Classification with class imbalance problem: A review

Article · January 2015

CITATIONS
183

READS
7,791

3 authors:

Aida Ali
Universiti Teknologi Malaysia
18 PUBLICATIONS **240** CITATIONS
SEE PROFILE

Siti Mariyam Shamsuddin
Universiti Teknologi Malaysia
403 PUBLICATIONS **4,127** CITATIONS
SEE PROFILE

Anca Ralescu
University of Cincinnati
226 PUBLICATIONS **1,929** CITATIONS
SEE PROFILE

Int. J. Advance Soft Compu. Appl, Vol. 5, No. 3, November 2013
ISSN 2074-2827

Classification with class imbalance problem: a review

Aida Ali[1,2], Siti Mariyam Shamsuddin[1,2], and Anca L. Ralescu[3]

[1]UTM Big Data Centre, Ibnu Sina Institute for Scientific and Industrial Research
Universiti Teknologi Malaysia, 81310 UTM Skudai,
Johor, Malaysia
e-mail: aida@utm.my, mariyam@utm.my
[2]Faculty of Computing, Universiti Teknologi Malaysia
81310 UTM Skudai,
Johor, Malaysia
e-mail: aida@utm.my, mariyam@utm.my
[3]School of Computing Science and Informatics, University of Cincinnati
email: anca.ralescu@uc.edu

Abstract

Most existing classification approaches assume the underlying training set is evenly distributed. In class imbalanced classification, the training set for one class (majority) far surpassed the training set of the other class (minority), in which, the minority class is often the more interesting class. In this paper, we review the issues that come with learning from imbalanced class data sets and various problems in class imbalance classification. A survey on existing approaches for handling classification with imbalanced datasets is also presented. Finally, we discuss current trends and advancements which potentially could shape the future direction in class imbalance learning and classification. We also found out that the advancement of machine learning techniques would mostly benefit the big data computing in addressing the class imbalance problem which is inevitably presented in many real world applications especially in medicine and social media.

Keywords: *Class imbalance problem, Imbalanced data sets, Imbalanced classification, Big data*

1 Introduction

In many domain applications, learning with class imbalance distribution happens regularly. Imbalanced class distribution in datasets occur when one class, often the one that is of more interest, that is, the positive or minority class, is insufficiently represented. In simpler term, this means the number of examples from positive class (minority) is much smaller than the number of examples of the negative class (majority). When rare examples are infrequently present, they are most likely predicted as rare occurrences, undiscovered or ignored, or assumed as noise or outliers which resulted to more misclassifications of positive class (minority) compared to the prevalent class.

Ironically, the smaller class (minority) is often of more interest and more importance, therefore it called for a strong urgency to be recognized. For example, in a medical diagnosis of a rare disease where there is critical need to identify such a rare medical condition among the normal populations. Any errors in diagnostic will bring stress and further complications to the patients. The physicians could not afford any incorrect diagnosis since this could severely affect the patients' wellbeing and even change the course of available treatments and medications. Thus, it is crucial that a classification model should be able to achieve higher identification rate on the rare occurrences (minority class) in datasets.

Studies on class imbalance classification has grown more emphasis only in recent years [1]. Reported works in classifications for class imbalance distribution come in many ranges of domain applications like fault diagnosis [2][3], anomaly detection [4], medical diagnosis [5][6], detection of oil spillage in satellite images [7], face recognition [8], text classification [9], protein sequence detection [10] and many others. The significant challenges of the class imbalance problem and its repeated incidence in practical applications of pattern recognition and data mining have engrossed many researchers that two workshops dedicated to research efforts in addressing the class imbalance problems were held at AAAI 2000 [11] and ICML 2003 [12] respectively.

This paper is organized as follows; Section 2 discusses the major challenges and limitations with class imbalance classification. Section 3 explains in details the main problems with imbalanced class datasets that hinder the performance of a classification algorithm, and how such problems affect the learning of class boundary for classification tasks. Section 4 takes on the overview of existing approaches in the research society in addressing the class imbalance learning and classification. A comparison table on various methods and techniques is presented for better explanation. Section 5 describes the output measurements widely adopted in evaluating the performance of a classification algorithm in classifying datasets with imbalance characteristics. Lastly, Section 6 describes the present trends and current development in class imbalance studies along with a discussion on how such trends and development might propagate the potential direction of research to further improve the learning and classification with class imbalance problem.

2 Learning with class imbalance problem

One of the main issues in learning with class imbalance distribution is that most standard algorithms are accuracy driven. In simpler term, this means that many classification algorithms operate by minimizing the overall error that is, trying to maximize the classification accuracy. However, in a class imbalance dataset, classification accuracy tells

very little about the minority class. Choosing accuracy as the performing criterion in class imbalance classification may gives inaccurate and misleading information about a classifier performance [13][14][15][16][17][18][19]. Consider a case scenario of a dataset with imbalance ratio of 1:100. The ratio suggests that for each example of minority class (positive), there are 100 majority (negative) class examples. A classification algorithm which tries to maximize accuracy to meet its objective-rule, will produce an accuracy of 99% just by correctly classifying all examples from the majority class but misclassify one example of the minority class.

Another concern with imbalanced class learning is that most standard classifiers assumed that the domain application datasets are equal in sight [1][20][21][22][23]. In reality, there are many datasets with class imbalance distribution presence, like has been mentioned earlier in the previous section. Many classification algorithms do not take into account the underlying distribution of the datasets thus generate inaccurate model representation in class-learning task. Such unwise attempt will then lead to deterioration in classification performance. Experiment from the studies in [22][24] found out that a majority of learning algorithms are designed around the notion that training sets are well balanced in distribution, which most of the time is not correct. The authors in [22] went to prove that in the case of feed-forward neural networks, class imbalance does hinder its performance especially when the class complexity increases.

In recent years, the size of data has rapidly grown to a larger volume due to advanced computer technologies and data mining. It is observed that data like genome study, protein sequence, DNA microarray, cloud computing, banking information, all exhibit higher volume than before with a growing number of features, sometimes up to thousands in number. Since domain applications like medical diagnosis and financial involved highly imbalance occurrences such as detecting a certain pattern in DNA microarray or recognizing fraudulent transactions in banking data, this motivates further advancement in imbalanced datasets management. Skewed distribution datasets with very high number of features called for effective feature selection strategies to evaluate the goodness of features since it is widely known that irrelevant, redundant and noise presence in feature space will hinder a classification performance [25][26][27][28][29][30][31][32].

Also, reported works in [17][21][33][34] pointed out that in most domain applications, the error cost are not similar, even though many current classifiers make the assumption that error cost from different classes are similar. For example, in real world scenario of tumor versus non-tumor, system OK versus system fault, fraud versus legitimate, all of these situations have different costs. If however, the cost matrix and class distribution is well-defined, the right threshold can be obtained easily. Unfortunately, the error cost is not easy to define even though with the help from field-experts, henceforth the error cost for these situations are uncommonly identified. Besides that, it is worth to note that even with well-balanced datasets, the cost is usually not known [17].

3 Challenges with class imbalance classification

Class imbalance happens when there are significantly lesser training examples in one class compared to other class. The nature of class imbalance distribution could occur in two situations; 1) when class imbalance is an intrinsic problem or it happens naturally. A naturally imbalanced class distribution happens in the case of credit card fraud or in rare disease detection. Another situation is 2) when the data is not naturally imbalanced, instead it

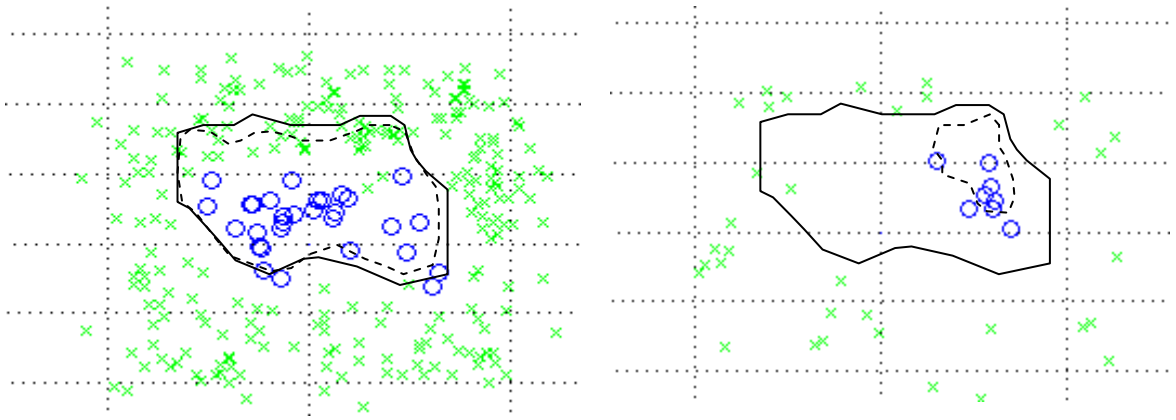
is too expensive to acquire such data for minority class learning due to cost, confidentiality and tremendous effort to find a well-represented data set, like a very rare occurrence of the failure of a space-shuttle. Class imbalance involves a number of difficulties in learning, including imbalanced class distribution, training sample size, class overlapping and small disjuncts. All these factors are explained in details in the following sections

3.1 Imbalanced class distribution

The imbalanced class distribution can be defined by the ratio of the number of instances of minority class to that of the majority class [1][17][21][33]. In certain domain problems, the imbalance ratio could be as extreme as 1:10000 [34]. The study of [35] investigated the correlation between ratio imbalance in training set with the classification results using decision tree classifier, and found out that a relatively balanced distribution between classes in datasets generally gives better results. However, as pointed out by [33], the degree of imbalance class distribution that will start to hinder the classification performance is still not explicitly known. An experiment from the study in [36] discovered that a balance distribution among classes is not a guarantee to improve a classifier performance since a 50:50 population ratio does not always the best distribution to learn from. This suggests that class imbalance distribution is not the only reason that deteriorates a classifier performance, other factors such as training sample size and class complexity also give influence [14].

