength range since the gravitational lensing theory predicts that the time delay must be the same regardless of the wavelength of observation [31,16,32].

4 Clustering and Topographic Mapping of Light Curves from Eclipsing Binary Stars [9]

A binary is a gravitationally bound pair of stars that orbit a common centre of mass. Astronomical observations suggest that almost half of the stars are in binary systems [12]. Thus, the study of such systems is a crucial element for understanding a significant proportion of stars. Amongst other reasons, binary stars are important to astrophysics because they allow calculation of fundamental quantities such as masses and radii. The increasing number of binary star discoveries provides samples for testing theoretical models for stellar formation and evolution. Also, by measuring their fundamental parameters they can serve as distance indicators to galaxies. Moreover, the study of binaries has led to the discovery of extrasolar planets. A particular subclass of binary stars are eclipsing binary stars. The luminosity of such stars varies over time and forms a graph called light curve. Light curves are important because they provide information on the characteristics of stars and help in the identification of their type. It is therefore of great importance for the astronomers to have tools for automated analysis of large repositories of light curves from binary systems, e.g. tools for grouping and topographic mapping of such systems.

While clustering of a data set is mainly concerned with finding natural groups of data items e.g. according to some ‘similarity’ measure, topographic mapping is concerned with the construction of low-dimensional maps of the data where the distances on the map between data items reveal the structure of their relationships. Both approaches can provide valuable tools in initial mining complex data sets. The Self-Organizing map (SOM) [17] and its extensions are examples of tools for topographic map construction.

Several probabilistic analogues of SOM have been proposed, seeking to address some of the limitations of SOM¹ (e.g. [2]). One formulates a mixture of Gaussian densities constrained on a smooth image of a low-dimensional latent space. Each point in the latent space is mapped via a smooth non-linear mapping to its image in the high-dimensional data space. This image plays the role of the mean of a local spherical Gaussian noise model capturing data points in its vicinity. Due to the smoothness of the mapping, when latent points are mapped to the high-dimensional data space, they retain their local neighborhood structure. The model is trained by maximizing the model likelihood for the given data. Once the model has been trained, each data item is represented in the latent space by a point given as the mean position of latent points weighted by their posterior probabilities (responsibilities) given the data item.

Such approaches provide a framework that is readily extensible to structured data by adopting alternative formulations of noise models in the place of Gaussian densities. Such extensions have been proposed in [34] for the visualization of symbolic sequences and in [10] for the visualization of tree-structured data, where hidden Markov models and hidden Markov tree models play the role of local noise models respectively. In the same spirit, in [9] we formulated a

¹ E.g. the absence of a principled cost function.

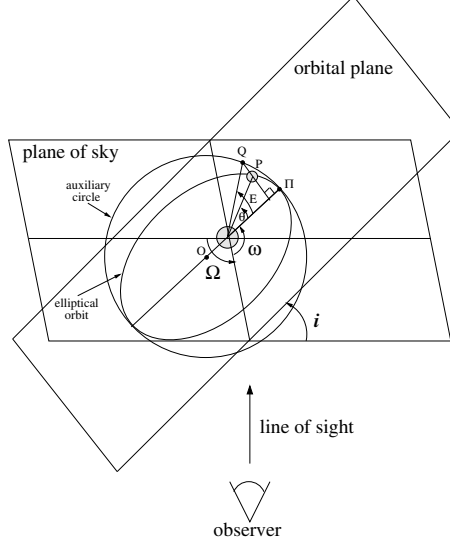


Fig. 4. Angles orientating the orbital plane with respect to the plane of sky, and angles associated with the orbits. Adapted from [14].

generative probabilistic model for clustering and topographic mapping of light curves from eclipsing binary stars. We adopted a specialized noise model based on the physical Kepler model that describes the generation of such light curves.

The core of the noise model is given by the basics of two-body gravitational dynamics, i.e. the Kepler model parametrized by the masses of the two stars (primary mass + mass ratio), the eccentricity of their relative orbit and other dynamical parameters, plus the geometry of viewing, i.e., the inclination of the orbit to the line of sight, as illustrated in figure 4. The overall model - a constrained and unconstrained mixture of such parametrized probabilistic models of fluxes from binary star models - can then be readily used for clustering and topographic mapping of binary systems, respectively. Since all the parameters are physically interpretable with universally known priors, the overall models were fitted in the MAP (maximum a-posteriori) framework.

The methodology was applied to a dataset of light curves from two resources:

- (1) The *Catalog and Archive of Eclipsing Binaries* at <http://ebola.eastern.edu/> which is a collection of light curves of the brightest eclipsing binary stars in the sky, compiled from the literature, based on several decades of photometric data;
- (2) *All Sky Automated Survey* at <http://archive.princeton.edu/~asas/>. This is a survey of the brightest stars visible in the sky in the Southern hemisphere from a dedicated telescope in Chile.

