A Novel Distribution Analysis for SMOTE Oversampling Method in Handling Class Imbalance

Dina Elreedy[(✉)] and Amir F. Atiya

Faculty of Engineering, Cairo University, Giza, Egypt
dinaelreedy@email.wustl.edu, amir@alumni.caltech.edu

Abstract. Class Imbalance problems are often encountered in many applications. Such problems occur whenever a class is under-represented, has a few data points, compared to other classes. However, this minority class is usually a significant one. One approach for handling imbalance is to generate new minority class instances to balance the data distribution. The Synthetic Minority Oversampling TEchnique (SMOTE) is one of the dominant oversampling methods in the literature. SMOTE generates data using linear interpolation between minority class data point and one its K-nearest neighbors. In this paper, we present a theoretical and an experimental analysis of the SMOTE method. We explore the accuracy of how faithful SMOTE method emulates the underlying density. To our knowledge, this is the first mathematical analysis of the SMOTE method. Moreover, we study the impacts of the different factors on generation accuracy, such as the dimension of data, the number of examples, and the considered number of neighbors K on both artificial, and real datasets.

Keywords: Class imbalance · Minority class · Over-sampling · SMOTE

1 Introduction

Imbalanced learning is encountered when one of the classes is represented fewer than others. Datasets may be naturally unbalanced such as medical diagnosis [15] and fraud detection [2], or data collection process may be too expensive such as detection of system failures. Yang et al. [22] have declared imbalanced learning as one of the ten most challenging problems in data mining. Handling class imbalance is challenging since there is a trade-off between the overwhelming influence of the majority class patterns, and an overemphasis on just a few minority class patterns.

Standard classifiers are biased towards the majority class examples while sacrificing minority class accuracy since such classifiers aim to maximize the overall classification accuracy without considering class distributions. The three main approaches for handling data imbalance problem in literature are: cost sensitive approach, algorithm level approach, and data level approach.

The cost sensitive approach uses cost matrices to set misclassification costs according to the importance of the class and degree of imbalance. Examples of work on the cost sensitive approach include AdaCost [7], and the work done by Chawla et al. [4].

The algorithm level approach adapts the classification algorithm's to handle the class imbalance problem. For example there is work on modifying the K nearest neighbor classifier (KNN) [23], other work on adapting decision trees [18], some approaches that modify support vector machines (SVM) [16], all these methods seeks to focus on minority class.

Finally, the data level approach is based on modifying the data distribution in order to balance minority and majority classes. Data level approach is the most popular approach for handling class imbalance, since it is a simple approach that can be applied independently of the classifier being used. Data level methods balance distributions by either removing some of the majority class data points (under-sampling), or adding more of minority class instances (over-sampling).

Under-sampling can be done randomly or using some heuristics such as: the condensed nearest neighbor rule [8] and one-sided selection [1]. However, under-sampling can be considered precarious since potential important information could be lost when removing majority class examples. On the other hand, over-sampling can be done by randomly replicating minority class patterns, or by generating new minority class patterns [1]. One of the most popular over-sampling methods is "Synthetic Minority Over-sampling Technique", or SMOTE [3]. SMOTE generates patterns from the minority class by performing a linear interpolation between a minority class pattern, and a randomly chosen one of its K-nearest neighbors. A detailed description of the SMOTE method is presented in Sect. 2.

Although there is much work in literature studying sampling methods handling class imbalance problem (see the reviews [15,19], and [12], most of this work provides empirical analysis only, and there is little work, if any, that provides a theoretical analysis of data sampling methods.

One of the empirical studies is done by Luengo et al. [20]. In this work, the authors analyze the behavior of different sampling methods including: SMOTE, its extension, SMOTE-ENN, and an evolutionary under-sampling method EUSCHC [11], by measuring the degree of feature overlapping of the different classes, and class separability and its geometrical properties. However, these measures do not consider distributional issues of the generated data.

Another empirical analysis is performed in [6], the authors analyze different under-sampling, over-sampling methods and hybrid methods using both over-sampling and under-sampling for Alzheimers disease dataset. Their experimental analysis includes: random over-sampling, SMOTE, random under-sampling and K-Medoids under-sampling, a proposed clustering-based under-sampling method. Their results show that the subtle methods such as SMOTE and K-Medoids outperform random over-sampling and random under-sampling.

