Comparative and Analysis of Classification Problems

B. Senthil Kumar
Assistant Professor, Department of Computer Science, Sree Narayana Guru College, K. G. Chavadi, Coimbatore, Tamilnadu, India

Dr. R. Gunavathi
Associate Professor & Head, Department of MCA, Sree Saraswathi Thyagaraja College, Pollachi, Tamil Nadu, India

Abstract – Data classification approaches are widely used in many real time applications; every classification approaches have two types of issues such as high dimensional data problem and class imbalance problem. The dataset in every domain may have the huge set of attributes or features and many classification algorithms are believe the training set is evenly distributed in every application. However, the data's are not always balanced. So, it creates many class imbalance problems. In high dimensional data environment, selecting optimal features are required for improving the detection accuracy. And similarly, uneven class distribution problems are also should be considered. A survey on existing approaches to handle the class imbalance and dimensionality issues are discussed. We discussed recent proposals and improvements which potentially could shape the future direction of class imbalance learning and dimensionality reduction for better classification. We also found out that the advancement of machine learning techniques would mostly benefit the health data computing in addressing the class imbalance and dimensionality problems which are inevitably presented in many real world applications especially in healthcare domain.

Index Terms – Data Mining, Classification, Class Imbalance, Feature Extraction, Dimensionality Reduction, Machine Learning.

1. INTRODUCTION

Data mining is the powerful tool to handle different types of high dimensional data. This is suitable for different types of applications such as medical, finance, education and so on. Based on the type of application, different types of data mining algorithms are used. In the paper [1] authors reviewed various data mining techniques and its application areas. When considering health care data, the complications and snag are high due to its size and uncertain nature. The health care data classification includes heart disease, diabetes and many, these data's are most popular research area in the in the data classification. Authors in [2] analyzed the survey and concluded that these types of domains develop dimensionality and class imbalance problems. High-dimensionality reduction is one of the major concerns, which is a part of data preprocessing. And this is an important task in several applications. Another issue in the classification process is it needs a complete training data. In this paper, different types of dimensionality reduction and class imbalance reduction techniques are discussed. The work of this paper is segmented into three parts; one is dimensionality reduction techniques, class imbalance classification techniques and finally the classification algorithms using dimensionality reduction techniques and class imbalance reduction techniques.

1.1 Classification

In data mining, classification is the supervised learning process, which is used to classify database records into a number of predefined classes based on training samples. In Health Care dataset, a patients electronic health records can be classified based on its features and the prior classes. It contains several iterations and feature classes. Based on the training samples the classifiers classify the Health Care records. The most popular classification algorithms in the healthcare such as K-nearest neighbor, Artificial neural networks, Fuzzy c-means, Naive Bayes classifier , Support vector machines, Decision Trees [3] etc.,

1.2 Dimensionality reduction

The high dimensional dataset often tough to perform accurate classification in the medical field, many classification algorithms were associated with the set of popular feature reduction techniques like SVD, PCA, LDA etc., these algorithms helps to handle the large datasets by eliminating redundant and unwanted attributes.

1.3 Class imbalance Problem

Many classification techniques assume that the given training dataset is evenly distributed among given class. But, the training dataset have uneven count of training samples at each class. The majority class is often assigned at the time of classification. So, it creates many class imbalance problems. The minor classes may often eliminate and misclassified into them. This occurs in many ranges of applications like medical diagnosis, financial data classifications, intrusion or anomaly detection approaches etc. many classifiers has the aim of increasing accuracy and reducing the iterations, but none have concentrated on the class imbalance with dimensionality reduction problems.

The paper is organized as follows: In Section II we discuss some existing work on dimensionality reduction techniques.

The class imbalance reduction techniques are presented in Section 3. Section 4 presents the problem summary with the classification algorithms. Section 5 describes the conclusion of the paper.

