Transient Detection Modelling for Gravitational-wave Optical Transient Observer (GOTO) Sky Survey

A. B. Tabacolde[1], T. Boongoen[1,2], N. Iam-On[1,2], J. Mullaney[3],U. Sawangwit[4] and K. Ulaczyk[5]
[1]School of Information Technology, Mae Fah Luang University, Thailand
[2] IQ-D Research Unit, School of Information Tehcnology, Mae Fah Luang University, Thailand
[3]Department of Physics and Astronomy, University of Sheffield, Sheffield, UK
[4]National Astronomical Research Institute of Thailand, Chiang Mai, Thialand
[5]Department of Physics, University of Warwick, Coventry, UK

ABSTRACT

Given the advancement of data acquisition and telescope technology, astronomy has joined the global trend of big data and artificial intelligence in recent years. The objective of GOTO is to identify optical counterparts to gravitational wave detections. This requires obtaining many images of the sky every night, which are then systematically processes and analysed to deliver 40-million observed sources. These sources are then compared against a reference set such that new bright sources can be extracted and used to form a set of counterpart candidates. Most of the candidates will not represent real cases, with their detected changes in brightness caused by errors in data collection and/or pre-processing. To this end, the handful of real candidates have to be correctly sifted from the false-positives to allow astronomers to effectively employ follow-up observations to verify their truth. The aforementioned problem falls nicely into data classification, where multiple physical measurements of candidates are explicated as independent variables with labels given by experts as the class variable. This research is set to explore conventional techniques to analyze this specific dataset, from data preparation through to model development and evaluation. The outcome of our research not only provides a baseline for future developments, but also provides a thorough review of data characteristics. It will be also proving useful for the GOTO project in terms of shaping the approach to acquire and store data.

Keywords

sky survey, transient detection, classification, imbalance class

1. INTRODUCTION

Transient astronomical events (referred to as simply ‘transients’ within the astronomical community), have recently emerged as one of the most exciting areas of study in astronomy [3]. Developments in telescope and detector technologies means that scouting deep space for transient sources no longer presents a significant challenge [6]. Recently, this has led to interdisciplinary collaborations, in order to help astronomers analyse the vast amounts of digital information received from sky surveys carried out by ground and space-based telescopes and, most recently, from the advanced Laser Interferometer Gravitational-wave Observatory (aLIGO [13]) which has detected gravitational wave transients.

Astronomers’ ability to detect transient events paves the way to the study of rare classes of extreme events. Examples include mergers of dense objects (such as neutron stars and black holes), the tidal disruption of stars by dormant super massive black holes, breakout shocks of Type II SNe, or megaflares on normal main sequence stars [5]. The Gravitational-wave Optical Transient Observer (GOTO[1]) is among the new breed of telescopes dedicated to detecting such events, with its primary science goal being that of identifying optical counterparts to gravitational events detected by aLIGO. GOTO is an international collaboration led by University of Warwick of UK and Monash University of Australia, with its facility housed at Roque de Los Muchachos observatory on La Palma, Canary Island. The observatory consists of an array of four state-of-the-art 0.5m-aperture, wide-field optical telescopes which can respond to alerts coming from gravitational wave detectors, i.e., LIGO and VIRGO. GOTO’s primary science goal necessitates a rapid response in order to track down optical counterparts to gravitational wave events, since the former are likely to fade on short timescales (hours to days). As such, GOTO has been designed specifically to detect these optical signatures as quickly as possible so as to provide astronomers with as much information on these sources before they rapidly fade.

Each night that GOTO observes, it delivers large volumes of data in the form of optical images. Each image contains roughly 20,000 astronomical sources, and GOTO will deliver roughly 400 such images per night. These images need to be quickly analysed in order to provide astronomers with useful, legible data. The management of this magnitude of data is one of the major challenges facing GOTO and all other wide-field sky surveys. Considering the enormous amounts of received data, statistical algorithms are basically needed to model certain characteristics in order to interpret the data as accurately as possible. Furthermore, with such large amounts of astronomical data is being delivered and processed every night, artificial intelligence is likely the only

ICMLC-2018 Macau, China

[1] https://goto-observatory.org/project/

route via which this can can be meaningfully classified and exploited. To this end, the current research project focuses on the problem of new source detection, which can be modelled as a binary classification task. The conventional data analysis methodology [1] is employed for data preparation, model development and assessment. To generalize the finding, the study includes classification models belonging to both single model and ensemble categories [16]. The well-known SMOTE technique [15] is also exploited to handle the challenge of class imbalance raised by the data.