A lot of methods have extended SMOTE technique [3] due to its simplicity and performance. For example, two variations of Borderline SMOTE are pre-

sented in [13], Borderline-SMOTE1 and Borderline-SMOTE2. In these methods only the minority examples near the classification boundary are over-sampled, since the near-boundary examples tend to be more informative.

Another model is the so-called Adaptive Synthetic Sampling Approach for Imbalanced Learning (ADASYN) [14]. It uses a weighted distribution for different minority class examples according to their level of difficulty in learning, where more synthetic data are generated for minority class points that are harder to learn. Difficulty of a minority example learning is determined by the class composition of the K nearest neighbors.

A recently developed over-sampling method named Sampling WIth the Majority (SWIM) handles extreme class imbalance [21]. The authors of that paper utilize the distribution of majority class to generate synthetic minority class samples in new under-represented regions of minority class. Their proposed method (SWIM) achieves that by generating synthetic data at the same Mahalanbois distance from the majority class as the minority class sample.

Although the SMOTE generation mechanism is extensively used in literature [13,14,21], and [17], the SMOTE method has a major drawback that it is not grounded on a solid mathematical theory [3]. Consequently, in this work, we aim to provide a comprehensive analysis of the SMOTE method. Specifically, our goals are the following:

– Develop a mathematical analysis of SMOTE, and test the degree of its emulation to the underlying distribution (by checking its moments).
– Provide a detailed experimental study of SMOTE, exploring the factors that affect its accuracy (in mimicking the distribution).

The paper is organized as follows: Sect. 2 introduces SMOTE method stating its advantages and potential drawbacks. Section 3 presents a mathematical analysis to derive the distribution of the patterns generated by SMOTE. Then, the experimental analysis of SMOTE is presented in Sect. 4. Finally, Sect. 5 concludes the paper and presents potential future work.

2   SMOTE Method

The SMOTE over-sampling procedure consists of the following simple steps:

– For each pattern X_0 from the minority class do the following:
  • Pick one of its K nearest neighbors X (belonging to the minority class also).
  • Create a new pattern Z on a random point on the line segment connecting the pattern and the selected neighbor, as follows:

$$Z = X_0 + w(X - X_0) \tag{1}$$

where w is a uniform random variable in the range $[0, 1]$.

Figure 1 shows an example of patterns generated by SMOTE. In contrast, this figure shows extra patterns generated from the original distribution. It can be observed that the SMOTE generated patterns are more contracted than the patterns generated from the true distribution. This is because the SMOTE generation process by linear interpolation causes them to be inward-placed. In addition, SMOTE generated patterns are allocated only on the line segments connecting the K neighbors, creating an unrealistic graph shape, where edges are studded with data points and internal portions are void of them. This problem is accentuated even more in higher dimensions. Figure 1 shows how SMOTE generated patterns cluster around some paths, with some empty spaces around them. However, means of original distribution and SMOTE generated examples' distribution are very close as shown in Fig. 2.
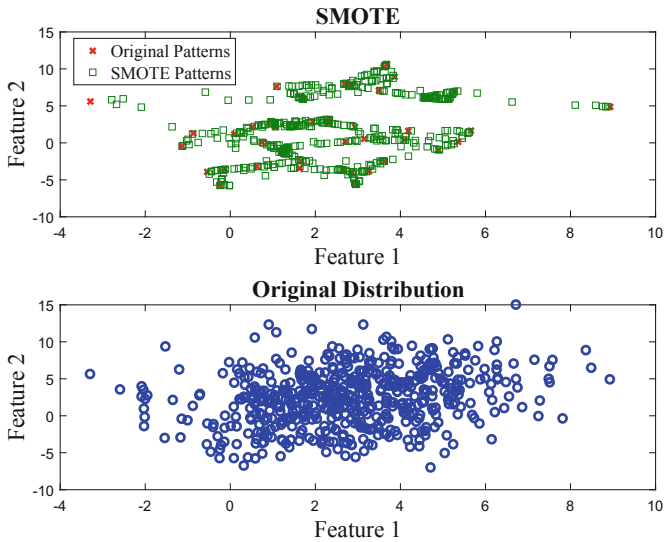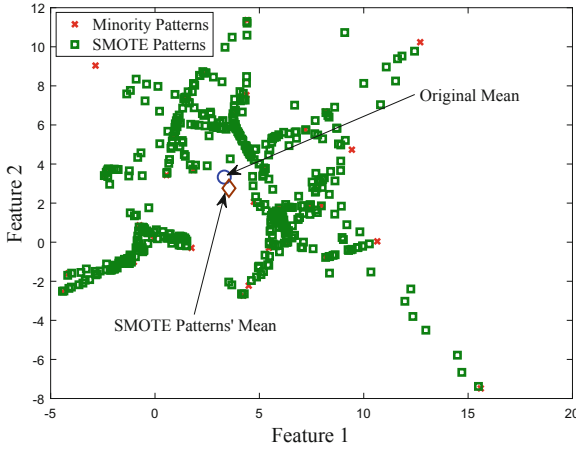