3.2 Lack of information caused by small sample size

In addition to imbalance class distribution, another primary reason why class imbalance classification is challenging is because of lack of data due to small sample size in training set. Inadequate number of examples will caused difficulties to discover regularities, that is, pattern uniformity especially in the minority class.



(a) (b)

Fig. 1: The impact of small sample size in class imbalance problem; (a) the solid line determines the true decision boundary and (b) the dashed line defines the estimated decision boundary

Figure 1 illustrates how lack of data affects the classification performance in class imbalance learning, in which Figure 1a explains how the a classifier builds an estimated decision boundary (dashed line) from a relatively larger number of examples from positive class

(minority) where as Figure 1b is the estimated decision boundary constructed by the learning classification algorithm resulted from insufficient number of examples from positive class (minority). It is demonstrated that when an adequate number of examples is available, the estimated decision boundary captures the region agreeably to the true decision boundary as opposed to when insufficient examples from positive class do not improve the decision boundary, instead draws an unsatisfactory region that does not cover well to the true boundary.

A reported work found out that as training sample size increases, the error rate of the imbalanced class classification reduces[37]. This is also confirmed by [36], which reported the similar results using Fuzzy Classifier. This discovery is explainable since a classifier builds better representation for classes with more available training sample since more information could be learned from variation of instances provided by larger number of training size. It also reveals that a classifier should not be affected much by high imbalance ratio providing that there is a large enough number of training data.

3.3 Class overlapping or class complexity

One of the leading problems in class imbalance classification is class overlapping occurrences in the datasets. Class overlapping or sometimes referred to as class complexity or class separability corresponds to the degree of separability between classes within the data [21]. The difficulty to separate the minority class from the majority class is the major factor that complicates the learning of the smaller class. When overlapping patterns are present in each class for some feature space, or sometimes even in all feature space, it is very hard to determine discriminative rules to separate the classes. The overlapping feature space caused the features to lose their intrinsic property thus making them redundant or irrelevant to help recognize good decision boundaries between classes.

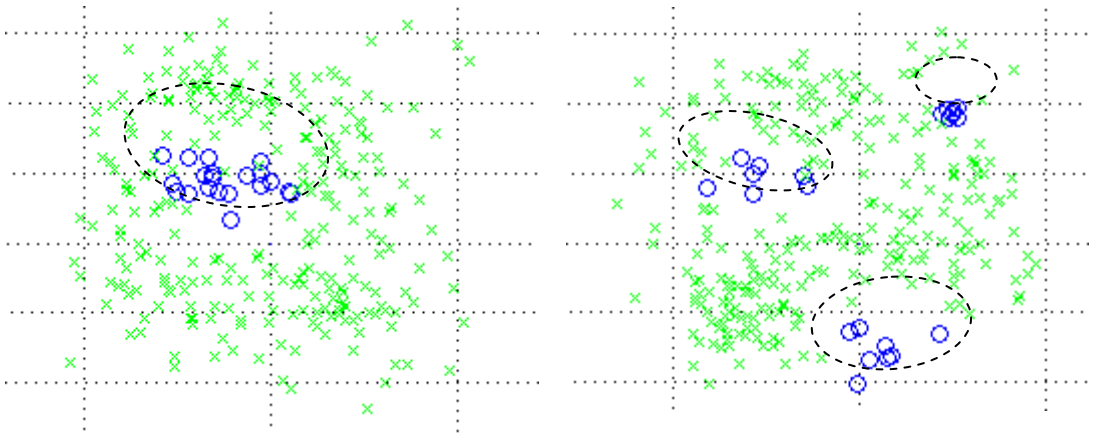
Previous work in [37] discovered that as the level of data complexity increases, the class imbalance factor begins to affect the generalization capability of a classifier. The work from [38] suggested that there is a relationship between overlap and imbalance in class imbalance classification however the level is not well-defined. Many investigations into class separability [39] [40][41][42][43], [44] [45],[16] and [46] bring evidences that class overlapping problem bring severe hindrance to a classifier performance compared to imbalanced class distribution. Standard classifiers which operate by trying to maximize accuracy in classification often fall into the trap of overlapping problem since those classifiers generally classified the overlapping region as belong to the majority class while assuming the minority class as noise [32].

3.4 Small disjuncts - within class imbalance

While in learning class imbalance classification, the imbalance ratio between minority class and majority class is obvious, sometimes an imbalance present within a single class might be overlooked. The within class imbalance or sometimes referred to as small disjunct appears when a class is comprised of several sub-clusters of different amount of examples [47][48][49]. Figure 2 below illustrates the concept of small disjuncts and overlapping class in class imbalance problem.

The studies of [35] and [50] explored the within class imbalance in minority class and claimed that the underrepresented minority class caused by small disjunct could be improved by applying a guided upsampling in respect to the minority class. [33] reported that small

disjuncts problem in class imbalance affects the classification performance because 1) it burdens a classifier in the task of concept learning of minority class and that 2) the occurrences of within class problem, most of the time is implicit. The within class problem is further signify because many current approaches to class imbalance mostly are more interested to solve the between-class problem and disregard the imbalance distribution within each class. Even though such situation provokes for more studies in solving within class problem, this study does not addressed the issue. Nevertheless, potential research direction reserved for future works is most definitely is of interest.



(a) (b)

Fig. 2: Example of Imbalance Between Classes (a) overlapping between classes (b) small disjunct – within class imbalance

4 Approaches in class imbalance classification

In general, there are two strategies [9][19][23][51][52][53][54][55] to handle class imbalance classification; 1) data-level approach and 2) algorithm-level approach. The methods at data-level approach adjust the class imbalance ratio with the objective to achieve a balance distribution between classes whereas at algorithm-level approach, the conventional classification algorithms are fine-tuned to improve the learning task especially relative to the smaller class. Table 1 provides a detailed summary on several notable previous works in class imbalance classification along with advantages and limitations of each strategy. Please note that we do not provide all reported literature due to lack of space.

4.1 Data level approach for handling class imbalance problem

Data-level approach or sometimes known as external techniques employs a pre-processing step to rebalance the class distribution. This is done by either employing under-sampling or over-sampling to reduce the imbalance ratio in training data. Under-sampling removes a smaller number of examples from majority class in order to minimize the discrepancy between the two classes whereas over-sampling duplicates examples from minority class [56].

4.1.1 Sampling

In 2002, a reported work of [56] proposed for an adaptive over-sampling technique named SMOTE (Synthetic Minority Over-sampling Technique) which has since gain popularity with

class imbalance classification. SMOTE adds new examples to minority class by computing a probability distribution to model the smaller class thus making the decision boundary larger in order to capture adjacent minority class examples. As proposed in [50], a new cluster-based over sampling that could simultaneously handle between-class imbalance and within-class imbalance, and a study in [57] came out with oversampling through the joining of boosting and data generation called DataBoost-IM algorithm. For undersampling scheme, the work of [58] proposed for a novel scheme that resample majority class using vector quantization to construct representative local models to train SVM. A cluster-based undersampling is put forward by the research in [59] that use clustering to choose representatives training examples to gain better accuracy prediction for minority class.

Nevertheless, the elimination of examples (down-sampling) from a class could lead to loss of potentially important information about the class, while in examples replication (over-sampling), the duplication only increase the number of examples but do not provide new information about the class, thus, it does not rectify the issue with lack of data [13][19][20][60]. However as disputed by [21] that when a limited number of examples from minority class is available, the estimated data distribution computed by the probability function might not be accurate. The authors in the same publication also highlighted that computational cost is higher when more examples are replicated from a class, besides leading to over-fitting [17][61].

Another study from [62] has demonstrated that sampling has the equivalent output to moving the decision threshold or modifying the cost matrix. Although there are many efforts in managing class imbalance problems through sampling, a study in [35] argued that there is no formal standard to define the most suitable class distribution, and experiments conducted by [20] discovered that often, a 50:50 imbalance ratio between minority class and majority class in training set does not always return the optimal classification performance. In addition, there is a knowledge gap in how does sampling is affected by within-class imbalance problem especially with random oversampling [33]. With within-class imbalance distribution (small disjuncts), a random oversampling method could replicate examples on certain regions but lesser on the others. Again, this brings to the question, which region should be concentrate on first? This issue cannot by systemically answered and further experiments are needed to provide satisfied feedback. Nevertheless, despite the disadvantages with sampling, sampling is still a well-known approach to handle imbalanced datasets compared to cost-sensitive learning algorithms [19]. Parts of its successful reasons are because many learning algorithms does not implement error-cost in learning process. Also, it is observed that nowadays many class imbalance datasets come in larger volume than before, thus motivating sampling in order to reduce the data size for a feasible learning task.

4.1.2 Feature selection

Besides sampling methods, another pre-processing step that is gaining popularity in class imbalance classification is feature selection. There are a few reported works on feature selection methods designed especially to address the problem of imbalanced class distribution. A suggestion in [27] proposed for a new class decomposition-based feature selection by applying feature selection on smaller pseudo-subclasses constructed from the partitioning of the majority class, and a new Hellinger distance-based approach for feature selection to address the high dimension class imbalance datasets. A reported study in [63] put forward an approach for feature subset selection that considers problem specifics and the property of learning algorithm for highly unbalanced class distribution, and discovered that Odds ratio is the most successful measurement fro Bayesian learning, nevertheless the

proposed method is only developed for Naive Bayes classifier. Authors from [64] described a new feature selection method to solve the problem with imbalanced text documents by exploiting the combination of most useful features from both classes, positive and negative. They then used multinomial Naive Bayes as classifiers and compared with traditional feature selection methods such as chi-square, correlation coefficient and odds-ratios.