The rest of this paper is organized as follows. Section 2 presents previous works related to the problem investigated by this research. Details of data analysis methodology are given in Section 3, including data generation, preparation and model creation. Following that, Section 4 provides the experimental results, with original and modified datasets. The paper is concluded in Section 5 with perspective of future work.

2. RELATED WORK

The identification and classification of sources in astronomical data is critical to astronomical research. Throughout this process there are several factors to consider such as source variability, the effects of noise in the data, fluctuations or maybe system errors. It is in this way that machine learning algorithms have been utilized successfully in classification processes. One of these is a supervised classification, wherein a set of data is trained and basically known as classifier, taken from an identified class of data source of which the functions are studied to determine the relationship between the observed data and the source of the data. Once a suitable training set has been established, it can be useful for predicting any future sources or objects based on the observed data. This makes machine learning an automated classification engine, where it can segregate false data sources from real events. The trustworthiness of this classification engine however, relies on a dataset which is properly trained and validated. In so doing, the classifier can provide an authentic result wherein the false-positives and false-negative rates are usually small [12]. The followings give a brief list of recent studies that apply classification modeling to various problems in the astronomy domain.

□ Random Forest as an ensemble of decision trees applied to detect point-source transients and moving ob- jects in the Dark Energy Survey Supernova (DES-SN) program [7].

□ A number of conventional classifiers like KNN and neural networks have been exploited for the task of photometric supernova classification [14].

□ Deep-HiTS [10] is a rotation-invariant convolutional neural network model introduced to classify images of transient candidates for the High cadence Transient Survey (HiTS).

□ Unlike the previous studies that rely mostly on an automated classification process, a joint force between human and machine [8] has been proposed to find transient sources in data from the Pan-STARRS1 telescope.

To train a data set in a new survey, foregoing simulated or limited commissioning data are usually considered. Furthermore, good quality and large-sized training sets can enhance success rates since transient events are usually very subtle and most of data detected are spurious. In addition, the determination of informative features has proven to be critical for classifiers to establish accurate feature-class relations, which lead to an effective classification [4]. This means that using a good set of data features will enhance the authenticity of the data interpretation. Fig 1(a) and (b) show the variance between bogus and real sources with simulated GOTO data (see Section 3).

Discriminating the real from the bogus data sources posed a major challenge for this study. On the other hand, with sufficient archived samples from the database, it has shed light on the identification procedure. This study is significant in as much as it contributes to the idea on how machine learning works in handling and classifying large amounts of astronomical data (i.e., astronomical 'Big Data'). Though the accuracy is not perfect, it is very evident that the success rate of the result in the data segregation is substantial. This provides insights into how to manipulate the training set in order to attain a more stable data classifier.

3. METHOD

The framework of this study follows the sequencing of procedures from data acquisition, through data processing and, ultimately, to final analysis. Basically, GOTO provides a massive source of raw data of detected optical sources. It is significant that the data goes through a preliminary preprocessing step to improve its quality before being delivered to the various types of supervised learning classification techniques employed here (Naive Bayes (NB), K-Nearest Neighbor (KNN), Decision Tree (J48) and Random Forest (RF)). This setting is to generalize the outcome, hence providing a reliable conclusion for practice and further development.

3.1 Data Acquisition

In this proof-of-concept study, we exclusively analyse simulated (as opposed to real) GOTO images. The benefit of using simulated images is that we know *a priori* which sources in the simulated images are real transients. As such, they provide an ideal testbed for our supervised machine learning algorithms, as we are readily able to compare the output of the algorithms against the known inputs. The pitfalls of using simulated data, however, are that they can provide somewhat idealised approximations of the real data. Thankfully, there exists off-the-shelf software packages that can simulate astronomical images, including approximations for the most commonly encountered complications, such as background noise and the point spread functions (PSF) of sources. For this study, we used the SkyMaker software [2] to generate our simulated images.