Figure 5 presents clustering of the fluxes into 6 clusters. For each cluster, we show all lightcurves assigned to it. The clusters represent a natural grouping of binary

systems into 4 groups according to their physical properties. The first three clusters group binaries with low eccentricity (symmetrical normalized fluxes). Cluster 2 represents binary systems with smaller primary and secondary stars and well-separated systems with large semi-major axis. On the other hand, cluster 3 groups binary systems with larger primary and secondary stars and systems with small separation between the primary and secondary stars. Cluster 1 represent binary systems in between the extremes characterized by clusters 2 and 3. Cluster 4 collects binaries of higher eccentricity. The binaries in cluster 4 have smaller primary and secondary stars and/or a large semi-major axis. Clusters 5 and 6 play the role of “garbage collectors” and collect binary systems that are either ‘atypical’ or represent low frequency binary star types in the data set that could not be naturally grouped together based on the given data.

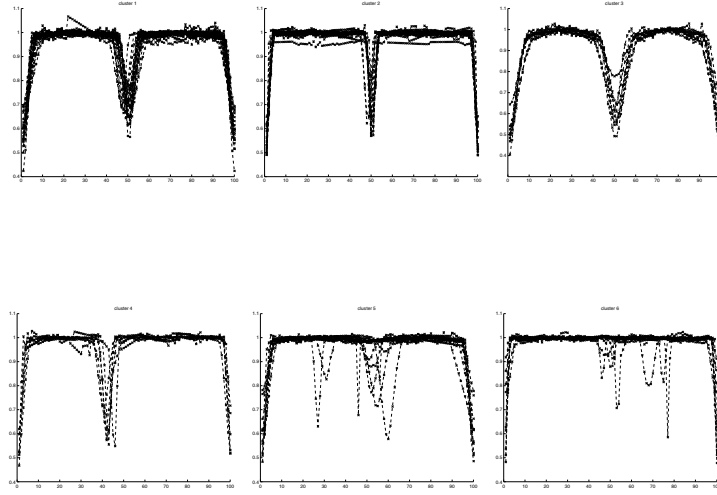


Fig. 5. Lightcurves belonging to each of the 6 clusters ordered from top left to bottom right

Visualization of the binary systems through their topographic mapping is shown in figure 6. The 2-dimensional manifold of binary star models (visualization space) is represented by the regular grid of fluxes generated at the corresponding positions on the visualization space. Superposed on this grid are some projections of the fluxes from real binary systems. A detailed discussion of this visualization is beyond the scope of this paper, we only mention that to the best of our knowledge, our methodology represents the first fully principled approach to physics based data driven automated clustering and topographic mapping of binary complexes.



Fig. 6. Topographic mapping of a set of binary systems

5 Conclusions

In view of the assembly of large (petabytes) astronomical data archives, as in the Virtual Observatory, where observations across the electromagnetic spectrum (i.e. radio, infrared, optical, X-rays etc) of planets, stars and galaxies, the need has arisen for the application of computational intelligence to the automated analysis of these large multivariate datasets. These data will be in the form of images and spectra, and the combination of a large number of observations of the same parts of sky will give rise to measurements of variability in the form of multivariate time series, as well as various catalogues of derived parameters, whether directly derived from the observations or through detailed physical modelling of the observations.

Fast automated methods of classification, parameter extraction, characterisation and visualisation of multi-dimensional and multi-type datasets are necessary for advances in this field, and they have to be tailored to the particular problems with the help of the knowledge of the domain. One particular requirement of astrophysics is the handling of systematic and random errors introduced by the process of measurement as well as the inherent diversity of various systems. Learning algorithms are essential in characterising the inputs of experts, since

many sub-fields still depend on visual characterisation of features in the observed morphology, spectra and time-series acquired in astronomical datasets. In this review, we have shown three diverse examples where such methods have been successful in understanding the physical problem, which in turn have helped in the development of new approaches in computational intelligence.