Another attempt at applying feature selection to solve class imbalance problem is from [65] who proposed for a ROC based feature selection, instead of classification accuracy to assess classification performances. As discussed in [66] a novel feature subset selection based on correlation measure to handle small sample in class imbalance problem. Authors in [67] proposed for a new correlation measure named CFS to measure the worth of a feature subset based on level of correlation to the class however as disputed in [68] that the underlying algorithm, CFS applied heuristic search which has quadratic property will lead to increase time complexity. Another work in [69] proposed for a straight-forward approach for feature selection in which the method computes the relevancy of a feature based upon the variance of the mean value of the minority class. A classifier is assumed as relevant when the mean of the minority class is similar to or equivalent to two standard deviations away from the mean of the majority class.

The study of [63] demonstrated that irrelevant features do not significantly improved classification performance and suggested that more features slow down induction process. Also, feature selection removes irrelevant, redundant or noisy data [70] which reflected in the problem of class complexity or overlapping in class imbalance. Feature selection is adopted in class imbalance classification mostly to define feature relevance to the target concept [10]. In class imbalance classification, feature selection is employed to measure the "goodness" of a feature. Feature selection helps to suggest highly-influential features which often provide intrinsic information and discriminant property for class separability, besides improving classification performance, decreases computational cost and gives better understanding on model representations [28][31][66][70][71][72][73]. However, as pointed out by [27] and [17], although feature selection is rather established in many pattern recognition and data mining domain applications, feature selection in class imbalance classification is underexplored and the lack of systematic approach to feature selection for imbalanced datasets opens to many research possibilities. It is also argued by many such as [74], [66] and [75] that sampling might not be enough to solve the challenge in class imbalance data.

In general, there are two approaches to apply feature selection algorithms in classification i.e. by adopting either method 1) filter or 2) wrapper. Filter method referred to pre-processing algorithms that measure the goodness of the feature subset by looking at the intrinsic features from the data. They are practically inexpensive since they do not depend on induction algorithm. Wrapper method, in contrast, wraps the feature selection process around the induction algorithm. Although they are computationally expensive compared to the former, they generally are better at predicting accuracy than filter methods [31][69][76][77]. Wrapper methods explore the feature subsets space using a learning algorithm to report on estimated classification accuracy so that each feature could be included or eliminated from the feature subset. Filter methods, on the other hand, choose a feature set to have a learning algorithm use it to learn a target concept in the training set. An advantage of wrapper methods is that the estimated classification accuracy is usually the best heuristic evaluation for the feature subsets [76]. Furthermore, when learning with class imbalance datasets, the heuristic measurement of the feature subsets serves as an open alternative for better fitness evaluation thus making this approach a more versatile option than filter methods. Also, it is argued that

in real world problem, providing that all resources and instruments are formally established, feature subset selection is only done once, that is, during the pre-processing stage, thus, the computationally expensive cost when the induction algorithm is in operation, does not influence the classification task.

4.2 Algorithm level approach for handling class imbalance problem

Generally, the algorithm-level methods could be categorised as dedicated algorithms that directly learn the imbalance distribution from the classes in the datasets, recognition-based one class learning classifications, cost-sensitive learning and ensemble methods. The following subsections will discuss each category in details.

4.2.1 Improved algorithm

One of the leading approach in managing the classification of datasets with class imbalance problem is when researchers developed a classification algorithm which is modified to fit the requirement to learn directly from the imbalanced class distribution. These type of algorithms learn about the imbalance distribution of the classes before extracting important information in order to develop a model based upon the target objective. There are recent literature on improved SVM to handle imbalanced data such as in the work of z-SVM [78] and GSVM-RU [79]. z-SVM uses the parameter z that moves the position of hyperplane which maximize the g-mean value while GSVM-RU applies granular computing to represents information as aggregates to improve classification efficiency.

Another attempt is by improving k-NN with Exemplar Generalization selectively enlarge the positive instances in the training sample which is referred to as exemplar positive instances to expand the decision boundary of the positive class [80]. The selected exemplar positive instances are determined by computing a set of positive pivot points and then generalized using Gaussian ball. The distance of nearest neighbours for each pivot positive instance are then computed as kNN classification to build the k Examplar-based Nearest Neighbour (kENN) classifier. A Class Conditional Nearest Neighbour Distribution (CCNND) algorithm uses nearest neighbour distances to represent the variability of class boundary by computing the relative distance properties of each class [81]. Through the relative distances, CCNND learns to extract the classification boundary that preserves high sensitivity to the positive class (minority class) straight from the data.

In addition, there are also reported works on Fuzzy to address the classification of imbalanced datasets. Hierarchical Fuzzy rule uses a linguistic rule generation method to construct initial rule base from which the hierarchical rule base (HRB) is extracted from [82]. Then the best cooperative rules from HRB are selected using genetic algorithm. Another study proposed Fuzzy Classifier which uses relative frequency distribution to generate membership degrees to each class before constructing corresponding fuzzy sets [20]. This study presented a new alternative approach to conventional Fuzzy since it is purely data driven while the later relies on trial and error method in constructing the *if-then* rules.

4.2.2 One-class learning

One-class learning algorithms are also known as recognition based methods, work by modelling the classifier on the representation of the minority class. [83] applied neural networks and proceed to learn only from the examples of minority class rather than trying to recognize the different patterns from examples of majority class and minority class. However, as pointed out by [33], an effective boundary threshold is the key point with this approach

since a strict threshold will separate apart the positive examples (minority class) while a lenient one will cover some negative examples (majority class) in the decision boundary. Furthermore, most machine learning algorithm like decision trees, Naive Bayes and k-Nearest Neighbourhood do not function with examples only from one class thus making this approach less popular and confined only to certain learning algorithms [16][20].

4.2.3 Cost sensitive learning

The different natures of domain applications with class imbalance datasets and misclassification cost being regarded as equal by many traditional learning algorithm motivate the studies in cost-sensitive learning. Cost-sensitive learning approaches are designed with the idea that an expensive cost is imposed on a classifier when a misclassification happens for example a classifier assigns larger cost to false negatives compared to false positives thus emphasizing any correct classification or misclassification regarding the positive class. Several studies on cost sensitive learning for imbalance class distribution includes the work from [84] which proposed for optimized cost sensitive SVM, [85] discussed a PSO-based cost sensitive neural network, and [86] whose work applied SVM for asymmetrical misclassification costs as have been listed in Table 1.

However it is argued that in most applications the real cost is not known [13][87][88], even with balance distribution datasets [89]. [21] and [33] both pointed out that most of the time the cost matrix is usually unavailable since there are large number of factors to consider on. Also, the work in [32] found out that cost sensitive learning may cause over-fitting problem during training. A recent study from [90] revealed that cost-sensitive learning gives equal performance with oversampling methods and there is no difference between both strategies. Moreover, the authors of [33] pointed out that when ‘real’ cost value cannot be obtained, an artificial value cost value is generated or searched for and the exploration for the effective cost will lead to overhead in cost-learning task.

4.2.4 Ensemble method

Ensemble learning is another option for class imbalance problem. These methods trained several classifiers on training data and their evaluations are aggregated to produce the final classification decision. In general ensemble methods can be described as boosting or boosting. Bagging stands for Bootstrap Aggregation is the approach to reduce the prediction variance by generating more examples for training set from original data. A classifier is induced for each of these training set examples by a chosen machine learning algorithm, therefore, there will be *k* number of classifiers for each *k* variations of the training set. The result is produced by combining the output all the classifiers. Boosting methods carry out experiments on training sets using several models to induce classifiers to produce output. Higher weights are assigned to each classifier for wrongly classified examples. The outputs are then updated using weighted average approach. The final decision is obtained by combining all classifiers [91][92].

AdaBoost [93], Bagging [94] and RandomForest [95] are among the popular ensemble learning methods. Many reported works like SMOTEBoost [96], RUSBoost [97], DataBoost-IM [57] and cost-sensitive boosting [98] employed boosting to handle class imbalance problem. SMOTEBoost and DataBoost-IM integrated data generation and boosting procedures to improve classification performance. SMOTEBoost adjusts the class distribution by replicating examples of minority class using SMOTE technique [56]. DataBoost-IM identified hard examples from both minority and majority classes in order to construct synthetic data points in training set to achieve a balance in class distribution and total weights

for every class. A cost-sensitive boosting method is developed by [98], where AdaBoost algorithm is incorporated into misclassification cost. Such integration allows weight-update for misclassified samples from minority class where a study in [99] proposed for a new linear programming boosting to handle uneven distribution in datasets by associating weight to examples in order to adjust the distribution of sample data. Several learning algorithms are developed by on many strategies of sampling. SMOTEBagging is proposed by [100] by duplicating examples in subset construction. In the contrary, underbagging by [101] added new subset in training set by randomly undersampling the majority class.