As its primary input, SkyMaker accepts a list of sources (i.e., stars, galaxies) cointaining the position (i.e., right ascension, RA, and declination, Dec) and brightness of each source. These three pieces of information is all that is required for stars (i.e, point sources). Galaxies in SkyMaker are represented by two cospatial ellipses (one for the bulge, the other for the disk) which are described by an additional seven parameters (the ratio of bulge-to-total light, bulge radius, bulge aspect ratio,bulge orientation on sky, disk radius, disk inclination and disk orientation on sky). For our simulations, we generated a source lists by querying two separate databases. For stars brighter than 17th magnitude, we used the USNO CCD Astrograph Catalog (UCAC) database, whereas for stars and galaxies fainter than 17th magnitude, we used the Sloan Digital Sky Survey (SDSS). This approach of combining two separate catalogues to generate our input lists was used to increase the dynamic range of our simulated images since bright stars saturate the SDSS detector and are thus under-represented in this catalogue, while UCAC does not go sufficiently deep for our purposes.

We queried these databases for all sources that would be covered by a single observation by one of the GOTO telescopes of a given patch of sky (given by its central coordinates). Each source's position (RA, Dec) was converted into a pixel coordinate (x,y) by referring to the on-sky position of the central pixel and the pixel scale (i.e., the on-sky angular size of each pixel, which is a constant 1.24 arcseconds per pixel). For the UCAC sources, the V-band magnitude was used, whereas G-band magnitudes were used for SDSS sources. At this stage, we simulate galaxies as a simple disk (i.e., not a bulge+disk combination), and thus only provide the three additional parameters that SkyMaker uses to simulate galaxy disks. Since we are primarily interested in transient sources, we do not expect that only simulating disks will have any significant impact on our study.

In addition to the source list, SkyMaker also requires an input configuration file. This provides the software with information such as the type of simulation that is required (e.g., include background noise or not) and the characteristics of the telescope. For the latter, the most important of these are the saturation level of the pixels (set to 65,535), the zeropoint of the telescope (i.e., the magnitude of a star that would result in one count per second; 23.5), PSF size (see below), pixel size (1.24 arcsec per pixel), CCD size in pixels (8176×6132). To increase the realism of the simulations, we allowed the PSF size (i.e., full-width half maximum, or FWHM) to vary randomly between observations, ranging from 0.8 to 3 arcseconds.

In order to simulate transient sources, we simulate two observations for each patch of sky. In the second observation, we inject new sources, radomly distributed across simulated image with brightnesses chosen randomly from a uniform distribution ranging from magnitude 14 to 19. Each simulated image is then processed using the LSST software stack [17], adapted to handle our simulated images. We use the output from the image differencing component of the stack as input to our machine learning algorithms.

3.2 Data Preparation

The dataset presented in this study contains a total of 7,891 records with 30 attributes. Each of these data instances has a variable flag, which indicates whether the simulated source is a transient (class 1) or not (class 0). There are 7, 873 data that belongs to class 0 and only 18 belongs to class 1. Examples of sources corresponding to class 0 and class 1 data are shown in Fig 1. Before setting the data into test, it is essential to ensure its quality, since this determines the reliability of the outcome and applications. The followings illustrate the flow in which the simulated dataset is pre-processed prior to the learning stage.

Data cleaning: Upon further examination of the original data, it was observed that there were incomplete values in it. So data cleaning is the next step in the process. The basic way to deal with bad and missing values is to eliminate or ignore certain records from the dataset. As a result, the data size was reduced to 5,989, resulting to a total of 5,973 which belongs to class 0 and the remaining 16 to class 1.

Data transformation: Selecting the set of informative attributes or features is crucial for classification modelling. The chosen ones will affect fairly, in terms of the accuracy of the predictive model that will be formulated. With the initial of 30 features, six of them (i.e., id, parent_id, RA, DEC, SdssCentroid_x, SdssCentroid_y) are inherently uninformative to learning models. They are excluded from the dataset, decreasing the number of features to 24 (see the left column of Fig 2 for the list of these).