References

1. Banerji, M., Lahav, O., Lintott, C.J., Abdalla, F.B., Schawinski, K., Bamford, S.P., Andreescu, D., Murray, P., Raddick, M.J., Slosar, A., Szalay, A., Thomas, D., Vandenberg, J.: Galaxy Zoo: reproducing galaxy morphologies via machine learning. *Monthly Notices of the Royal Astronomical Society* 406, 342–353 (2010)
2. Bishop, C.M., Svensén, M., Williams, C.K.I.: GTM: The generative topographic mapping. *Neural Computation* 10(1), 215–234 (1998)
3. Burud, I., Magain, P., Sohy, S., Hjorth, J.: A novel approach for extracting time-delays from lightcurves of lensed quasar images. *Astronomy and Astrophysics* 380(2), 805–810 (2001)
4. Cuevas-Tello, J.C., Tiño, P., Raychaudhury, S., Yao, X., Harva, M.: Uncovering delayed patterns in noisy and irregularly sampled time series: An astronomy application. *Pattern Recognition* 43(3), 1165–1179 (2010)
5. Cuevas-Tello, J.C., Tiño, P., Raychaudhury, S.: How accurate are the time delay estimates in gravitational lensing? *Astronomy and Astrophysics* 454, 695–706 (2006)
6. Cuevas-Tello, J.C., Tiño, P., Raychaudhury, S.: A Kernel-Based Approach to Estimating Phase Shifts Between Irregularly Sampled Time Series: An Application to Gravitational Lenses. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) *ECML 2006. LNCS (LNAI)*, vol. 4212, pp. 614–621. Springer, Heidelberg (2006)
7. Eigenbrod, A., Courbin, F., Vuissoz, C., Meylan, G., Saha, P., Dye, S.: COSMOGRAIL: The COSmological MONitoring of GRAVItational Lenses. I. How to sample the light curves of gravitationally lensed quasars to measure accurate time delays. *Astronomy and Astrophysics* 436, 25–35 (2005)
8. Fardal, M.A., Babul, A., Geehan, J.J., Guhathakurt, P.: Investigating the andromeda stream - ii. orbital fits and properties of the progenitor. *Monthly Notices of the Royal Astronomical Society* 366, 1012–1028 (2006)
9. Gianniotis, N., Tiño, P., Spreckley, S., Raychaudhury, S.: Topographic Mapping of Astronomical Light Curves via a Physically Inspired Probabilistic Model. In: Alippi, C., Polycarpou, M., Panayiotou, C., Ellinas, G. (eds.) *ICANN 2009, Part I. LNCS*, vol. 5768, pp. 567–576. Springer, Heidelberg (2009)
10. Gianniotis, N., Tiño, P.: Visualisation of tree-structured data through generative probabilistic modelling. In: Verleysen, M. (ed.) *European Symposium on Artificial Neural Networks*, pp. 97–102. D-Facto (2007)
11. Girolami, M., He, C.: Probability density estimation from optimally condensed data samples. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 1253–1264 (2003)
12. Guinan, E.F., Engle, S.G.: The brave new world of binary star studies. *Astrophysics Space Science* 304, 5–11 (2006)
13. Harva, M., Raychaudhury, S.: Bayesian estimation of time delays between unevenly sampled signals. In: *IEEE International Workshop on Machine Learning for Signal Processing*, pp. 111–116. IEEE (2006)