The study from [13] has extensively investigated the ensemble learning techniques in relative to binary-class problem. An empirical comparison has been conducted and analysed various ensemble algorithms from many strategies 1) classic ensembles such as AdaBoost, AdaBoost.M1, Bagging; 2) cost-sensitive boosting like AdaC2; 3) boosting-based ensembles such as RUSBoost, SMOTEBoost, MSMOTEBoost; 4) bagging-based ensembles like UnderBagging, OverBagging, SMOTEBagging; 5) hybrid ensemble such as EasyEnsemble, BalanceCascade on 44 UCI Machine Learning datasets. AUC results revealed that RUSBoost, SMOTEBagging and UnderBagging return better classification compared to other ensemble algorithms particularly RUSBoost, being the least computational complex among other methods.

Even though ensemble methods are more versatile compared to cost-sensitive learning and other algorithm level approaches caused of its independency of base classifier, nevertheless, when building ensembles, innovating diverse classifiers while preserving their regularity with the training data is the crucial factor to ensure accuracy. While diversity in ensemble methods has an extensive theoretical principle in regression problems, when it comes to classification, the concept of diversity is still largely undefined [102], [103], [100] and [104]. The review from [105] also pointed out that understanding the classifier error diversity is not an effortless task and its grounded framework is formally incomplete, and the complexity issue grow higher with the use of more classifiers [13].

4.2.5 Hybrid approach

Besides the one-class learning, cost-sensitive methods and ensemble approaches, a new breed of classification algorithms have been devised for handling class imbalance datasets in recent years. Most of them employ more than one machine learning algorithms to improve the classification quality, often through the hybridization with other learning algorithms to achieve better results. The hybridization is designed with the idea to alleviate the problem in sampling, feature subset selection, cost matrix optimization and fine-tuning the classical learning algorithms.

In cost-sensitive learning, there are several published works like [106] who demonstrated the work of combining cost-sensitive learning and sampling using SMOTE algorithm [56] to improve the performance of SVM. There is also a reported work from [85] that put forward a PSO-based cost sensitive neural network for imbalanced class datasets. A recent work from [86] proposed for a SVM with Asymmetrical Misclassifications Cost whereas like [107] which use neural network to train on cost-sensitive classification.

Besides optimization of cost matrix, multiple studies dedicated to improving sampling and feature subset selection are also reported. A work of [108] used PSO to optimize feature selection for SVM and ANN in classifying the highly unbalanced data of power transformers fault diagnosis. Another work from [84] investigated the optimization of cost-sensitive SVM with PSO training using imbalanced evaluation criteria i.e. G-mean and AUC to find optimal

feature subset. Authors in [109] proposed for a hybrid method incorporating random over-sampling, decision tree, Particle Swam Optimization (PSO) and feature selection to address highly imbalanced datasets on a zoo dataset. Although the previous work use sampling with decision tree to improve effectiveness, this leads to complexity issue and an overhead in ensure the successfulness in parameter selection. To address these issues, [110] discussed a novel decision tree algorithm named Hellinger Distance Decision Trees (HDDT) that apply Hellineger distance for splitting criterion. A method called ACOSampling which applied an ant colony to optimize undersampling for the classification of highly imbalanced microarray data has been proposed by [111].

There are also reported studies that hybridize classifiers in order to improve classification qualities with class imbalance problem. [5] proposed to train NN with back-propagation method with mean square error as objective and compared with NN classifier trained with PSO in handling class imbalanced medical datasets. A study that address class overlap and imbalance using hybrid approaches by applying cost function to address class imbalance and Gabriel graphs editing to lessen the effect of class overlapping proble, both strategies are trained on back-propagation neural network [51]. There are also published work that applied k-NN for highly imbalanced datasets of medical decision [112] and an investigation on the effect of between class imbalance and within class imbalance on the performance of k-NN by [43]. Also, [113] proposed for a new F-measure based classifier instead of accuracy to address class overlapping and imbalance problem.

While most hybrid methods in class imbalance classifications focus more on neural networks, SVM and decision tree, only several literatures from fuzzy rule are dedicated to highly imbalanced distribution datasets. Fuzzy linguistic [114] investigated the behaviour of linguistic fuzzy rule based classification systems for imbalanced datasets while [115] proposed for a novel neuro-fuzzy network algorithm to produce multiple decision rules for real world banking data. Another work from [116] applied fuzzy classifier e-algorithm for fault detection in power distribution imbalanced data, and [6] used GA to help with fuzzy rule extraction to detect Down's syndrome in fetus.

Table 1: Previous Works on Class Imbalance Classification

Solutions	Strength	Weakness
Data-level Approach		
Sampling [117] MLSMOTE [63] Diversified sensitivity-based undersampling [111] ACOsampling with Ant Colony [56] SMOTE [60] Evolutionary undersampling	Straight forward approach and widely used in many domain applications	Risk of over fitting
Cost-sensitive Boosting [27] Cost sensitive linguistic fuzzy rule	Straight-forward technique especially if	Additional learning cost due to exploration for effective cost matrix

[98] Cost sensitive Boosting	the cost error is known	especially when real cost are not known
Feature Selection [70] Minority class feature selection [118] Density-based feature selection [65] FAST; roc-based feature selection [66] CFS; correlation feature selection	Helpful in alleviating class overlapping problem	Extra computational cost due to included pre-processing task
Algorithm-level Approach		
Improved Algorithm [119] Argument-based rule learning [64] Dissimilarity-based learning [20] Fuzzy Classifier [78] z-SVM [82]Hierarchical Fuzzy rule [81] Class conditional nearest neighbour distribution [80] k-NN with Examplar Generalization [120] Weighted nearest-neighbour classifier	Effective methods due to modified algorithms to learn exclusively from imbalance class distribution	Might need pre-processing tasks to balance out skewed class distribution
One-class Learning [83] One class learning [81] Class Conditional Nearest Neighbor Distribution (CCNND)	Simple methods	Not efficient when applied with classification algorithms that must learn from prevalent class
Cost-sensitive Learning [121] Near Bayesian SVM [84] Cost sensitive learning with SVM [85] Cost sensitive NN with PSO [86] SVM for Adaptively Asymmetrical Misclassification Cost	Simple, fast processing method	Ineffective if real cost are not available Extra cost introduced if cost exploration is needed when error cost is not known
Ensemble Method [122] SMOTE and feature selection ensemble [123] Ensemble GA [124] Ensemble for financial problem [125] Boosting with SVM ensemble [97] RUSBoost	Versatile approaches	Complexity grows with the use of more classifiers Diversity concept is difficult to achieve

[96] SMOTEBoost		
Hybrid Approach		
[66] FTM-SVM [113] F-measure based learning [114] Linguistic fuzzy rule [116] Fuzzy classifier e-algo [6] Fuzzy rule extraction with GA [115] Neuro fuzzy [51] Neural net medical data [126] Neural networks with SMOTE [112] Case-based classifier kNN for medical data [5] NN trained with BP and PSO for medical data [127] Dependency tree kernels [128] Exploiting cost sensitive in tree [71] [10] Undersampling and GA for SVM	Gaining popularity in class imbalance classification Symbiosis learning through combination with other learning algorithms	Needs careful design evaluation to compliment the differences between applied methods

The Fuzzy Classifier (FC) proposed by [45] is a classification algorithm that learns directly from the data and its underlying distribution. The FC is a data-driven algorithm whiles other fuzzy classifiers methods, most of the time, depend very much on trial and error approach to construct fuzzy sets. Even though, the later approaches benefit from the use of linguistic terms in the if-then rules, they have their own restrictions since the estimation of membership functions, most of the times, are very difficult to determine, unless the fuzzy rules are already known and established. The conventional Fuzzy has the advantage of using ambiguous linguistic term in the rules, however, has difficulties in membership function estimation unless the rules are already known and established. Besides that, the many ways of interpreting fuzzy rules and the defuzzification of output prove to be very challenging to solve especially when there is insufficient expert knowledge to define them [129][130].

Many earlier literatures reported on class imbalance classifications using decision tree, neural networks and SVM. For a decision tree, pruning seems to have severe effect on its performance since there is a high chance that an entire small class could be lost. Furthermore, when the smaller class is given high penalty such as system fault for a space shuttle, the very rare occurrence of such incident could result the class being mapped as a leaf in the tree structure, thus not much rules could be learned from [20][33]. Methods like C4.5, although a straight-forward and easy approach to adopt, tends to lead to overfitting problem. Moreover, over-sampling using this method will end with less pruning which results in generalization issue [125]. Besides that, most decision trees techniques like C4.5 and CART will produce complex representation because of replication problem.

Other classifiers like Naive Bayes classifier and some Neural Networks, provide a membership degree to how much an example belong to a class. This ranking approach is effective to classification reasoning, since Naive Bayes has strong assumption of child nodes independence, however [1] argued that this approach is not effective when the features examined have complex, non linear relationships between each other. Neural Network is a sophisticated machine learning in determining classification boundary, nevertheless the performance of ANN is largely dependent on the complexity of its architecture; the selection of ANN structure and parameters are usually subjective, lack of theoretical recommendations for training data set size, overtraining which later leads to memorization instead of data generalization, and the need of fine tuning a large number of parameters and their initialization e.g. starting weights, amount of cases, and the quantity of training cycles [131]). ANN also is notoriously known in treating smaller class as noise or outliers thus needing much of a pre-processing strategy to rebalance the class distribution [20].

k-NN is still one of the successful machine learning method in classification, nevertheless large-scaled data with complex, non-liner relationships available today poses a new problem. Another popular method to classification is SVM, however despite that, SVM is not without fault. Researchers like [132] and [133] pointed out that the major problem with SVM is the selection of the kernel function parameters, and its high algorithmic complexity and its need on expensive memory for the quadratic programming in large-scaled computation process.