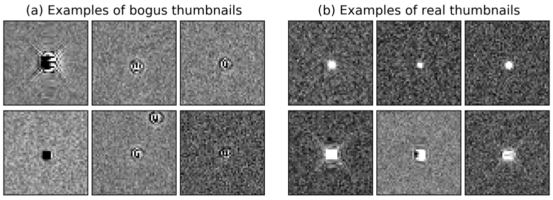
(a) Examples of bogus thumbnails
(b) Examples of real thumbnails

Figure 1: Image examples of bogus and real sources

Nonetheless, the resulting set of features is not that remarkable, so it is essential to transform the features in order to obtain a more consistent and stable data once it goes through the classification modelling process. For instance, some attributes expressing absolute positions on an image-based observation can be aggregated to form distance, width or height, which can be more informative. Fig 2 summarizes the process of transformation from the previous set to the final collection of 16 features.

Data normalization: After the data has been cleaned and pre-processed, the next stage is the data transformation by normalization using the min-max normalization technique. In particular, feature values are scaled with the range of [0, 1], using the following: $v_x^* = \frac{v_x - min_x}{max_x - min_x}$, where v_x is a value of feature f_x and v_x^* is its corresponding in the normalized scale, min_x and max_x are the minimum and maximum values of f_x..

3.3 Classification Modeling

One of the challenges in this research is handling the imbalance data. The number of instances of real sources is much smaller than the artifacts. The ratio between two classes (1 as real sources and 0 otherwise) can be as high as 1 to 400 or 500. As such, accuracy levels obtained from conventional classifiers are generally high for class 0 and low for the other, which is the event to be detected. Our first attempt in solving the imbalance data is to divide the majority class (i.e., class 0) into smaller portions and test per set with classifiers. This leads to observations with different $x : y$, where x denotes the number of instances belonging to class 1 and y is the number of class 0 instances. For instance, the original ratio is 16 : 5973. Specific to class 1, Fig 3 presents the initial results in terms of true positives $\in \{0, 16\}$. From this figure, KNN and J48 classificaiton models commonly perform better as the difference between two classes become smaller (i.e., 16 : 200 to 16 : 600). Note that the goodness of KNN drops faster than the other, as the context of local neighbors is overwhelmed with class 0 instances.

One of the typical approaches to tackle imbalanced-class problem is at data level [11]. It modifies data distribution by either reducing the number of instances from the majority class, or increasing the number of instances belonging to other. Given the aforementioned results that indicate the possible advantage of modifying the original data distribution, the SMOTE technique [15] that has been widely