2. SURVEY ON DIMENSIONALITY REDUCTION TECHNIQUES

The effective feature selection methods are the success of data mining applications, specifically in disease diagnosis applications. Data classification consists of three major steps such as data pre-processing, data analysis and result interpretation. The dimensionality reduction process is falls under the data pre-processing. This section describes the most popular and standard approaches for dimensionality reduction. The dimensionality reduction techniques are classified into linear, non-linear techniques, Sampling dimensionality reduction and Similarity measure dimensionality reduction.

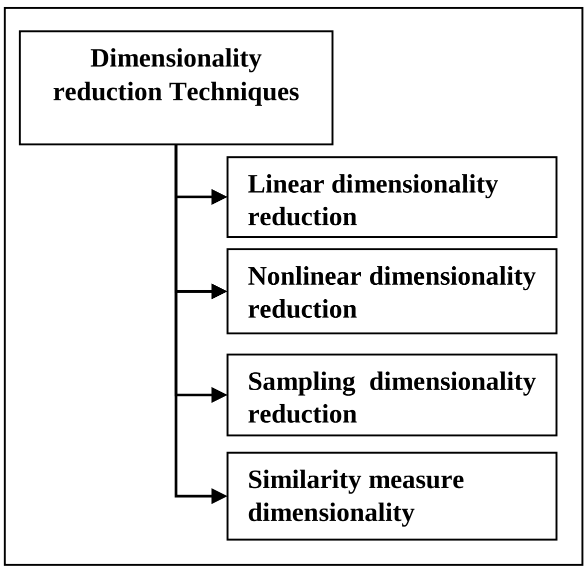

Fig 1.0 Dimensionality reduction technique

2.1 Linear dimensionality reduction

Principal Component Analysis (PCA) is a well established method for dimensionality reduction. It derives new variables (in decreasing order of importance) that are linked by linear combinations of the original variables and are uncorrelated. Several models and techniques for data reduction based on PCA have been proposed in [4]. Authors in [5] proposed a maximum likelihood approach to the multi-size PCA problem. The covariance based approach was extended to estimate errors within the resulting PCA decomposition. Instead of making all the vectors of fixed size and then computing a covariance matrix, they directly estimate the covariance matrix from the multi-sized data using nonlinear optimization.

In paper [6] authors studied the recognition accuracy and the execution times of two different statistical dimensionality reduction methods applied to the biometric image data, which are: PCA and linear discriminant analysis (LDA). The learning algorithm that has been used to train and recognize the images is a support vector machine with linear and polynomial kernel functions. The main drawback of reducing dimensionality with PCA is that it can only be used if the original variables are correlated and homogeneous, if each component is guaranteed to be independent and if the dataset is normally distributed. If the original variables are not normalized, PCA is not effective.

The Sparse Principal Component Analysis (SPCA) [7] is an improvement of the classical method of PCA to overcome the problem of correlated variables using the LASSO technique. LASSO is a promising variable selection technique, producing accurate and sparse models. SPCA is based on the fact that PCA can be written as a regression problem where the response is predicted by a linear combination of the predictors. Therefore, a large number of coefficients of principal components become zero, leading to a modified PCA with sparse loading. Many studies on data reduction based on SPCA have been presented.

In [8] authors proposed an iterative algorithm named sparse PCA via regularized SVD (sPCArSVD) that uses the close connection between PCA and singular value decomposition (SVD) of the data matrix and extracts the PCs through solving a low rank matrix approximation problem.

Authors in [9] proposed a method based on sparse principal component analysis for finding an effective sparse feature principal component (PC) of multiple physiological signals. This method identifies an active index set corresponding to the non-zero entries of the PC, and uses the power iteration method to find the best direction. Singular Value Decomposition (SVD) is a powerful technique for dimensionality reduction. It is a particular case of the matrix factorization approach and it is therefore also related to PCA. The key issue of an SVD decomposition is to find a lower dimensional feature space by using the matrix product $U \times S \times V$, where U and V are two orthogonal matrices and S is a diagonal matrix with $m \times m$, $m \times n$, and $n \times n$ dimensions, respectively. SVD retains only r _ n positive singular values of low effect to reduce the data, and thus S becomes a diagonal matrix with only r non-zero positive entries, which reduces the dimensions of these three matrices to $m \times r$, $r \times r$, and $r \times n$, respectively.