It is also discovered that although not many researchers tend to explore on developing new algorithms which learn to adapt into imbalanced class distribution, this approach, most of the times, has the least expensive computing cost. This is due to no strong need of pre-processing requirement such as sampling methods in order to adjust the imbalance between class, and also modification on the algorithm itself to learn from the selected training sample in a specific manner as well as discarding irrelevant information in order to build better class representation.

5 Performance measures

Since the normal metric of overall accuracy in describing a classifier performance is no longer sufficient [20][37][134], the confusion matrix and its derivations will be used to summarise the performance results. For a binary-class problem, the confusion matrix comprises of four results from classification outputs that reports the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) like as has been illustrated in Figure 3 below. In the experiment, ‘positive’ refers to the minority class while ‘negative’ denotes the majority class. These four values provide to more detailed analysis and objective assessment which are then use to measure the performance of all classifiers in classifying the six data sets described previously.

		Predicted (Classified as)	
		Positive	**Negative**
Actual (Really is)	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Fig. 3: Confusion Matrix for A Binary Class Problem

For a two class classifier, a confusion matrix consists of information about actual and predicted classification return by a classifier. Often, a classifier performance is evaluated based on the information obtained from the confusion matrix. The entries in the confusion matrix are denoted as;

- True Positive (TP) refers to the number of positive examples which are correctly predicted as positives by a classifier
- True Negative (TN) denotes as the number of negative examples correctly classified as negatives by a classifier
- False Positive (FP), often referred to as *false alarm*; defines as the number of negative examples incorrectly classified as positives by a classifier
- False Negative (FN), sometimes known as *miss*; is determined as the number of positive examples incorrectly assigned as negatives by a classifier

However, by analysing the four entries in the confusion matrix is not enough in determining the performance of a classifier. Therefore, several derivatives based from the previously discussed confusion matrix are used in evaluating a classifier in this study. These performance metrics from the confusion matrix are:

- Sensitivity or true positive rate / recall is denoted as;

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

  Sensitivity refers to the ability of a classifier in correctly identifying positive class as such. It ranges from 0 to 1 with 1 being the perfect score.

- Specificity or true negative rate is determined as;

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

  Specificity denotes the ability a classifier in correctly identifying negative class as such. The perfect score is 1 and 0 is the worst measure.

- Accuracy is denoted as;

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

  Accuracy is a proportion of true results (both true positives and true negatives) in the population.

- G-mean (geometric mean)

$$G - mean = \sqrt{Sensitivity \times Specificity} \quad (4)$$

  G-mean or geometric mean introduced by [135] is for indicating the ability of a classifier in balancing the classification between positive class accuracy and negative class accuracy. By taking the G-mean of both sensitivity and specificity together, a low score for G-mean denotes a classifier that is highly biased towards one single class, and vice-versa.

- F-measure

$$F_\beta measure = \frac{(1+\beta^2) \times Recall \times Precision}{\beta^2 \times Recall + Precision} \quad (5)$$

where

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Precision is a measure of exactness, which is the proportion of observations from positive class correctly classified as positive. That is, the number of correct classification for positive class. It tells how well a classifier removes negative class from being misclassified as positive class.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Recall is a measure of completeness. It refers to the proportion of observations from positive class that should be returned, in other words it describes how well a classifier learns the positive class.

and β is a coefficient that balances the relative importance between precision and recall. β is set to 1 following the common practice in which the F_1 is often used for classification [136]. The F-measure with β equivalent to 1, means the recall and precision are evenly weighted. In simpler term, since Precision tells what percent of positive predictions were correct and Recall defines what percent of positive cases did a classifier catch, the F-measure shows the trade-off between the precision and recall regarding the positive class. It ranges between 0 and 1 with the score 1 is the best value. Since F-measure tells the trade-off between precision and recall, it indicates whether a classifier obtains high recall by sacrificing precision or vice versa by giving the classifier a low score.

In addition, for better visualization purpose, many researchers adopt another performance measure i.e. the Receiving Operating Characteristic (ROC) curve. The idea is to estimate the performance of a binary classifier by varying the discrimination threshold. The graphical plot of a ROC curve is make up by plotting the false positive rate (FP) on the x-axis and true positive rate (TP) on the y-axis. Every threshold value then construct a pair of measurements of FP versus TP. The perfect score is when a model achieved 1 true positive rate and 0 false positive rate. Hence, a good classification model should yield points near the upper left coordinate as shown in Figure 4. A ROC curve tells the trade-off between true positives and false positives. Each point in the ROC space represents a prediction result from a confusion matrix. A ROC curve can also be depicted as sensitivity versus (1-specificity) plot since true positive rate is equivalent to sensitivity and false positive rate can be written as 1-specificity. Additionally, the area under the ROC curve (AUC) shows the performance of the classifier model. The AUC score reduces the ROC curve to a single measure performance metric. It is also similar to the probability that a classifier model will rank a randomly selected positive sample higher than a randomly selected negative sample.

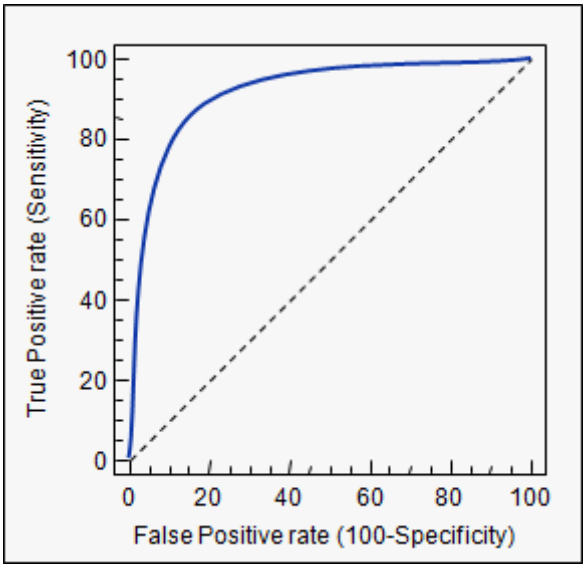


Fig. 4: An example of a receiver operating characteristic (ROC) plot

6 Current trends and future direction in class imbalance learning and classification

In the research society, it is observed that there are more reported works in binary class imbalance problem but only a handful of studies are conducted to find solution with multiple class imbalance problem. This most likely happen because of two reasons; 1) class imbalanced dataset with binary class is more prevalent in most domain applications like in anomaly detection and 2) the higher degree of complication when multiple class imbalance problem is concerned [33][137].

Besides that, more published works evidently has began to use state-of-the-art machine learning algorithm like SVM, GA and PSO whereas most earlier studies revolve around classifier induced by decision tree methods. It is also noticed that there is an increase of researches on hybridization techniques to achieve better classifier algorithms, as well as improving sampling and feature selection tasks to further understand class representation. Examples of hybrid learning come from many such as improving under sampling with ant colony optimization [111], evolutionary computing in feature selection [138],[139] and building ensembles [122],[124] to address the issues with the class imbalance classification.

Many reported works such as from [44], [39], [32], [16], [41], [140] and [38] have discovered that class overlapping severely hinders a classifier performance more than class imbalance. To solve the issues of class overlapping, many studies have adopted the alternative solution by incorporating sampling strategy in the pre-processing task. However, it is pointed out by many such as [74], [66] and [75] that sampling might not be enough to solve the challenge of overlapping problem in class imbalance data. The studies in [44][67][71][73][77][118][141][142][143][144] highlighted that class overlapping problem is caused by irrelevant or redundant features, and feature selection is one of the strategies used to address this issue.

The complexity of how the data is distributed in the multidimensional feature space coupled with very small number of observations forming the minority class makes it very challenging

to distinguish the minority class from the majority class. The task to learn the minority class from a very small training size using very high number of features has become a burden to the classifier. Worse, not many classifiers are designed to handle large amount of irrelevant features. There exist many cases where employing all available features is no guarantee that better classification output can be achieved due to redundancy problem, that is, features are partially or completely irrelevant to objective learning [68], [26], [25] and [30] . In the case of class imbalance data, irrelevant features may sometimes happen because of class overlapping problem. This situation occurs when data points fall into overlap region of the majority class and minority class. When this occurrence happens in many feature space, the recognition of the two classes becomes very difficult. Hence, there is a strong urgency in searching out for strong, dominant feature subset from all available features especially in the classification of multivariate data set.

Since not all features are able to give discriminant information that could well-separate a single class from an adverse class, the processing involved in taking in all the features is becoming an overhead to a classifier especially in its execution time and memory usage. The need for minimising computing cost when handling large volume of data with multivariate features bring forth the necessity of good feature selection in order to assist classification tasks. The advantages of feature selection come in manifolds, as it should help to improve prediction performance by preventing overfitting problem, provide quicker process and more cost-effective model in term of storage requirements and training times, assist in visualization of underlying data distribution in feature space for better understanding and resisting the curse of dimensionality for classification improvement [31], [29]. These listed advantages justify the need for a feature selection strategy to address the issues face in the classification of class imbalance data set with high dimensional feature space.

It is also revealed that there are strong inclinations among researchers whose studies are moving towards learning imbalanced classification of higher dimensional datasets especially from the medical domain applications. A few reported works show that the direction is leaning towards solving real world problem such as genomic datasets [143], learning protein sequence[10], breast cancer gene expression microarray study [145] and neuropsychological data [54] among others. In addition, the boom of social media in the global era in various applications has enabled for sentiment analysis and opinion mining from the crowd [146][147]. Many previous works such as sentiment analysis on Twitter media[148], live reactions on sports from Twitter [149], sentiment classification from social media [150] and sentiment analysis on online shopping review [151] reported class imbalanced problem does exist in such domain. In social media, the multiple class label issue in text categorization and image annotation for example, complicates the problem. This problem is further accentuates by the rapid growth of computer technologies and data mining which allow large volume of data to be stored and analyze.