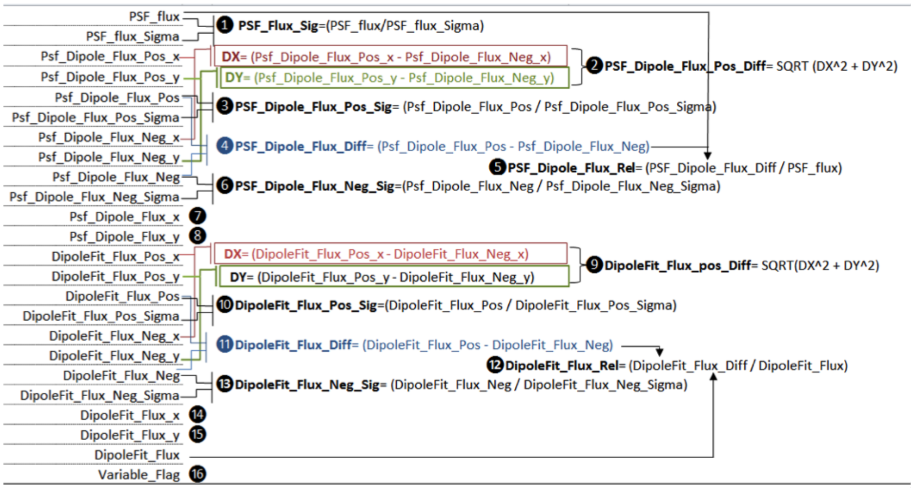
PSF_flux
PSF_flux_Sigma
Psf_Dipole_Flux_Pos_x
Psf_Dipole_Flux_Pos_y
Psf_Dipole_Flux_Pos
Psf_Dipole_Flux_Pos_Sigma
Psf_Dipole_Flux_Neg_x
Psf_Dipole_Flux_Neg_y
Psf_Dipole_Flux_Neg
Psf_Dipole_Flux_Neg_Sigma
Psf_Dipole_Flux_x 7
Psf_Dipole_Flux_y 8
DipoleFit_Flux_Pos_x
DipoleFit_Flux_Pos_y
DipoleFit_Flux_Pos
DipoleFit_Flux_Pos_Sigma
DipoleFit_Flux_Neg_x
DipoleFit_Flux_Neg_y
DipoleFit_Flux_Neg
DipoleFit_Flux_Neg_Sigma
DipoleFit_Flux_x 14
DipoleFit_Flux_y 15
DipoleFit_Flux
Variable_Flag 16
1 PSF_Flux_Sig=(PSF_flux/PSF_flux_Sigma)
DX= (Psf_Dipole_Flux_Pos_x - Psf_Dipole_Flux_Neg_x)
DY= (Psf_Dipole_Flux_Pos_y - Psf_Dipole_Flux_Neg_y)
2 PSF_Dipole_Flux_Pos_Diff= SQRT (DX^2 + DY^2)
3 PSF_Dipole_Flux_Pos_Sig= (Psf_Dipole_Flux_Pos / Psf_Dipole_Flux_Pos_Sigma)
4 PSF_Dipole_Flux_Diff= (Psf_Dipole_Flux_Pos - Psf_Dipole_Flux_Neg)
5 PSF_Dipole_Flux_Rel= (PSF_Dipole_Flux_Diff / PSF_flux)
6 PSF_Dipole_Flux_Neg_Sig=(Psf_Dipole_Flux_Neg / Psf_Dipole_Flux_Neg_Sigma)
DX= (DipoleFit_Flux_Pos_x - DipoleFit_Flux_Neg_x)
DY= (DipoleFit_Flux_Pos_y - DipoleFit_Flux_Neg_y)
9 DipoleFit_Flux_pos_Diff= SQRT(DX^2 + DY^2)
10 DipoleFit_Flux_Pos_Sig=(DipoleFit_Flux_Pos / DipoleFit_Flux_Pos_Sigma)
11 DipoleFit_Flux_Diff= (DipoleFit_Flux_Pos - DipoleFit_Flux_Neg)
12 DipoleFit_Flux_Rel= (DipoleFit_Flux_Diff / DipoleFit_Flux)
13 DipoleFit_Flux_Neg_Sig= (DipoleFit_Flux_Neg / DipoleFit_Flux_Neg_Sigma)

Figure 2: Details of data transformation.

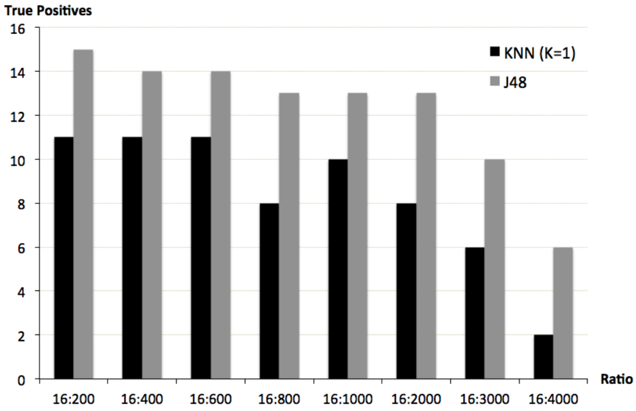

Figure 3: Initial class 1-specific true positives with respect to different ratios of instances.

accepted for such a task is employed to increase the size of the minority class.

4. RESULT

This section presents the design of experiments including the modification of dataset using SMOTE, followed by the experimental results obtained with different classifiers.