Fig. 1. SMOTE generated vs. original patterns

Another problem is that SMOTE could generate patterns in the decision regions of the majority class, this is more likely to occur in case of overlapping classes.

3   Theoretical Analysis of SMOTE

In this section, we present a theoretical analysis for SMOTE method in order to provide some mathematical basis. The success of SMOTE as a valid sampling algorithm hinges on its ability to generate patterns obeying a distribution close to the true one. We will investigate this issue here. Since the mean vector and the covariance matrix are the two major parameters characterizing any distribution,

Fig. 2. Original distribution mean vs. SMOTE generated patterns' mean

we derive approximate formulas for the mean and the covariance matrix of patterns generated using SMOTE, and compare them with the true distribution's parameters.

Let $\Delta = X - X_0$, then:

$$Z = X_0 + w\Delta \tag{2}$$

where w is a uniformly generated number in $[0, w^*]$. When w^* equals to zero, distances would be zero since generated patterns are identical to original patterns. The parameter w^*, typically greater than or equal one, allows us to both extrapolate and interpolate on the line connecting the pattern X_0 and its randomly selected neighbor X. If $w^* = 1$ then this reverts back to the original SMOTE (applying only interpolation). If $w^* > 1$, then we can go beyond point X_0, i.e. we are allowing some level of extrapolation.

The basic idea for the analysis is approximating the probability density of minority class $p(X)$ using Taylor series around the point X_0 as proposed in [10]. The final approximation of the mean and covariance matrix of the generated pattern vector Z are given by Eqs. (3) and (4) respectively.

$$E[Z] \approx \mu_{X_0} + \frac{Cw^{*2}}{2} \int_{X_0} p(X_0)^{\frac{-2}{d}} \frac{\partial p(X_0)}{\partial X} \, \mathrm{d}X_0 \tag{3}$$

where $[\frac{\partial p(X)}{\partial X}]^T = (\frac{\partial p(X)}{\partial x_1}, ..., \frac{\partial p(X)}{\partial x_d})$.

$$\Sigma_Z = \Sigma_{X_0} + \frac{Cw^{*2}}{3} \int_{X_0} p(X_0)^{1-\frac{2}{d}} \, \mathrm{d}X_0 I$$

$$+ \frac{C^2 w^{*2}}{3} \int_{X_0} p(X_0)^{\frac{-2}{d}} \frac{\partial p(X_0)}{\partial X} \, \mathrm{d}X_0 \int_{X_0} p(X_0)^{\frac{-2}{d}} \frac{\partial p(X_0)}{\partial X}^T \, \mathrm{d}X_0$$

$$+ \frac{Cw^*}{2} \Big[\int_{X_0} p(X_0)^{-\frac{2}{d}} \frac{\partial p(X_0)}{\partial X} [(X_0 - \mu_{X_0})^T] \, \mathrm{d}X_0$$

$$+ \int_{X_0} p(X_0)^{-\frac{2}{d}} (X_0 - \mu_{X_0}) \frac{\partial p(X_0)}{\partial X}^T \, \mathrm{d}X_0 \Big] \quad (4)$$

where d is the dimension of the pattern vector, μ_{X_0} is the true mean vector of the minority class, Σ_{X_0} is the true covariance function, $p(X_0)$ is the class-conditionl density at point X_0, I is the identity matrix, and C is calculated as follows:

$$C = \frac{N! \Gamma \left(1 + \frac{2}{d}\right)^{\frac{2}{d}} \Gamma \left(K + \frac{2}{d} + 1\right)}{\pi K! (d+2) \Gamma \left(N + \frac{2}{d} + 1\right)} \quad (5)$$