Large volume of data set remains one of the major challenges in the classification domain. Data sets expand rapidly in number and also in attribute / feature wise. Since the computing world is moving toward big data realm, class imbalance problem in big data is inevitable. When many features are needed in describing a data point, the features will spread into complicated surfaces in multidimensional space. Multiple features sometimes have non-linear and complex relationship between each other which complicates knowledge extraction and data mining process. Several works in imbalanced big data are reported in [153][154][155] where most of them focused on developing machine learning algorithms under MapReduce framework. It is foreseen that increasing demand of big data applications from real world will

most probably call for better advancement in machine learning algorithm for imbalanced big data management.

In the future, the rapid development of big data computing most probably will shape the way classification tasks are performed and with anomaly patterns exist in most real world problem, class imbalance problem is inevitable. Encouraged by the previous driving factors, it is also interesting to note that with machine learning techniques being more accepted across domains, the marriage between machine learning and big data will see more frameworks [156] and systematic mapping [157] as the potential research direction. There is a strong need to address such problem and based upon the current trends and development, we will most likely see new issues and innovative contributions open up to a new frontier in the area of class imbalance learning and classification.

7 Conclusion

This paper presents an overview on class imbalance classification and the inevitable challenges that come with it. It describes the main issues that hinder the classifier performance in managing highly imbalanced datasets and the many factors that contribute to the class imbalance problems. Research gap in previous works are discussed along with justifications to this research attempt. Lastly, the observed current trends together with recent advancements in class imbalance classification are presented. This paper also suggests several potential developments in the domain such as the machine learning for big data computing and the boom of sentiment analysis from social media that could fuel the future research direction.

ACKNOWLEDGEMENTS.

The authors would like to express appreciation to the Ministry of Higher Education of Malaysia (MOHE), UTM Big Data Centre of Universiti Teknologi Malaysia and Y.M.Said for their support in this study. The authors greatly acknowledge the Research Management Centre, UTM and Ministry of Higher Education for the financial support through Fundamental Research Grant Scheme (FRGS) Vot. No. J130000.7828.4F606

References

[1] Kotsiantis, S., D. Kanellopoulos, and P. Pintelas, Handling imbalanced datasets: a review. GESTS International Transactions on Computer Science and Engineering, 2006. Vol 30(No 1): p. 25-36.

[2] Yang, Z., et al., Association rule mining-based dissolved gas analysis for fault diagnosis of power transformers. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2009. 39(6): p. 597-610.

[3] Zhu, Z.-B. and Z.-H. Song, Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis. Chemical Engineering Research and Design, 2010. 88(8): p. 936-951.

[4] Tavallaee, M., N. Stakhanova, and A.A. Ghorbani, Toward credible evaluation of anomaly-based intrusion-detection methods. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2010. 40(5): p. 516-524.

[5] Mazurowski, M.A., et al., Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. Neural networks : the official journal of the International Neural Network Society, 2008. 21(2-3): p. 427-436.

[6] Soler, V., et al. Imbalanced Datasets Classification by Fuzzy Rule Extraction and Genetic Algorithms. in Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on. 2006.

[7] Kubat, M. and S. Matwin. Addressing the curse of imbalanced training sets: one-sided selection. in ICML. 1997.

[8] Yi-Hung, L. and C. Yen-Ting. Total margin based adaptive fuzzy support vector machines for multiview face recognition. in Systems, Man and Cybernetics, 2005 IEEE International Conference on. 2005.

[9] Li, Y., G. Sun, and Y. Zhu. Data imbalance problem in text classification. in Information Processing (ISIP), 2010 Third International Symposium on. 2010. IEEE.

[10] Al-Shahib, A., R. Breitling, and D. Gilbert, Feature selection and the class imbalance problem in predicting protein function from sequence. Applied Bioinformatics, 2005. 4(3): p. 195-203.

[11] Japkowicz, N. in Proc AAAI 2000 Workshop on Learning from Imbalanced Data Sets. 2000. AAAI Tech Report WS-00-05.

[12] Chawla, N.V., N. Japkowicz, and A. Kotcz. in Proc ICML 2003 Workshop on Learning from Imbalanced Data Sets. 2003.

[13] Galar, M., et al., A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2012. 42(4): p. 463-484.

[14] Japkowicz, N. Class Imbalance: Are We Focusing on the Right Issue? in Notes from the ICML Workshop on Learning from Imbalanced Data Sets II. 2003.

[15] Chawla, N.V., *Data mining for imbalanced datasets: An overview*, in *Data mining and knowledge discovery handbook*2005, Springer. p. 853-867.

[16] Batista, G.E., R.C. Prati, and M.C. Monard, *Balancing strategies and class overlapping*, in *Advances in Intelligent Data Analysis VI*2005, Springer. p. 24-35.

[17] Visa, S., Ralescu, A. Issues in mining imbalanced data sets - a review paper. in Proceedings of the Midwest Artificial Intelligence and Cognitive Science Conference. 2005. Dayton.

[18] He, H. and E.A. Garcia, Learning from imbalanced data. Knowledge and Data Engineering, IEEE Transactions on, 2009. 21(9): p. 1263-1284.

[19] Vaishali, G., An Overview Of Classification Algorithms For Imbalanced Datasets. Int Journal of Emerging technology and Advanced Engineering, 2012. 2(4).

[20] Visa, S., Fuzzy Classifiers for Imbalanced Data Sets, in Department of Electrical and Computer Engineering and Computer Science2006, Univeristy of Cincinnati: Cincinnati.

[21] Nguyen, G.H., A. Bouzerdoum, and S.L. Phung, Learning Pattern Classification Tasks with Imbalanced Data Sets. 2009.

[22] Japkowicz, N. The class imbalance problem: Significance and strategies. in Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI'2000). 2000. Citeseer.

[23] Guo, X., et al. On the class imbalance problem. in Natural Computation, 2008. ICNC'08. Fourth International Conference on. 2008. IEEE.

[24] Japkowicz, N. Learning from imbalanced data sets: a comparison of various strategies. in AAAI workshop on learning from imbalanced data sets. 2000.

[25] Almuallim, H. and T.G. Dietterich. Learning with Many Irrelevant Features. in AAAI. 1991.

[26] Dash, M. and H. Liu, Feature selection for classification. Intelligent data analysis, 1997. 1(3): p. 131-156.

[27] Yin, L., et al., Feature selection for high-dimensional imbalanced data. Neurocomputing, 2013. 105(0): p. 3-11.

[28] Van Hulse, J., et al. Feature selection with high-dimensional imbalanced data. in Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on. 2009. IEEE.

[29] Saeys, Y., I. Inza, and P. Larrañaga, A review of feature selection techniques in bioinformatics. Bioinformatics, 2007. 23(19): p. 2507-2517.

[30] John, G.H., R. Kohavi, and K. Pfleger. Irrelevant Features and the Subset Selection Problem. in ICML. 1994.

[31] Guyon, I. and A. Elisseeff, An introduction to variable and feature selection. The Journal of Machine Learning Research, 2003. 3: p. 1157-1182.

[32] Weiss, G.M., Mining with rarity: a unifying framework. ACM SIGKDD Explorations Newsletter, 2004. 6(1): p. 7-19.

[33] Sun, Y., A.K. Wong, and M.S. Kamel, Classification of imbalanced data: A review. International Journal of Pattern Recognition and Artificial Intelligence, 2009. 23(04): p. 687-719.

[34] Chawla, N.V., N. Japkowicz, and A. Kotcz, Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor. Newsl., 2004. 6(1): p. 1-6.

[35] Weiss, G.M. and F. Provost, Learning when training data are costly: The effect of class distribution on tree induction. Journal of Artificial Intelligence Research, 2003. 19: p. 315-354.

[36] Visa, S., Ralescu, A. , The Effect of Imbalanced Data Class Distribution on Fuzzy Classifiers - Experimental Study. Fuzzy Systems 2005 FUZZ05 The 14th IEEE International Conference on, 749-754, 2005.

[37] Japkowicz, N. and S. Stephen, The class imbalance problem: A systematic study. Intelligent data analysis, 2002. 6(5): p. 429-449.

[38] Denil, M. and T. Trappenberg, *Overlap versus imbalance*, in *Advances in Artificial Intelligence*2010, Springer. p. 220-231.

[39] Prati, R.C., G.E. Batista, and M.C. Monard, *Class imbalances versus class overlapping: an analysis of a learning system behavior*, in *MICAI 2004: Advances in Artificial Intelligence*2004, Springer. p. 312-321.

[40] García, V., et al., *Combined Effects of Class Imbalance and Class Overlap on Instance-Based Classification*, in *Intelligent Data Engineering and Automated Learning – IDEAL 2006*, E. Corchado, et al., Editors. 2006, Springer Berlin Heidelberg. p. 371-378.

[41] García, V., et al., *When overlapping unexpectedly alters the class imbalance effects*, in *Pattern Recognition and Image Analysis*2007, Springer. p. 499-506.

[42] García, V., J. Sánchez, and R. Mollineda, *An empirical study of the behavior of classifiers on imbalanced and overlapped data sets*, in *Progress in Pattern Recognition, Image Analysis and Applications*2007, Springer. p. 397-406.

[43] García, V., R.A. Mollineda, and J.S. Sánchez, On the k-NN performance in a challenging scenario of imbalance and overlapping. Pattern Analysis and Applications, 2008. 11(3-4): p. 269-280.

[44] Xiong, H., J. Wu, and L. Liu. Classification with Class Overlapping: A Systematic Study. in The 2010 International Conference on E-Business Intelligence. 2010.