4.1 Experimental Design

The experiments are arranged with the following two cases, each of which will be assessed using 10-fold cross validation:

- Before SMOTE, with the original instance distribution between two classes (5,973 of class 0 and 16 of class 1)

- After SMOTE, with the modified instance distribution between the two classes (5,973 of class 0 and 2,112 of class 1). This is achieved by applying SMOTE to increase the number of instances belonging to class 1 (using the setting of K = 5 that is recommended by the original work).

Once the aforementioned sets of data are obtained, they are used to model classifiers. Four classical alternatives are exploited and evaluated for the present research: KNN (K Nearest Neighbors, K = 1), NB or Naive Bayes, J48 (decision tree), and RF or Random Forest (ensemble of trees), respectively. Note that the default ensemble size of RF is 10, while the size of randomly selected feature set is determined by $\log_2 D + 1$ (D is the number of features or 16 in this study). In addition, the classification results will be evaluated using the two classical indices of Precision (P) and Recall (R).

$$P = \frac{TP}{TP + FP}, \tag{1}$$

$$R = \frac{TP}{TP + FN}, \tag{2}$$

where TP, FP and FN represent true positives, false positives and false negatives, respectively.

4.2 Experimental Result

Given the aforementioned setting, the corresponding results are shown in Table 1. From these, the problem of imbalance class is obvious and significantly degrades the benefit of a decision tree model like J48. In fact, in the Before SMOTE case, the resulting tree structure is meaningless with only a root node, which is naturally dictated by class 0. As a result, both precision and recall for class 1 are zero. This situation becomes marginally better with RF that summarizes the predictions obtained from an ensemble of ten trees. Despite the class 1 precision being perfect, the corresponding recall is extremely low. In other words, RF is able to predict instances of class 1 correctly, but there are a lot more instances RF have not recognised as class 1. With this observation, another investigation is conducted for RF with an ensemble size of 100 as to further investigated the case. Precision and recall measures remain the same for both classes, thus setting the bound to which an ensemble technique can achieve.

In contrast, both J48 and RF become more effective for the binary classification in the After SMOTE case, with high precision and recall values. Fig 4 presents the top layers of a tree generated by J48, with the entire tree consisting of 74 leaf nodes at the maximum depth of 16. Note that RF can slightly enhance the quality measures of J48. Likewise, as the ensemble size of RF is enlarged to 100, better results are obtained (class 0: P = 0.999, R = 0.997, class 1: P = 0.992, R = 0.998).

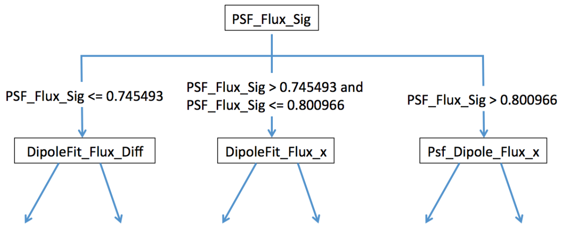

Figure 4: Example of decision tree generated by J48 in the After SMOTE case.

This trend of improvement is also witnessed with the other classifiers, KNN and NB. Specific to KNN, the classification process works extremely well for class 0 in both dataset variations. While it underperforms for class 1 in the original case, this same classifier is excellent in the other. As KNN relies on the prediction information held by neighbours of an instance under question, any local neighbouring context in the original data is simply overwhelmed by class 0 instances. Therefore, it is not a surprise to observe the difference between classifying two classes. As the previous condition is largely violated through simulated class 1 instances by SMOTE, KNN is able to show its true potential. One of the key factors to KNN is the number of nearest neighbours or K, which is set to one initially. Fig 5 presents the precision measures for both classes with respect to increasing values of K. Note that the recall measures remain pretty much the same for different K values investigated here. As such, the class 0 precision is rather consistent with the neighbouring size gets larger, while the suitable K should b around 1-4 for class 1.

Table 1: The quality of four classifiers as class-specific Precision (P) and Recall (R) measures, with respect to two evaluation cases (Before and After SMOTE).