If the true probability density is multivariate Gaussian, then the approximations can be simplified further to the following:

$$E[Z] \approx \mu_{X_0} \quad (6)$$

$$\Sigma_Z = \Sigma_{X_0} + \left[(2\pi)^{\frac{1-d}{2}} \frac{Cw^{*2}}{3} \det^{\frac{1-d}{2d}} (\Sigma_{X_0}) \left(\frac{d}{2d-1}\right)^{\frac{d}{2}} \right. $$

$$\left. - 2\pi C w^* \det^{\frac{1}{d}} (\Sigma_{X_0}) \left(\frac{d}{d-2}\right)^{\frac{d+2}{2}} \right] I \quad (7)$$

From Eq. (7), since the fraction $\frac{d}{d-2}$ is greater than one for any $d > 0$ and $d \neq 2$, hence the second term of the generated examples' covariance matrix Σ_Z would be negative and accordingly, the covariance matrix of SMOTE generated examples Σ_Z would be more contracted (since diagonal elements are smaller) than that of original minority class examples Σ_{X_0}.

From the above formulas one can observe the following:

– The mean vector of SMOTE-generated patterns is very close to the true one.
– The covariance matrix has some discrepancy. It is more contractive than the true one, because of the identity matrix times constant that is subtracted from the true covariance matrix (see Eq. 7). This agrees with the intuitive argument discussed last section, which argues that the SMOTE generation mechanism locates the patterns more inwards.

In order to measure how the covariance matrix of SMOTE-generated patterns diverges from the original covariance matrix, we define Total Variances Difference (TVD) measure. This measure helps us learn the amount and the polarity of the

difference between synthetic and original covariance matrices. It is defined as the difference between the traces of the two covariance matrices. We normalize TVD by dividing by trace of original covariance matrix.

$$TVD = \frac{trace(\Sigma_Z) - trace(\Sigma_{X_0})}{trace(\Sigma_{X_0})} \qquad (8)$$

where the trace of the covariance matrix represents the summation of individual features' variances.

4   Experimental Analysis of SMOTE

4.1   Experiments

To have a more detailed understanding of the quality of SMOTE sampling, and its influencing factors, we set out a simulation study. In these experiments, we generate artificial datasets from multivariate Gaussian distributions, apply SMOTE over-sampling, then estimate the SMOTE-sampled examples' distribution, and compare it to the original distribution.

   To have the analysis general enough, we consider 20 different distributions with different parameters. In all cases we consider the zero mean case, because the mean constitutes a shift in the center of operations, and will therefore be insignificant. However, we consider a variety of 20 different covariance matrices Σ_{X_0} varying between diagonal and off-diagonal ones. For the diagonal matrices, we sample the diagonal elements (eigenvalues) of the covariance matrix sampled from uniform distribution ranging from above zero to 40. Similarly, for the off-diagonal matrices, we first generate a diagonal matrix, named D, where its diagonal elements are randomly sampled. Then, we compute the covariance matrix Σ_{X_0} using the following equation:

$$\Sigma_{X_0} = RDR^T \qquad (9)$$

where R is an orthonormal matrix that is uniformly sampled.

   We studied the effect of the same influencing parameters considered in the previous section, namely the number of original minority examples N, the dimension d, and the K parameter of the KNN. We have separately varied each of the influencing factors, while fixing the others, and in each case we documented the accuracy in the distribution of the generated points. While varying each parameter, the others are set at their "default values", which are as follows: $N = 100, \quad d = 10, \quad K = 5$. We have used over-sampling rate $R = 1$, the over-sampling rate can be defined as the amount of data points generated for each minority pattern.

   Additionally, in these experiments, we have set $w^* = 1$ as used in standard SMOTE method [3] since we are interested in analyzing SMOTE method. However, for w^* can be set greater than 1, so that we can allow some extrapolation which could compensate the contraction of covariance matrix caused by SMOTE.