Many studies on data reduction have been presented which are built upon SVD, such as the ones used in [10]. Authors in [11] developed a dimensionality reduction approach by applying the sparsified singular value decomposition (SSVD). Their paper demonstrates how SSVD can be used to identify and remove nonessential features in order to facilitate the feature selection phase, to analyze the application limitations and the computational complexity. However, the application of SSVD on large datasets showed a loss of accuracy and makes it difficult to compute the eigenvalue decomposition of a matrix product ATA, where A is the matrix of the original data.

2.2. Nonlinear dimensionality reduction

A vast literature devoted to nonlinear techniques has been proposed to resolve the problem of dimensionality reduction, such as manifold learning methods, e.g., Locally Linear Embedding (LLE), Isometric mapping (Isomap), Kernel PCA (KPCA), Laplacian Eigenmaps (LE), and a review of these methods is summarized in [12].

KPCA in [13] is a nonlinear generalization of PCA in a high-dimensional kernel space constructed by using kernel functions. By comparing with PCA, KPCA computes the principal eigenvectors using the kernel matrix, rather than the covariance matrix. A kernel matrix is done by computing the inner product of the data points. LLE in [14] is a nonlinear dimensionality reduction technique based on simple geometric intuitions. This algebraic approach computes the low-dimensional neighborhood preserving embeddings. The neighborhood is preserved in the embedding based on a minimizing cost function in input space and output space, respectively. Isomap in [15] explores an underlying manifold structure of a dataset. The geodesic distance is determined as the length of the shortest path along the surface of the manifold between two data points. It first constructs a neighborhood graph between all data points based on the connection of each point to all its neighbors in the input space. Then, it estimates geodesic distances of all pairs of points by calculating the shortest path distances in the neighborhood graph. Finally, multidimensional scaling (MDS) is applied to the arising geodesic distance matrix to find a set of low-dimensional points that greatly match such distances.

2.3. Sampling dimensionality reduction

Other widely used techniques are based on sampling. They are used for selecting a representative subset of relevant data from a large dataset. In many cases, sampling is very useful because processing the entire dataset is computationally too expensive. In general, the critical issue of these strategies is the selection of a limited but representative sample from the entire dataset. Various random, deterministic, density biased sampling, pseudo-random number generator and sampling from non-uniform distribution strategies exist in the literature [16]. However, very little work has been done on the Pseudorandom number generator and sampling from nonuniform distribution strategies, especially in the multi-dimensional case with heterogeneous data. Naive sampling methods are not suitable for noisy data which are part of real-world applications, since the performance of the algorithms may vary unpredictably and significantly.

The random sampling approach effectively ignores all the information present in the samples which are not part of the reduced subset [17]. An advanced data reduction algorithm should be developed in multi-dimensional real-world datasets, taking into account the heterogeneous aspect of the data. Both approaches [18] [19] are based on sampling and a probabilistic representation from uniform distribution strategies.

The authors of [20] proposed a method to reduce the complexity of solving Partially Observable Markov Decision Processes (POMDP) in continuous state spaces. The paper uses sampling techniques to reduce the complexity of the POMDPs by reducing the number of state variables on the basis of samples drawn from these distributions by means of a Monte Carlo approach and conditional distributions.

The authors in [21] applied dimensionality reduction to a recent movement representation used in robotics, called Probabilistic Movement Primitives (ProMP), and they addressed the problem of fitting a low-dimensional, probabilistic representation to a set of demonstrations of a task. The authors fitted the trajectory distributions and estimated the parameters with a model-based stochastic using the maximum likelihood method. This method assumes that the data follow a multivariate normal distribution which is different from the typical assumptions about the relationship between the empirical data. The best we can do is to examine the sensitivity of results for different assumptions about the data distribution and estimate the optimal space dimension of the data.