Classifier	Class	Index	Before SMOTE	After SMOTE
NB	0	P	0.998	0.911
		R	0.672	0.251
	1	P	0.005	0.305
		R	0.563	0.931
KNN	0	P	0.998	0.999
		R	0.999	0.995
	1	P	0.25	0.986
		R	0.125	0.997
J48	0	P	0.997	0.993
		R	1	0.989
	1	P	0	0.969
		R	0	0.979
RF	0	P	0.997	0.998
		R	1	0.997
	1	P	1	0.991
		R	0.063	0.995

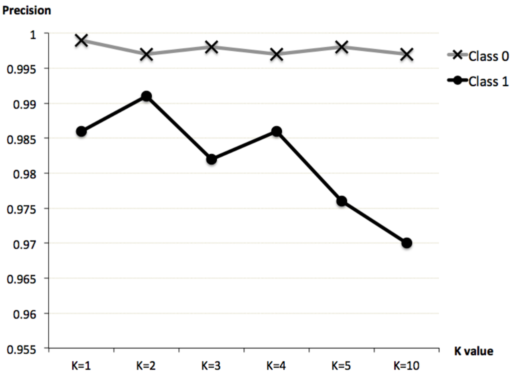

Figure 5: Precision measures of KNN with different K values.

Given the results from the NB classifier shown in Table 1, the data modification using SMOTE lefts the predictive performance of class 1, with the precision hugely increases from 0.005 to 0.305 and the recall from 0.563 to 0.931. Nonetheless, the downturn occurs with those of class 0, such that the precision is slightly reduced while the recall is lower by more than half. Unlike the KNN model that is based on local information, NB can be considered as a global counterpart as it concludes class-specific probabilities on a set of feature-class statistics summarized across the whole dataset. Yet, NB is not capable of differentiating between features like J48, which selects the most informative ones to form top layers of a tree. The modified SMOTE's class distribution may improve the discriminative power of a few features, but not all. As they are considered independent and blindly integrated, NB cannot raise its performance up to the level the others exhibit. To provide an alternative interpretation to NB, another classifier named Bayesian Network or BN [9] has also been investigated with the following results: Before SMOTE (class 0: P = 0.997, R = 1; class 1: P = 0, R = 0.) and After SMOTE (class 0: P = 0.994, R = 0.978; class 1: P = 0.941, R = 0.982). It is noteworthy that BN encodes relations between different features through network representation, while such a relationship is totally ignored by the NB.

These are in line with the statistics observed with other classifiers included in the current research. It strongly suggests that the core assumption of indepence among features causes NB the ability to deal with this imbalance problem accuracetly. Besides the obervations discussed thus far, the final point worth examining is the association between the size of simulated instances for the minority class and classification quality. To answer this, Fig 6 and 7 present precision values of the four classifiers specific to class 0 and class 1, which are categorized by three sizes of the minority class (class 1 = 1,056, class 1 = 2,112 and class 1 = 4,224). Note that the results in the instance that the size of class 1 equals 2,112 has been given in Table 1.

According to these figures, having more instances of the minority class helps to boost the classification performance of most classifiers. This is clear for class 1, even with the NB classifier that reaches the value of 0.465. But for the case of class 0, the precision of NB gets lower as the minority class becomes larger, while the opposite takes place for other models. Based on these, it is possible to recommend expanding the size of minority class to be roughly the same

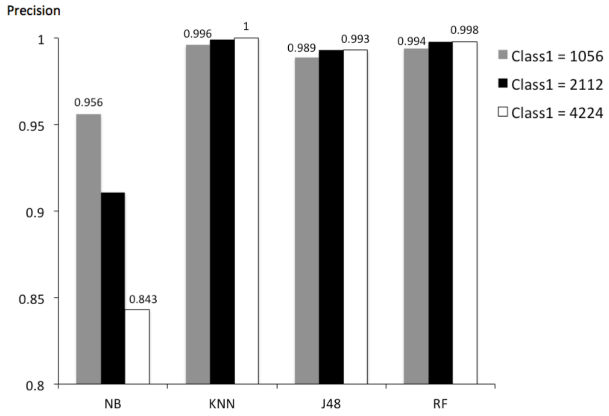

Figure 6: Precision values (for class 0) with different sizes of the minority class.