Algorithm 1. The experimental procedure for the SMOTE empirical analysis

   repeat
      Randomly generate N patterns from the original distribution.
      repeat
         Apply SMOTE to the generated N original patterns.
         Obtain the sample mean vector and covariance matrix of the SMOTE generated examples.
      until L runs executed.
      Average the sample mean vector and covariance matrix over the L inner runs.
   until M runs executed.
   Average mean and covariance estimates over the M outer runs to get final estimates of generated patterns' distribution.
   Compare the final estimated SMOTE generated patterns' distribution to the original distribution using the TVD distribution distance metric.

In order to estimate expectation and covariance of the SMOTE generated patterns, we apply the following procedure:

To measure how close the distribution of the SMOTE-generated patterns to the true distribution, we use the total variances difference (TVD) described and used last section. In our experiments, we set the outer number of runs M to 1000, and the inner number of runs L is set to 1000.

The following figures present the divergence of both empirical and theoretical estimates from the true distribution measured in terms of TVD metric described in last section. Figure 3 shows TVD when exploring the effect of the dimension d. As mentioned before, we fix all other factors at their default values, while varying the dimension. Similarly, Fig. 4 shows the TVD metric for the case of varying the number of minority samples N. Also, Fig. 5 show the TVD metric for the case of varying the K "number of neighbors".

It can be observed from the presented results that SMOTE behavior when varying different factors is similar in case of evaluating this behavior using our mathematical analysis and experimentally.

4.2   Experiments Using Real Data

In the other set of experiments we have applied a similar set-up as discussed on three real world UCI datasets. This provides a test for situations where the distribution is not necessarily Gaussian, and to justify that the derived conclusions apply to more complex situations, since real datasets could be noisy, and they could have sub-concepts for the minority class patterns.

We considered datasets that are originally large. This is in order to have an accurate estimate of the mean and covariance matrix. However, since SMOTE is used primarily for smaller datasets [3], we consider only a small subset (like 50 or 100) of the data, and perform the sampling using these. For example, assume that the dataset has about 10,000 points. We compute the mean and covariance matrix from the 10,000 points and assume these to be approximately the true
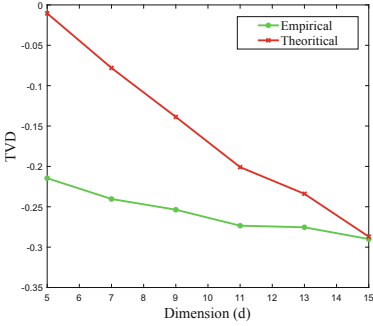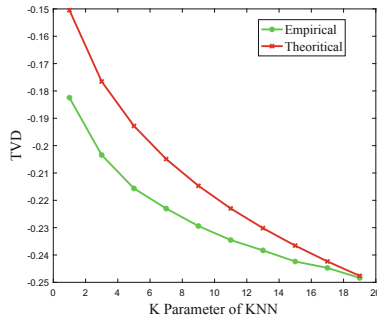
Fig. 3. TVD between empirical and theoretical estimates, and the true distribution versus dimension d
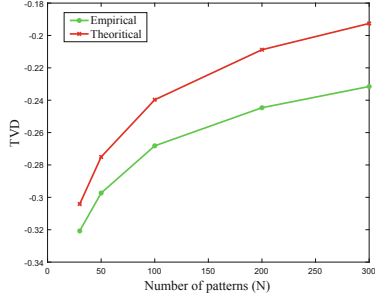
Fig. 4. TVD between empirical and Theoretical estimates, and the true distribution versus number of patterns N

Fig. 5. TVD between empirical and theoretical estimates, and the true distribution versus K parameter of KNN in SMOTE

ones (due to the large number of points). Consider that we test the case of number of patterns $N = 100$. In such a situation we select 100 patterns randomly from the 10,000 original data points. We perform the SMOTE generation experiments on these 100 selected points. Then we repeat with a different selection of the $N = 100$ data points M times, thus implementing the outer loop of the simulation experiment along the lines discussed above for the artificial data sets.

Table 1 shows the sizes and the dimensions of the considered datasets. Adult and Default datasets are UCI datasets [9] and the third dataset, credit card, is a Kaggle dataset developed by [5]. Table 2 shows the empirical estimates of the total variance difference (TVD) metric for varying dimensionality d, where N_f indicates the total number of features for every dataset as indicated in Table 1. It can be observed from Table 2 that as dimensionality increases, the distribution distance in terms of the TVD metric is enlarged, which supports the theoretical, and empirical results on artificial data presented in Fig. 3.