2.4. Similarity measure dimensionality reduction

There are other widely used methods for data reduction based on similarity measures [22]. According to [23], the presence of redundant or noisy features degrades the classification performance, requires huge memory, and consumes more computational time. Authors in [24] proposes a three-stage dimensionality reduction technique for microarray data classification using a comparative study of four different classifiers, multiple linear regression (MLR), artificial neural network (ANN), k-nearest neighbor (k-NN), and naive Bayesian classifier to observe the improvement in performance. In their experiments, the authors reduce the dimension without compromising the performance of such models.

Authors in [25] proposed a dimensionality reduction method that employs classification approaches based on the k-nearest neighbor rule. The effectiveness of the reduced set is measured in terms of the classification accuracy. This method attempts to derive a minimal consistent set, i.e., a minimal set which correctly classifies all the original samples discussed the ongoing work in the field of pattern analysis for bio-medical signals (cardio-synchronous waveform) using a Radio Frequency Impedance Interrogation (RFII) device for the purpose of user identification.

They discussed the feasibility of reducing the dimensions of these signals by projecting them into various sub-spaces while still preserving inter-user discriminating information, and they compared the classification performance using traditional dimensionality reduction methods such as PCA, independent component analysis (ICA), random projections, or k-SVD-based dictionary learning. In the majority of cases, the authors see that the space obtained based on classification carries merit due the dual advantages of reduced dimension and high classification.

Developing effective clustering methods for high-dimensional datasets is a challenging task [26] studied the topic of dimensionality reduction for k-means clustering that encompasses the union of two approaches:

1) A feature selection-based algorithm selects a small subset of the input features and then the k-means is applied on the selected features.

2) A feature extraction-based algorithm constructs a small set of new artificial features and then the kmeans is applied on the constructed features. The first feature extraction method is based on random projections and the second is based on fast approximate SVD factorization. Authors in [27] developed a tensor factorization based on a clustering algorithm (k-mean), referred to as Dimensionality Reduction Assisted Tensor Clustering (DRATC). In this algorithm, the tensor decomposition is used as a way to learn low-dimensional representation of the given tensors and, simultaneously, clustering is conducted by coupling the approximation and learning constraints, leading to the PCA Tensor Clustering and Non-negative Tensor Clustering models. In this study, we develop a sampling-based dimensionality reduction technique that can deal with very high-dimensional datasets. The proposed approach takes into account the heterogeneous aspects of the data, and it models the different multivariate data distributions using the theory of Copulas. It maintains the integrity of the original information and reduces effectively and efficiently the original high-dimensional datasets.

3. SURVEY ON CLASS IMBALANCE PROBLEMS

In general, there are two strategies 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 1.0 provides a detailed summary on several notable previous works in class imbalance classification along with advantages and limitations of each strategy. Fig 2.0 shows the classification of class imbalance approaches.

3.1. 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 oversampling 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 [28]. The existing study [29] 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 [30] which reflected in the problem of class complexity or overlapping in class imbalance. A method called ACO Sampling which applied an ant colony to optimize undersampling for the classification of highly imbalanced microarray data has been proposed by [31].