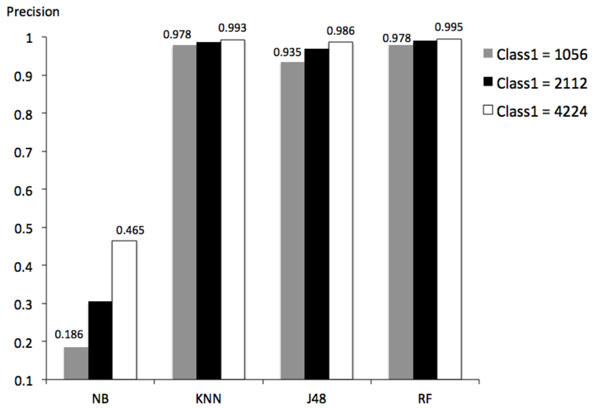

Figure 7: Precision values (for class 1) with different sizes of the minority class.

as the majority's. And to get the best out of such a setting, a classifier should be able to be selective on features or bring local information into consuderation.

5. CONCLUSION

This paper has presented a classification modeling for transient detection, as the first step to explore the full potentail of data acquired by the GOTO project. The study exhibits processing stages in accordance with the standard methodology of data analysis. With the original dataset, the problem of imbalance class has been recognized and resolved using a data-level approach. In particular, the SMOTE technique is used to simulate additional instances of the minority class. As such, the quality of several conventional classifiers have been generally improved, with the preferred model characteristics leaning towards the use of neighboring information or relationships between features. Given this outcome, one of the obvious future works is to generalize the current results with more data. In addition, other techniques belonging to the data-level approach to imbalance classification are to be assessed for the current task. Of course, other approaches such as algorithm-level and ensemble would also provide the interesting research areas to be fully studied.

6. ACKNOWLEDGMENTS

This work is funded by ST/P005594/1 - Newton STFC-NARIT: Using astronomy surveys to train Thai researchers in Big Data analysis. The authors are grateful to the members of the project team for their contribution, whilst taking full responsibility for the views expressed in this paper.

7. REFERENCES

[1] C. C. Aggarwal. *Data Mining*. Springer, 2013.

[2] E. Bertin. SkyMaker: astronomical image simulations made easy. *Memorie della Societa Astronomica Italiana*, 80:422, 2009.

[3] L. du Buisson et al. Machine learning classification of SDSS transient survey images. *Monthly Notices of the Royal Astronomical Society*, 454(2):2026–2038, 2015.

[4] A. D'Isanto et al. An analysis of feature relevance in the classification of astronomical transients with machine learning methods. *Monthly Notices of the Royal Astronomical Society*, 457(3):3119–3132, 2016.

[5] A. L. Bailer-Jones et al. Finding rare objects and building pure samples: probabilistic quasar classification from low-resolution gaia spectra. *Monthy Notices of the Royal Astronomical Society*, 391(4):1838–1853, 2008.

[6] A. M. Meisner et al. Searching for Planet Nine with coadded wise and neowise-reactivation images. *The Astronomical Journal*, 153(2):65, 2017.

[7] D. A. Goldstein et al. Automated transient identification in the dark energy survey. *The Astronomical Journal*, 150(3):82, 2015.

[8] D. E. Wright et al. A transient search using combined human and machine classifications. *Monthly Notices of the Royal Astronomical Society*, 472(2):1315–1323, 2017.

[9] Friedman et al. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.

[10] G. Cabrera-Vives et al. Deep-HiTS: Rotation invariant convolutional neural network for transient detection. *The Astrophysical Journal*, 836(1):97, 2017.

[11] G. Haixianga et al. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems With Applications*, 73:220–239, 2017.

[12] H. Brink et al. Using machine learning for discovery in synoptic survey imaging data. *Monthly Notices of the Royal Astronomical Society*, 435(2):1047–1060, 2013.

[13] J. Aasi et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, 2015.

[14] M. Lochner et al. Photometric supernova classification with machine learning. *The Astrophysical Journal*, 225(2):31, 2016.

[15] N. Chawla et al. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357, 2002.

[16] Y. Ren et al. Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational Intelligence Magazine*, 11(1):41–53, 2016.

[17] M. Juric. The LSST data management system. *ArXiv e-prints*, 2015.