In addition, Table 3 demonstrates the empirical estimates of the total variance difference (TVD) metric for varying number of patterns N. It could be noted from Table 3 that for the three considered datasets, increasing number of minority class patterns generates samples closer to the original distribution, which agrees with the theoretical, and empirical results on artificial datasets shown in Fig. 4.

Finally, Table 4 presents the empirical estimates of the total variance difference (TVD) metric for varying the K parameter of KNN in SMOTE. It can be observed that increasing K results in increasing the TVD, which means that the generated patterns incur more divergence away from the original distribution. These results agree with the theoretical and empirical results represented in Fig. 5. A further discussion on the impact of the K parameter of the KNN used in the SMOTE method is provided in Sect. 4.3.

For Tables 2, 3 and 4, only empirically estimated TVD values have been computed. The theoretical estimates as defined in Eq. (4) are hard to compute because the underlying density function $p(X_0)$ is unknown and probability densities are very hard to estimate with a reasonable error, especially for high dimensions, even in case of large data sets.

Table 1. Real world datasets description

Dataset	Number of minority class patterns	Dimension
Adult	22,654	14
Credit card	284,315	28
Client default	23,364	23

Table 2. TVD for SMOTE versus dimensionality d for the real world datasets

Dataset	$d = 3$	$d = 5$	$d = 10$	$d = N_f$
Adult	−0.111	−0.172	−0.234	−0.252
Credit card	−0.107	−0.162	−0.226	−0.275
Client default	−0.106	−0.165	−0.178	−0.227

4.3 Commentary on the Results

From the presented results, we can observe that different variables affect the accuracy in similar directions, whether based on the theoretical or the experimental results. This validates and makes these findings more general. In summary, we observe the following:

Table 3. TVD for SMOTE versus number of patterns N for the real world datasets

Dataset	$N = 50$	$N = 100$	$N = 200$	$N = 500$
Adult	-0.098	-0.064	-0.031	-0.01
Credit card	-0.3192	-0.298	-0.284	-0.265
Client default	-0.197	-0.141	-0.0950	-0.0824

Table 4. TVD for SMOTE versus K parameter of KNN in SMOTE for the real world datasets

Dataset	$K = 1$	$K = 5$	$K = 10$	$K = 20$
Adult	-0.012	-0.023	-0.036	-0.068
Credit card	-0.253	-0.275	-0.286	-0.298
Client default	-0.066	-0.099	-0.118	-0.156

- We find the TVD always negative, indicating the contractive nature of SMOTE method.
- The faithfulness of SMOTE-sampling in emulating the true density deteriorates with higher dimension d. As mentioned, whether generating from a density or estimating parameters, handling higher dimension becomes more challenging.
- The accuracy improves as the number of minority examples N is higher, exhibiting a steep decline as N becomes very small. The reason is that for higher N the K-nearest neighbor patterns become closer to each other. This has us dealing with a region of similar density function value. Going too far means going to regions of markedly different density values, and hence less "representative" generated patterns.
- The faithfulness improves with smaller K (of the KNN), becoming the best at having a single neighbor $K = 1$. But, as we mentioned, a drawback of very small K, such as $K = 1$ is that the generated examples will generally be very close to the original examples, making them highly correlated with the original examples, and lessening their contribution in improving classification performance and other estimation tasks. As a general guide, selecting K in the range of 4 to 6 seems to be a sensible choice. This would be a trade-off to avoid the high errors of large K, and the correlation issue for very small K.

5   Conclusion

In this paper, we provide a theoretical and experimental analysis of the Synthetic Minority over-sampling TEchnique (SMOTE) method. SMOTE is an effective over-sampling method that generates extra examples from the minority class in order to combat class imbalance. In this work, we investigate the distribution of the SMOTE generated patterns and analyze how it deviates from the true distribution. In addition, we study how the different factors, such as: dimensions, the

number of minority patterns and the number of neighbors affect the divergence from the original distribution. We apply our experiments on both synthetic, and real datasets. The theoretical and the empirical results generally agree, and they should be a useful guide for using the SMOTE generation. As a disclaimer, this work considers only faithfulness in generating according to the true density. We do not consider how this affects classification, as this is out of scope of this work. However, an important first step in classification is to have accurate generation of patterns. A possible future work is to consider how this affects classification performance. Another possible direction to explore is to find methods or variants that would undo the contractive nature of SMOTE.