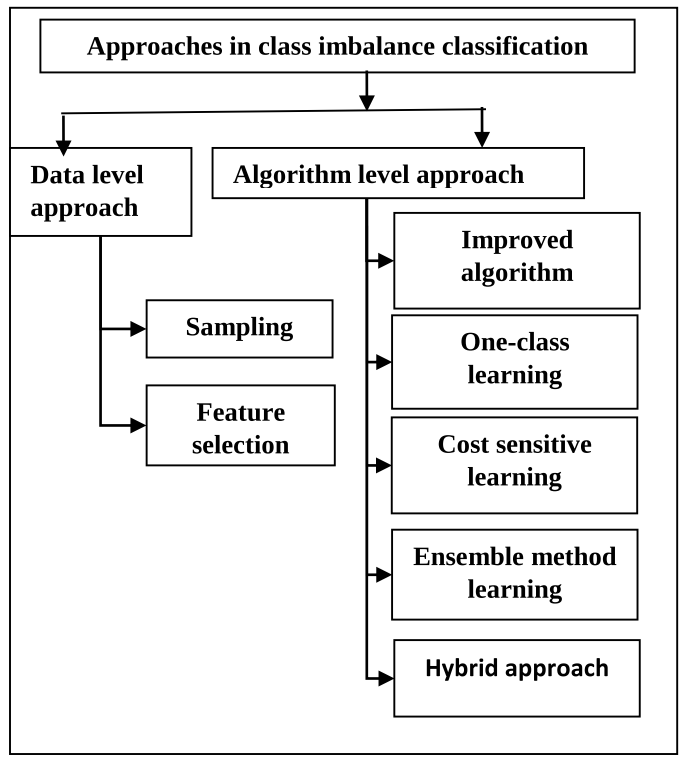

Fig 2.0 class imbalance classification approaches

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 [32]. For undersampling scheme, the several algorithms use clustering to choose representatives training examples to gain better accuracy prediction for minority class. Nevertheless, the elimination of examples from a class could lead to loss of potentially important information about the class, while in examples replication (oversampling), 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 [33].

Feature selection removes irrelevant, redundant or noisy data [34] 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. Authors from [35] 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 chisquare, correlation coefficient and odds-ratios.

Authors	Year	Name	Advantages	Disadvantages
Technique	**A. Data-level Approach**			
	1. Sampling			
29	1999	MLSMOTE	Straight forward approach and widely used in many domain applications	Risk of over fitting
30	1999	Diversified sensitivity-based undersampling		
31	2013	ACOsampling with Ant Colony		
32	2002	SMOTE		
33	2009	Evolutionary undersampling		
	2. Feature Selection			
34	2011	Minority class feature selection	Helpful in alleviating class overlapping problem	Extra computational cost due to included preprocessing task
35	2012	Density-based feature selection		
36	2008	FAST; roc-based feature selection		
37	2010	CFS; correlation feature selection		

Table 1.0 data level class imbalance reduction approaches

Another attempt at applying feature selection to solve class imbalance problem is from earlier concept who proposed for a ROC based feature selection, instead of classification accuracy to assess classification performances. As discussed in [36] [37] a novel feature subset selection based on correlation measure to handle small sample in class imbalance problem.

B. Algorithm level approach for handling class imbalance problem

Generally, the algorithm-level methods could be categorized 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 table 3.0 summarizes the categories of algorithm level approaches.

The cost effective learning with class imbalance dataset is always a big challenge in several machine learning algorithms. The cost sensitive approaches are widely studied in the literature [48], which optimized with the earlier classifiers to reduce the false alarm. Cost effective learning algorithms are developed on numerous classifiers like optimized cost sensitive SVM [49], neural network. Authors in [50] applied SVM for reducing the cost listed in table 3.0. Ensemble learning is another option for class imbalance issues and these techniques trained many classifiers on training samples and their assessments are combined to produce the final classification decision. The ensemble approaches can be described as boosting or bagging. 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 *n* number of classifiers for each *n* variations of the training samples. The outcome is produced by combining the output all the classifiers. The Boosting approaches 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.