[45] Visa, S., Ralescu, A. Learning Imbalanced And Overlapping Classes Using Fuzzy Sets. in Proceedings of the International Conference of Machine Learning, Workshop on Learning from Imbalanced data Sets (II): Learning with Imbalanced Data Sets II. 2003. Washington

[46] Tomašev, N. and D. Mladenić, Class imbalance and the curse of minority hubs. Knowledge-Based Systems, 2013. 53(0): p. 157-172.

[47] Japkowicz, N., *Concept-learning in the presence of between-class and within-class imbalances*, in *Advances in Artificial Intelligence*2001, Springer. p. 67-77.

[48] Prati, R.C., G.E. Batista, and M.C. Monard, *Learning with class skews and small disjuncts*, in *Advances in Artificial Intelligence–SBIA 2004*2004, Springer. p. 296-306.

[49] Weiss, G.M. The impact of small disjuncts on classifier learning. in Data Mining. 2010. Springer.

[50] Jo, T. and N. Japkowicz, Class imbalances versus small disjuncts. SIGKDD Explor. Newsl., 2004. 6(1): p. 40-49.

[51] Alejo, R., et al., A hybrid method to face class overlap and class imbalance on neural networks and multi-class scenarios. Pattern Recognition Letters, 2013. 34(4): p. 380-388.

[52] Fatourechi, M., et al. Comparison of evaluation metrics in classification applications with imbalanced datasets. in Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on. 2008. IEEE.

[53] Stefanowski, J. and S. Wilk, *Selective pre-processing of imbalanced data for improving classification performance*, in *Data Warehousing and Knowledge Discovery*2008, Springer. p. 283-292.

[54] Nunes, C., et al., *Class Imbalance in the Prediction of Dementia from Neuropsychological Data*, in *Progress in Artificial Intelligence*, L. Correia, L. Reis, and J. Cascalho, Editors. 2013, Springer Berlin Heidelberg. p. 138-151.

[55] Phung, S.L., A. Bouzerdoum, and G.H. Nguyen, Learning pattern classification tasks with imbalanced data sets. 2009.

[56] Chawla, N.V., et al., SMOTE: synthetic minority over-sampling technique. arXiv preprint arXiv:1106.1813, 2002.

[57] Guo, H. and H.L. Viktor, Learning from imbalanced data sets with boosting and data generation: the DataBoost-IM approach. ACM SIGKDD Explorations Newsletter, 2004. 6(1): p. 30-39.

[58] Yu, T., et al. A hierarchical VQSVM for imbalanced data sets. in Neural Networks, 2007. IJCNN 2007. International Joint Conference on. 2007. IEEE.

[59] Yen, S.-J. and Y.-S. Lee, Cluster-based under-sampling approaches for imbalanced data distributions. Expert Systems with Applications, 2009. 36(3): p. 5718-5727.

[60] García, S. and F. Herrera, Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. Evolutionary Computation, 2009. 17(3): p. 275-306.

[61] Provost, F. Machine learning from imbalanced data sets 101. in Proceedings of the AAAI'2000 workshop on imbalanced data sets. 2000.

[62] Maloof, M.A. Learning when data sets are imbalanced and when costs are unequal and unknown. in ICML-2003 workshop on learning from imbalanced data sets II. 2003.

[63] Mladenic, D. and M. Grobelnik. Feature selection for unbalanced class distribution and naive bayes. in ICML. 1999.

[64] Zheng, Z., X. Wu, and R. Srihari, Feature selection for text categorization on imbalanced data. ACM SIGKDD Explorations Newsletter, 2004. 6(1): p. 80-89.

[65] Chen, X.-w. and M. Wasikowski. Fast: a roc-based feature selection metric for small samples and imbalanced data classification problems. in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. 2008. ACM.

[66] Wasikowski, M. and X.-w. Chen, Combating the small sample class imbalance problem using feature selection. Knowledge and Data Engineering, IEEE Transactions on, 2010. 22(10): p. 1388-1400.

[67] Hall, M.A. and L.A. Smith. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. in FLAIRS Conference. 1999.

[68] Yu, L. and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. in ICML. 2003.

[69] Cuaya, G., A. Muñoz-Meléndez, and E.F. Morales, *A minority class feature selection method*, in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*2011, Springer. p. 417-424.

[70] Cuaya, G., A. Muñoz-Meléndez, and E.F. Morales. A Minority Class Feature Selection Method. in Proceedings of the 16th Iberoamerican Congress Conference on Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 2011. Springer-Verlag.

[71] Alibeigi, M., S. Hashemi, and A. Hamzeh, Unsupervised Feature Selection Based on the Distribution of Features Attributed to Imbalanced Data Sets. International Journal of Artificial Intelligence and Expert Systems, 2011. 2(1): p. 133-144.

[72] Jamali, I., M. Bazmara, and S. Jafari, Feature Selection in Imbalance data sets. International Journal of Computer Science Issues, 2012. 9(3): p. 42-45.

[73] Kamal, A.H., et al., Feature Selection for Datasets with Imbalanced Class Distributions. International Journal of Software Engineering and Knowledge Engineering, 2010. 20(02): p. 113-137.

[74] Longadge, R. and S. Dongre, Class Imbalance Problem in Data Mining Review. arXiv preprint arXiv:1305.1707, 2013.

[75] Khoshgoftaar, T.M., K. Gao, and N. Seliya. Attribute selection and imbalanced data: Problems in software defect prediction. in Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on. 2010. IEEE.

[76] Das, S. Filters, wrappers and a boosting-based hybrid for feature selection. in ICML. 2001. Citeseer.

[77] Kamal, A.H., et al. Feature selection with biased sample distributions. in Information Reuse & Integration, 2009. IRI'09. IEEE International Conference on. 2009. IEEE.

[78] Imam, T., K. Ting, and J. Kamruzzaman, *z-SVM: An SVM for Improved Classification of Imbalanced Data*, in *AI 2006: Advances in Artificial Intelligence*, A. Sattar and B.-h. Kang, Editors. 2006, Springer Berlin Heidelberg. p. 264-273.

[79] Tang, Y., et al., SVMs modeling for highly imbalanced classification. Trans. Sys. Man Cyber. Part B, 2009. 39(1): p. 281-288.

[80] Li, Y. and X. Zhang, *Improving k nearest neighbor with exemplar generalization for imbalanced classification*, in *Advances in knowledge discovery and data mining*2011, Springer. p. 321-332.

[81] Kriminger, E., J.C. Principe, and C. Lakshminarayan. Nearest Neighbor Distributions for imbalanced classification. in Neural Networks (IJCNN), The 2012 International Joint Conference on. 2012.

[82] Fernández, A., M.J. del Jesus, and F. Herrera, Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets. International Journal of Approximate Reasoning, 2009. 50(3): p. 561-577.

[83] Japkowicz, N., C. Myers, and M. Gluck. A novelty detection approach to classification. in IJCAI. 1995.

[84] Cao, P., D. Zhao, and O. Zaiane, *An Optimized Cost-Sensitive SVM for Imbalanced Data Learning*, in *Advances in Knowledge Discovery and Data Mining*2013, Springer. p. 280-292.

[85] Cao, P., D. Zhao, and O. Zaiane, A PSO-based Cost-Sensitive Neural Network for Imbalanced Data Classification. 2011.

[86] Wang, X., et al. Using SVM with Adaptively Asymmetric Misclassification Costs for Mine-Like Objects Detection. in Machine Learning and Applications (ICMLA), 2012 11th International Conference on. 2012. IEEE.

[87] Weiss, G.M., K. McCarthy, and B. Zabar. Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs? in DMIN. 2007.

[88] Lin, H.-T. Cost-sensitive classification: Status and beyond. in Workshop on Machine Learning Research in Taiwan: Challenges and Directions. 2010.

[89] Visa, S., Ralescu, A., Data-Driven Fuzzy Sets For Classification. International Journal of Advanced Intelligence Paradigms 2008 - Vol. 1, No.1 pp. 3 - 30, 2008.

[90] López, V., et al., Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. Expert Systems with Applications, 2012. 39(7): p. 6585-6608.

[91] Kotsiantis, S. and P. Pintelas, Combining bagging and boosting. International Journal of Computational Intelligence, 2004. 1(4): p. 324-333.

[92] Machová, K., et al., A comparison of the bagging and the boosting methods using the decision trees classifiers. Computer Science and Information Systems, 2006. 3(2): p. 57-72.

[93] Friedman, J., T. Hastie, and R. Tibshirani, Additive Logistic Regression: A Statistical View of Boosting. Ann. Statist, 2000(28(2)): p. 337-374.

[94] Breiman, L., Bagging Predictors. Machine Learning, 1996. 24: p. 123-140.

[95] Breiman, L., Random Forests. Machine Learning, 2001. 45: p. 5-32.

[96] Chawla, N.V., et al., *SMOTEBoost: Improving prediction of the minority class in boosting*, in *Knowledge Discovery in Databases: PKDD 2003*2003, Springer. p. 107-119.

[97] Seiffert, C., et al., RUSBoost: A hybrid approach to alleviating class imbalance. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2010. 40(1): p. 185-197.

[98] Sun, Y., et al., Cost-sensitive boosting for classification of imbalanced data. Pattern recognition, 2007. 40(12): p. 3358-3378.

[99] Leskovec, J. and J. Shawe-Taylor. Linear programming boosting for uneven datasets. in ICML. 2003.

[100] Wang, S. and X. Yao. Diversity analysis on imbalanced data sets by using ensemble models. in Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on. 2009. IEEE.