References

1. Batista, G., Prati, R., Monard, M.: A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explor. Newslett. **6**(1), 20–29 (2004)
2. Chan, P.K., Ave, L., York, N.: Distributed data mining in credit card fraud detection. IEEE Intell. Syst. Appl. **14**(6), 67–74 (1999)
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**(1), 321–357 (2002)
4. Chawla, N.V., Cieslak, D.A., Hall, L.O., Joshi, A.: Automatically countering imbalance and its empirical relationship to cost. Data Min. Knowl. Discov. **17**(2), 225–252 (2008)
5. Dal Pozzolo, A., Caelen, O., Johnson, R.A., Bontempi, G.: Calibrating probability with undersampling for unbalanced classification. In: 2015 IEEE Symposium Series on Computational Intelligence, pp. 159–166. IEEE (2015)
6. Dubey, R., Zhou, J., Wang, Y., Thompson, P.M., Ye, J., Alzheimer's Disease Neuroimaging Initiative: Analysis of sampling techniques for imbalanced data: an n = 648 ADNI study. NeuroImage **87**, 220–241 (2014)
7. Fan, W., Stolfo, S.J., Zhang, J., Chan, P.K.: AdaCost: misclassification cost-sensitive boosting. ICML **99**, 97–105 (1999)
8. Fayed, H., Atiya, A.F.: A novel template reduction approach for the-nearest neighbor method. IEEE Trans. Neural Netw. **20**(5), 890–896 (2009)
9. Frank, A., Asuncion, A.: UCI machine learning repository, vol. 213. School of Information and Computer Science, University of california, Irvine (2010). http://archive.ics.uci.edu/ml
10. Fukunaga, K., Hostetler, L.: Optimization of k nearest neighbor density estimates. IEEE Trans. Inf. Theory **19**(3), 320–326 (1973)
11. García, S., Herrera, F.: Evolutionary undersampling for classification with imbalanced datasets: proposals and taxonomy. Evol. Comput. **17**(3), 275–306 (2009)
12. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: Review of methods and applications. Expert Syst. Appl. **73**, 220–239 (2016)
13. Han, H., Wang, W.-Y., Mao, B.-H.: Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 878–887. Springer, Heidelberg (2005). https://doi.org/10.1007/11538059_91

14. He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: IEEE International Joint Conference on Computational Intelligence, IJCNN 2008, pp. 1322–1328. IEEE (2008)

15. He, H., Garcia, E.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)

16. Imam, T., Ting, K.M., Kamruzzaman, J.: z-SVM: an SVM for improved classification of imbalanced data. In: Sattar, A., Kang, B. (eds.) AI 2006. LNCS (LNAI), vol. 4304, pp. 264–273. Springer, Heidelberg (2006). https://doi.org/10.1007/11941439_30

17. Jian, C., Gao, J., Ao, Y.: A new sampling method for classifying imbalanced data based on support vector machine ensemble. Neurocomputing **193**, 115–122 (2016)

18. Liu, W., Chawla, N.V.: A robust decision tree algorithm for imbalanced data sets. In: SDM, vol. 10, pp. 766–777. SIAM (2010)

19. Longadge, R., Dongre, S.: Class imbalance problem in data mining review. arXiv preprint arXiv:1305.1707 (2013)

20. Luengo, J., Fernández, A., García, S., Herrera, F.: Addressing data complexity for imbalanced data sets: analysis of smote-based oversampling and evolutionary undersampling. Soft Comput. **15**(10), 1909–1936 (2011)

21. Sharma, S., Bellinger, C., Krawczyk, B., Zaiane, O., Japkowicz, N.: Synthetic oversampling with the majority class: a new perspective on handling extreme imbalance. In: 2018 IEEE International Conference on Data Mining (ICDM), pp. 447–456. IEEE (2018)

22. Yang, Q., Wu, X.: 10 challenging problems in data mining research. Int. J. Inf. Technol. Decis. Making **5**(4), 597–604 (2006)

23. Zhang, X., Li, Y.: A positive-biased nearest neighbour algorithm for imbalanced classification. In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds.) PAKDD 2013. LNCS (LNAI), vol. 7819, pp. 293–304. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37456-2_25