Authors	Year	Name	Advantages	Disadvantages
	B. Algorithm level approach			
	1. Improved Algorithm			
38	2010	Argument-based rule learning	Effective methods due to modified algorithms to learn exclusively from imbalance class distribution	Might need preprocessing tasks to balance out skewed class distribution
39	2004	Dissimilarity-based learning		
40	2006	Fuzzy Classifier		
41	2006	z-SVM		
42	2009	Hierarchical Fuzzy rule		
43	2012	Class conditional nearest neighbour distribution		
44	2011	k-NN with Examplar Generalization		
45	2011	Weighted nearest-neighbour classifier		
	2. One-class Learning			
46	1995	One class learning	Simple methods	Not efficient when applied with classification algorithms that must learn from prevalent class
47	2012	Class Conditional Nearest Neighbor Distribution (CCNND)		
	3. Cost-sensitive Learning			
48	2005	Near Bayesian SVM	Simple, fast processing method	Ineffective if real cost are not available Extra cost introduced if
49	2013	Cost sensitive learning with SVM		
50	2011	Cost sensitive NN with PSO		

51	2012	SVM for Adaptively Asymmetrical Misclassification Cost		cost exploration is needed when error cost is not known

	1. **Ensemble Method**			
52	2013	SMOTE and feature selection	Versatile approaches	Complexity grows with the use of more classifiers Diversity concept is difficult to achieve
53	2006	Ensemble GA		
54	2014	Ensemble for financial problem		
55	2006	Boosting with SVM ensemble, RUSBoost, SMOTEBoost		
	2. **Hybrid Approach**			
56	2010	FTM-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
57	2013	F-measure based learning		
58	2008	Linguistic fuzzy rule		
59	2007	Fuzzy classifier e-algo		
60	2008	Fuzzy rule extraction with GA		
61	2006	Neuro fuzzy		
62	2013	Neural net medical data		
63	2010	Neural networks with SMOTE		
64	2012	Case-based classifier kNN for medical data		
65	2008	NN trained with BP and PSO for		
66	2004	Dependency tree kernels		
67	2000	Exploiting cost sensitive in tree		
68	2005	Undersampling and GA for SVM		

Table 2.0 Algorithm Level Class Imbalance Reduction Approaches

3. CONCLUSIONS

In this paper, we surveyed various techniques and algorithms for effective classifier in terms of two problems such as class imbalance and dimensionality reduction. In the research it is observed that many work falls under the class imbalance issue reduction and dimensionality reduction issues, but there is none of the approaches were concentrated on both issues. So creating a classifier for the high-dimensional data with class imbalance problem will be an interesting area for the future research. Many class imbalance problems are still not sufficient for multiple class imbalance problems.

This survey presents an overview on dimensionality reduction and class imbalance classification with the possible issues and outcomes. It describes the main issues that hamper the classifier performance due to these two problems. Research gap in previous works are discussed along with merits and demerits. At last, the observed current trends together with recent advancements in class imbalance classification and effective feature selection for reducing dimensionality are presented. This paper also suggests several potential developments in the domain such as the machine learning for health care dataset analysis.

REFERENCES

[1] Liao, Shu-Hsien, Pei-Hui Chu, and Pei-Yuan Hsiao. "Data mining techniques and applications–A decade review from 2000 to 2011." *Expert systems with applications* 39.12 (2012): 11303-11311.

[2] Senthil Kumar, B & Gunavathi R., Dr. (2016). A Survey on Data Mining Approaches to Diabetes Disease Diagnosis and Prognosis. IJARCCE. 5. 463-467. 10.17148/IJARCCE.2016.512105.

[3] Wu, Xindong, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan et al. "Top 10 algorithms in data mining." Knowledge and information systems 14, no. 1 (2008): 1-37.

[4] Shi, Feiyu, Menghua Zhai, Drew Duncan, and Nathan Jacobs. "Mpca: Em-based pca for mixed-size image datasets." In *Image Processing (ICIP), 2014 IEEE International Conference on*, pp. 1807-1811. IEEE, 2014.

[5] Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." *Wiley interdisciplinary reviews: computational statistics* 2, no. 4 (2010): 433-459.