14. Hilditch, R.W.: An introduction to close binary stars. Cambridge University Press (2001)
15. Inada, N., Oguri, M., Becker, R.H., Shin, M.-S., Richards, G.T., Hennawi, J.F., White, R.L., Pindor, B., Strauss, M.A., Kochanek, C.S., Johnston, D.E., Gregg, M.D., Kayo, I., Eisenstein, D., Hall, P.B., Castander, F.J., Clocchiatti, A., Anderson, S.F., Schneider, D.P., York, D.G., Lupton, R., Chiu, K., Kawano, Y., Scranton, R., Frieman, J.A., Keeton, C.R., Morokuma, T., Rix, H.-W., Turner, E.L., Burles, S., Brunner, R.J., Sheldon, E.S., Bahcall, N.A., Masataka, F.: The Sloan Digital Sky Survey Quasar Lens Search. II. Statistical Lens Sample from the Third Data Release. *Astronomical Journal* 135, 496–511 (2008)
16. Kochanek, C., Schechter, P.: The Hubble constant from gravitational lens time delays. In: Freedman, W.L. (ed.) *Measuring and Modeling the Universe, from the Carnegie Observatories Centennial Symposia*, vol. 2, p. 117. Cambridge University Press (2004)
17. Kohonen, T.: The self-organizing map. *Proceedings of the IEEE* 78(9), 1464–1480 (1990)
18. Kundic, T., Turner, E.L., Colley, W.N., Gott-III, J.R., Rhoads, J.E., Wang, Y., Bergeron, L.E., Gloria, K.A., Long, D.C., Malhorta, S., Wambsganss, J.: A robust determination of the time delay in 0957+561A,B and a measurement of the global value of Hubble’s constant. *Astrophysical Journal* 482(1), 75–82 (1997)
19. Lahav, O., Naim, A., Buta, R.J., Corwin, H.G., de Vaucouleurs, G., Dressler, A., Huchra, J.P., van den Bergh, S., Raychaudhury, S., Sodre Jr., L., Storrie-Lombardi, M.C.: Galaxies, Human Eyes, and Artificial Neural Networks. *Science* 267, 859–862 (1995)
20. McConnachie, A.W., Irwin, M.J., Ibata, R.A., Dubinski, J., Widrow, L.M., Martin, N.F., Côté, P., Dotter, A.L., Navarro, J.F., Ferguson, A.M.N., Puzia, T.H., Lewis, G.F., Babul, A., Barmby, P., Bienaymé, O., Chapman, S.C., Cockcroft, R., Collins, M.L.M., Fardal, M.A., Harris, W.E., Huxor, A., Mackey, A.D., Peñarrubia, J., Rich, R.M., Richer, H.B., Siebert, A., Tanvir, N., Valls-Gabaud, D., Venn, K.A.: The remnants of galaxy formation from a panoramic survey of the region around M31. *Nature* 461, 66–69 (2009)
21. Nolan, L.A., Harva, M.O., Kabán, A., Raychaudhury, S.: A data-driven Bayesian approach for finding young stellar populations in early-type galaxies from their ultraviolet-optical spectra. *Monthly Notices of the Royal Astronomical Society* 366, 321–338 (2006)
22. Nolan, L.A., Raychaudhury, S., Kabán, A.: Young stellar populations in early-type galaxies in the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 375, 381–387 (2007)
23. Oguri, M.: Gravitational Lens Time Delays: A Statistical Assessment of Lens Model Dependences and Implications for the Global Hubble Constant. *Astrophysical Journal* 660, 1–15 (2007)
24. Ovaldsen, J.E., Teuber, J., Schild, R.E., Stabell, R.: New aperture photometry of QSO 0957+561; application to time delay and microlensing. *Astronomy and Astrophysics* 402(3), 891–904 (2003)
25. Pelt, J., Hjorth, J., Refsdal, S., Schild, R., Stabell, R.: Estimation of multiple time delays in complex gravitational lens systems. *Astronomy and Astrophysics* 337(3), 681–684 (1998)
26. Pelt, J., Kayser, R., Refsdal, S., Schramm, T.: Time delay controversy on QSO 0957+561 not yet decided. *Astronomy and Astrophysics* 286(1), 775–785 (1994)
27. Pelt, J., Kayser, R., Refsdal, S., Schramm, T.: The light curve and the time delay of QSO 0957+561. *Astronomy and Astrophysics* 305(1), 97–106 (1996)

28. Pijpers, F.P.: The determination of time delays as an inverse problem - the case of the double quasar 0957+561. *Monthly Notices of the Royal Astronomical Society* 289(4), 933–944 (1997)
29. Pindor, B.: Discovering Gravitational Lenses through Measurements of Their Time Delays. *Astrophysical Journal* 626, 649–656 (2005)
30. Press, W.H., Rybicki, G.B., Hewitt, J.N.: The time delay of gravitational lens 0957+561, I. Methodology and analysis of optical photometric Data. *Astrophysical Journal* 385(1), 404–415 (1992)
31. Refsdal, S.: On the possibility of determining the distances and masses of stars from the gravitational lens effect. *Monthly Notices of the Royal Astronomical Society* 134, 315–319 (1966)
32. Saha, P.: Gravitational Lensing. In: Murdin, P. (ed.) *Encyclopedia of Astronomy and Astrophysics*. Nature Publishing, London (2001)
33. Sun, J., Kaban, A., Raychaudhury, S.: Robust Mixtures in the Presence of Measurement Errors. In: *International Conference on Machine Learning*, pp. 847–854 (June 2007)
34. Tiño, P., Kaban, A., Sun, Y.: A generative probabilistic approach to visualizing sets of symbolic sequences. In: *KDD 2004: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 701–706. ACM Press (2004)
35. Vincent, P., Bengio, Y.: Manifold parzen windows. *Advances in Neural Information Processing Systems* 15, 825–832 (2003)
36. Wadadekar, Y.: Estimating Photometric Redshifts Using Support Vector Machines. *Publications of the Astronomical Society of the Pacific* 117, 79–85 (2005)
37. Walsh, D., Carswell, R.F., Weymann, R.J.: 0957 + 561 A, B - Twin quasistellar objects or gravitational lens. *Nature* 279, 381–384 (1979)
38. Wang, X., Tiño, P., Fardal, M., Raychaudhury, S., Babul, A.: Fast Parzen Window Density Estimator. In: *Proceedings of the International Joint Conference on Neural Networks - IJCNN 2009, Atlanta, Georgia*, pp. 473–475. IEEE Computer Society (2009)
39. Chen, S., Hong, X., Harris, C.J.: A forward-constrained regression algorithm for sparse kernel density estimation. *IEEE Transactions on Neural Networks* 19, 193–198 (2008)
40. Zhang, K., Kwok, J.T.: Simplifying mixture models through function approximation. *Advances in Neural Information Processing Systems* 15, 825–832 (2006)