[101] Liu, Y., et al., A study in machine learning from imbalanced data for sentence boundary detection in speech. Computer Speech & Language, 2006. 20(4): p. 468-494.

[102] Brown, G., Diversity in neural network ensembles, in Phd Thesis2004, University of Birmingham.

[103] Ueda, N. and R. Nakano. Generalization error of ensemble estimators. in Neural Networks, 1996., IEEE International Conference on. 1996. IEEE.

[104] Krogh, A. and J. Vedelsby, Neural network ensembles, cross validation, and active learning. Advances in neural information processing systems, 1995: p. 231-238.

[105] Brown, G., et al., Diversity creation methods: a survey and categorisation. Information Fusion, 2005. 6(1): p. 5-20.

[106] Akbani, R., S. Kwek, and N. Japkowicz, *Applying support vector machines to imbalanced datasets*, in *Machine Learning: ECML 2004*2004, Springer. p. 39-50.

[107] Zhou, Z.-H. and X.-Y. Liu, Training cost-sensitive neural networks with methods addressing the class imbalance problem. Knowledge and Data Engineering, IEEE Transactions on, 2006. 18(1): p. 63-77.

[108] Lee, T.-F., M.-Y. Cho, and F.-M. Fang, Features selection of SVM and ANN using particle swarm optimization for power transformers incipient fault symptom diagnosis. International Journal of Computational Intelligence Research, 2007. 3(1): p. 60-65.

[109] Lee, C.Y., et al. A hybrid algorithm applied to classify unbalanced data. in Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on. 2010.

[110] Cieslak, D.A., et al., Hellinger distance decision trees are robust and skew-insensitive. Data Mining and Knowledge Discovery, 2012. 24(1): p. 136-158.

[111] Yu, H., J. Ni, and J. Zhao, ACOSampling: An ant colony optimization-based undersampling method for classifying imbalanced DNA microarray data. Neurocomputing, 2013. 101(0): p. 309-318.

[112] Malof, J.M., M.A. Mazurowski, and G.D. Tourassi, The effect of class imbalance on case selection for case-based classifiers: An empirical study in the context of medical decision support. Neural Networks, 2012. 25(0): p. 141-145.

[113] Martino, M.D., et al., Novel Classifier Scheme for Unbalance Problems. Pattern Recognition Letters, 2013.

[114] Fernández, A., et al., A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. Fuzzy Sets and Systems, 2008. 159(18): p. 2378-2398.

[115] Hung, C.-M. and Y.-M. Huang, Conflict-sensitivity contexture learning algorithm for mining interesting patterns using neuro-fuzzy network with decision rules. Expert Systems with Applications, 2008. 34(1): p. 159-172.

[116] Le, X., C. Mo-Yuen, and L.S. Taylor, Power Distribution Fault Cause Identification With Imbalanced Data Using the Data Mining-Based Fuzzy Classification E-Algorithm. Power Systems, IEEE Transactions on, 2007. 22(1): p. 164-171.

[117] Yang, T.-N. and S.-D. Wang, Robust algorithms for principal component analysis. Pattern Recognition Letters, 1999. 20(9): p. 927-933.

[118] Alibeigi, M., S. Hashemi, and A. Hamzeh, DBFS: An effective Density Based Feature Selection scheme for small sample size and high dimensional imbalanced data sets. Data & Knowledge Engineering, 2012. 81–82(0): p. 67-103.

[119] Luukka, P., Nonlinear fuzzy robust PCA algorithms and similarity classifier in bankruptcy analysis. Expert Systems with Applications, 2010. 37(12): p. 8296-8302.

[120] Candès, E.J., et al., Robust principal component analysis? Journal of the ACM (JACM), 2011. 58(3): p. 11.

[121] Jolliffe, I., *Principal component analysis*2005: Wiley Online Library.

[122] Yang, P., et al., *Ensemble-based wrapper methods for feature selection and class imbalance learning*, in *Advances in Knowledge Discovery and Data Mining*2013, Springer. p. 544-555.

[123] Yu, E. and S. Cho, Ensemble based on GA wrapper feature selection. Computers & Industrial Engineering, 2006. 51(1): p. 111-116.

[124] Liao, J.-J., et al., An ensemble-based model for two-class imbalanced financial problem. Economic Modelling, 2014. 37(0): p. 175-183.

[125] Liu, Y., A. An, and X. Huang, *Boosting prediction accuracy on imbalanced datasets with SVM ensembles*, in *Advances in Knowledge Discovery and Data Mining*2006, Springer. p. 107-118.

[126] Jeatrakul, P., K.W. Wong, and C.C. Fung, Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm, in Proceedings of the 17th international conference on Neural information processing: models and applications - Volume Part II2010, Springer-Verlag: Sydney, Australia. p. 152-159.

[127] Culotta, A. and J. Sorensen. Dependency tree kernels for relation extraction. in Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. 2004. Association for Computational Linguistics.

[128] Drummond, C. and R.C. Holte. Exploiting the cost (in) sensitivity of decision tree splitting criteria. in ICML. 2000.

[129] Nauck, D. and R. Kruse, Obtaining interpretable fuzzy classification rules from medical data. Artificial Intelligence in Medicine, 1999. 16(2): p. 149-169.

[130] Angelov, P., E. Lughofer, and X. Zhou, Evolving fuzzy classifiers using different model architectures. Fuzzy Sets Syst., 2008. 159(23): p. 3160-3182.

[131] Kapetanovic, I.M., S. Rosenfeld, and G. Izmirlian, Overview of Commonly Used Bioinformatics Methods and Their Applications. Annals of the New York Academy of Sciences, 2004. 1020(1): p. 10-21.

[132] Burges, C.J., A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998. 2(2): p. 121-167.

[133] Suykens, J.A., et al., *Least squares support vector machines*. Vol. 4. 2002: World Scientific.

[134] Wang, L.X. and J.M. Mendel, Generating fuzzy rules by learning from examples. Systems, Man and Cybernetics, IEEE Transactions on, 1992. 22(6): p. 1414-1427.

[135] Vapnik, V.N., *Statistical Learning Theory*1998, New York: John Wiley and Sons.

[136] Lewis, D. and W. Gale. A sequential algorithm for training text classifiers. in Proc. 17th Annual Intl. ACM SIGIR Conf. on R&D in Information Retrieval. 1994.

[137] Fernández, A., et al., Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. Knowledge-Based Systems, 2013.

[138] Xue, B., M. Zhang, and W.N. Browne. Multi-objective particle swarm optimisation (PSO) for feature selection. in Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference. 2012. ACM.

[139] Orriols-Puig, A. and E. Bernadó-Mansilla, Evolutionary rule-based systems for imbalanced data sets. Soft Computing, 2009. 13(3): p. 213-225.

[140] García, V., et al., *Combined effects of class imbalance and class overlap on instance-based classification*, in *Intelligent Data Engineering and Automated Learning–IDEAL 2006*2006, Springer. p. 371-378.

[141] Martín-Félez, R. and R.A. Mollineda, *On the suitability of combining feature selection and resampling to manage data complexity*, in *Current Topics in Artificial Intelligence*2010, Springer. p. 141-150.

[142] Hall, M.A., Correlation-based feature selection for machine learning, 1999, The University of Waikato.

[143] Lin, W.-J. and J.J. Chen, Class-imbalanced classifiers for high-dimensional data. Briefings in bioinformatics, 2013. 14(1): p. 13-26.

[144] Zhang, H., et al., Feature selection for optimizing traffic classification. Computer Communications, 2012. 35(12): p. 1457-1471.

[145] Lusa, L., Class prediction for high-dimensional class-imbalanced data. BMC Bioinformatics, 2010. 11(1): p. 523.

[146] Liu, B., *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*2015: Cambridge University Press.

[147] Liu, B., Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 2012. 5(1): p. 1-167.

[148] Hassan, A., A. Abbasi, and D. Zeng. Twitter Sentiment Analysis: A Bootstrap Ensemble Framework. in Social Computing (SocialCom), 2013 International Conference on. 2013.

[149] Guerra, P.C., W. Meira Jr, and C. Cardie. Sentiment analysis on evolving social streams: How self-report imbalances can help. in Proceedings of the 7th ACM international conference on Web search and data mining. 2014. ACM.

[150] Li, S., et al. Active learning for imbalanced sentiment classification. in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012. Association for Computational Linguistics.

[151] Burns, N., et al., *Sentiment Analysis of Customer Reviews: Balanced versus Unbalanced Datasets*, in *Knowledge-Based and Intelligent Information and Engineering Systems*, A. König, et al., Editors. 2011, Springer Berlin Heidelberg. p. 161-170.

[152] Fan, M., et al., Intrinsic dimension estimation of data by principal component analysis. arXiv preprint arXiv:1002.2050, 2010.

[153] López, V., et al., *Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data.* Fuzzy Sets and Systems, 2015. **258**: p. 5-38.

[154] del Río, S., et al., *A MapReduce Approach to Address Big Data Classification Problems Based on the Fusion of Linguistic Fuzzy Rules.* International Journal of Computational Intelligence Systems, 2015. **8**(3): p. 422-437.

[155] del Río, S., et al., *On the use of MapReduce for imbalanced big data using Random Forest.* Information Sciences, 2014. **285**: p. 112-137.

[156] Hasan, S., S.M. Shamsuddin, and N. Lopes, Machine Learning Big Data Framework and Analytics for Big Data Problems. Int. J. Advance Soft Compu. Appl, 2014. 6(2).

[157] Wienhofen, L., B.M. Mathisen, and D. Roman, Empirical Big Data Research: A Systematic Literature Mapping. arXiv preprint arXiv:1509.03045, 2015.