[6] Kerdprasop, Nittaya, Ratiporn Chanklan, Anusara Hirunyawanakul, and Kittisak Kerdprasop. "An empirical study of dimensionality reduction methods for biometric recognition." In *Security Technology (SecTech), 2014 7th International Conference on*, pp. 26-29. IEEE, 2014.

[7] Jenatton, Rodolphe, Guillaume Obozinski, and Francis Bach. "Structured sparse principal component analysis." In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 366-373. 2010.

[8] Shen, Haipeng, and Jianhua Z. Huang. "Sparse principal component analysis via regularized low rank matrix approximation." *Journal of multivariate analysis* 99.6 (2008): 1015-1034.

[9] Shen, Haipeng, and Jianhua Z. Huang. "Sparse principal component analysis via regularized low rank matrix approximation." *Journal of multivariate analysis* 99, no. 6 (2008): 1015-1034.

[10] Yin, Yingjie, De Xu, Xingang Wang, and Mingran Bai. "Online state-based structured SVM combined with incremental PCA for robust visual tracking." *IEEE transactions on cybernetics* 45, no. 9 (2015): 1988-2000.

[11] Wall, Michael E., Andreas Rechtsteiner, and Luis M. Rocha. "Singular value decomposition and principal component analysis." *A practical approach to microarray data analysis*. Springer US, 2003. 91-109.

[12] Gisbrecht, A., & Hammer, B. (2015). Data visualization by nonlinear dimensionality reduction. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5, 51–73.

[13] Kuang, F., Zhang, S., Jin, Z., & Xu, W. (2015). A novel svm by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection. Soft Computing, (pp. 1–13)

[14] Hettiarachchi, R., & Peters, J. (2015). Multi-manifold LLE learning in pattern recognition. Pattern Recognition, 48, 2947–2960.

[15] Zhang, T., Du, Y., Huang, T., & Li, X. (2016). Stochastic simulation of geological data using isometric mapping and multiple-point geostatistics with data incorporation. Journal of Applied Geophysics, 125, 14–25.

[16] Rubinstein, R. Y., & Kroese, D. P. (2011). Simulation and the Monte Carlo method volume 707. John Wiley & Sons

[17] Whelan, M., Khac, N. A. L., Kechadi, M. et al. (2010). Data reduction in very large spatio-temporal datasets. In Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE), 2010 19th IEEE International Workshop on (pp. 104–109). IEEE.

[18] Mejía-Lavalle, Manuel, et al. "Toward Optimal Pedagogical Action Patterns by Means of Partially Observable Markov Decision Process." *Mexican International Conference on Artificial Intelligence*. Springer, Cham, 2016.

[19] McGhan, Catharine LR, Ali Nasir, and Ella M. Atkins. "Human intent prediction using markov decision processes." *Journal of Aerospace Information Systems* (2015).

[20] Fakoor, R., & Huber, M. (2012). A sampling-based approach to reduce the complexity of continuous state space POMDPs by decomposition into coupled perceptual and decision processes. In Machine Learning and Applications (ICMLA), 2012 11th International Conference on (pp. 687–692). IEEE volume 1.

[21] Colom´e, A., Neumann, G., Peters, J., & Torras, C. (2014). Dimensionality reduction for probabilistic movement primitives. In Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on (pp. 794– 800). IEEE.

[22] Wencheng, P. (2010). Theory and application of parameter self-optimizing intelligent sampling method. In Signal Processing (ICSP), 2010 IEEE 10th International Conference on (pp. 66–69). IEEE.

[23] Pirolla, F. R., Felipe, J., Santos, M. T., & Ribeiro, M. X. (2012). Dimensionality reduction to improve contentbased image retrieval: A clustering approach. In Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on (pp. 752–753). IEEE.

[24] Zhang, J., Nie, X., Hu, Y., Liu, S., Tian, Y., & Wu, L. (2010). A method for land surveying sampling optimization strategy. In Geoinformatics, 2010 18th International Conference on (pp. 1–5). IEEE.

[25] Deegalla, S., Bostr¨om, H., & Walgama, K. (2012). Choice of dimensionality reduction methods for feature and classifier fusion with nearest neighbor classifiers. In Information Fusion (FUSION), 2012 15th International Conference on (pp. 875–881). IEEE.

[26] Boutsidis, C., Zouzias, A., Mahoney, M. W., & Drineas, P. (2015). Randomized dimensionality reduction for-means clustering. Information Theory, IEEE Transactions on, 61, 1045–1062.

[27] Sun, Y., Gao, J., Hong, X., Guo, Y., & Harris, C. J. (2014). Dimensionality reduction assisted tensor clustering. In Neural Networks (IJCNN), 2014 International Joint Conference on (pp. 1565–1572). IEEE.

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

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

[30] 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

[31] Chawla, N.V., et al., SMOTE: synthetic minority over-sampling technique. arXiv preprint arXiv:1106.1813, 2002.
[32] García, S. and F. Herrera, Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. Evolutionary Computation, 2009. 17(3): p. 275-306.
[33] Yin, L., et al., Feature selection for high-dimensional imbalanced data. Neurocomputing, 2013. 105(0): p. 3-11.
[34] 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.
[35] 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.
[36] 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.
[37] 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.
[38] Luukka, P., Nonlinear fuzzy robust PCA algorithms and similarity classifier in bankruptcy analysis. Expert Systems with Applications, 2010. 37(12): p. 8296-8302.
[39] .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.
[40] Visa, S., Fuzzy Classifiers for Imbalanced Data Sets, in Department of Electrical and Computer Engineering and Computer Science2006, Univeristy of Cincinnati: Cincinnati.
[41] 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.
[42] 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.
[43] 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.
[44] Li, Y. and X. Zhang, Improving k nearest neighbor with exemplar generalization for imbalanced classification, in Advances in knowledge discovery and data mining2011, Springer. p. 321-332.
[45] Candès, E.J., et al., Robust principal component analysis? Journal of the ACM (JACM), 2011. 58(3): p. 11.
[46] Japkowicz, N., C. Myers, and M. Gluck. A novelty detection approach to classification. In IJCAI. 1995.
[47] 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.
[48] Jolliffe, I., Principal component analysis2005: Wiley Online Library.
[49] Cao, P., D. Zhao, and O. Zaiane, An Optimized Cost-Sensitive SVM for Imbalanced Data Learning, in Advances in Knowledge Discovery and Data Mining2013, Springer. p. 280- 292.
[50] Cao, P., D. Zhao, and O. Zaiane, A PSO-based Cost-Sensitive Neural Network for Imbalanced Data Classification. 2011.
[51] 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.
[52] Yang, P., et al., Ensemble-based wrapper methods for feature selection and class imbalance learning, in Advances in Knowledge Discovery and Data Mining2013, Springer. 544-555.
[53] Yu, E. and S. Cho, Ensemble based on GA wrapper feature selection. Computers & Industrial Engineering, 2006. 51(1): p. 111-116.
[54] .Liao, J.-J., et al., An ensemble-based model for two-class imbalanced financial problem. Economic Modeling, 2014. 37(0): p. 175-183.
[55] Liu, Y., A. An, and X. Huang, Boosting prediction accuracy on imbalanced datasets with SVM ensembles, in Advances in Knowledge Discovery and Data Mining2006, Springer. p. 107-118.
[56] 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.
[57] Martino, M.D., et al., Novel Classifier Scheme for Unbalance Problems. Pattern Recognition Letters, 2013.
[58] 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.
[59] 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.
[60] 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.
[61] 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.
[62] 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.
[63] 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.
[64] 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.
[65] 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.
[66] 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.
[67] Drummond, C. and R.C. Holte. Exploiting the cost (in) sensitivity of decision tree splitting criteria. in ICML. 2000.
[68] 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