

Heteroscedastic-embedded Ensemble for Imbalanced Massive Data Classification

Hong Li

Yunnan University of Finance and Economics

Ge Peng

Baoshan University

Jia Wang

Yunnan University of Finance and Economics

Xingui Xu (✉ xu_xingui@126.com)

Yunnan University of Finance and Economics

Kun Qian

Yunnan University of Finance and Economics

Research Article

Keywords: Imbalanced learning, skewed dataset, heteroscedastic uncertainty, data classification

Posted Date: October 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2141598/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Heteroscedastic-embedded Ensemble for Imbalanced Massive Data Classification

Hong Li^a, Ge Peng^{b*}, Jia Wang^a, Xingui Xu^a and Kun Qian^a

^a*School of Information, Yunnan University of Finance and Economics, Kunming 650221, China*

^b*School of Bigdata, Baoshan University, Baoshan 678000, China*

ABSTRACT

The imbalanced learning methods aim to learn the unbiased models from massive class-imbalanced datasets. However, due to the uncertainty of data distributions affected by noise and borderline samples, the models based on heuristic assumptions or meta learning often suffer from the lack of stability and practicability for real-world classification tasks which often uses the skewed dataset of low-quality. In order to effectively cope with these issues, in this work, we propose a novel heteroscedastic-embedded ensemble (HEE) for imbalanced massive data classification. We first design an effective task sensing strategy with data-dependent heteroscedastic to adaptively guide the sampler's focus on informative samples, which makes the learning method much more robust for noisy data. Then, similar to the learning style of human beings, i.e., easy to difficult, the HEE framework gradually builds a strong ensemble classifier through interactions between harmonizing data resampling and model-training. The simulation results and analysis on both synthetic and real-world tasks demonstrate the effectiveness, robustness, and transferability of the proposed method.

Keywords—Imbalanced learning, skewed dataset, heteroscedastic uncertainty, data classification.

* Corresponding author. e-mail: pg_memf@126.com

1. Introduction

The highly imbalanced massive data classification has been a focus of research community as a result of extensive applications in many fields, such as customer churn forecast, fraud detection, and medical diagnosis [1-3]. However, due to the inherent skewness in class distributions which leads to canonical classification, the resulting model is biased for the majority classes. Most of the existing learning methods [4-7,17,26] perform poorly in terms of classification accuracy, stability, and computation efficiency. Meanwhile, the minority classes commonly yield higher interests, from both machine learning and practical perspectives [13]. Unfortunately, the real-world datasets are often flawed, due to the addition of noise, missing values, and inconsistent labeling. Thus, such highly imbalanced, large-scale, and defective datasets give rise to serious challenges in the classification tasks.

Generally, imbalanced data classification learning (IDCL) methods attempt to modify the underlying distribution by means of data resampling [6-8] or alleviate the bias towards majority classes by reweighting [9-12] in the learning process. Recently, the ensemble methods [14-18] are presented which combine one of the previous resampling or reweighting strategy for reducing the learning variance. These methods have achieved satisfactory performance. However, in practice, these algorithms persistently have few limitations, including: **(a)** Low accuracy and instability unstable performance. A major reason is that most resampling models blindly select data samples simultaneously, which leads to a large number of noise residuals. The reweighting cost sensitive learning is often designed in a heuristic way, however, these

assumptions are not always true. Another factor that contributes in the instability and inaccuracy of classification is the high sensitivity towards outliers; **(b)** Poor applicability. Many distance-based methods require predefined distance metric, which is usually unavailable in real-world datasets due to missing values. Meanwhile, the construction of cost matrix usually requires some domain expert knowledge, which is usually unavailable in practice. **(c)** High computing cost. Most of these methods need to compute the distance between instances' feature space, which is very time-consuming and even impossible for massive datasets.

In order to mitigate the aforementioned limitations, we propose a heteroscedastic embedded ensemble method to adaptively under-sample and train the classifier. The proposed method is inspired by two observations. The first is derived from the field of computer vision, where heteroscedastic quantifying uncertainty [21-22] is used to observe what a model does not know. This information is then used to make these learning methods more robust to noisy data. The second inspiration originates from the human-beings learning process of "easy to difficult", which is developed for optimal data-distribution weighting and classifier learning in our method. The proposed HEE framework is presented in Fig.1. Interestingly, we model the architecture similar to a closed-loop feedback control system, where the "feedback" part implements data-dependent task sensing and the other "control" part is online classifier learning. By learning the property of data distribution structure with heteroscedastic, the sample sensing function guides the sampler to always focus on the informative samples and make the learning method robust to noisy data. Additionally, the optimal

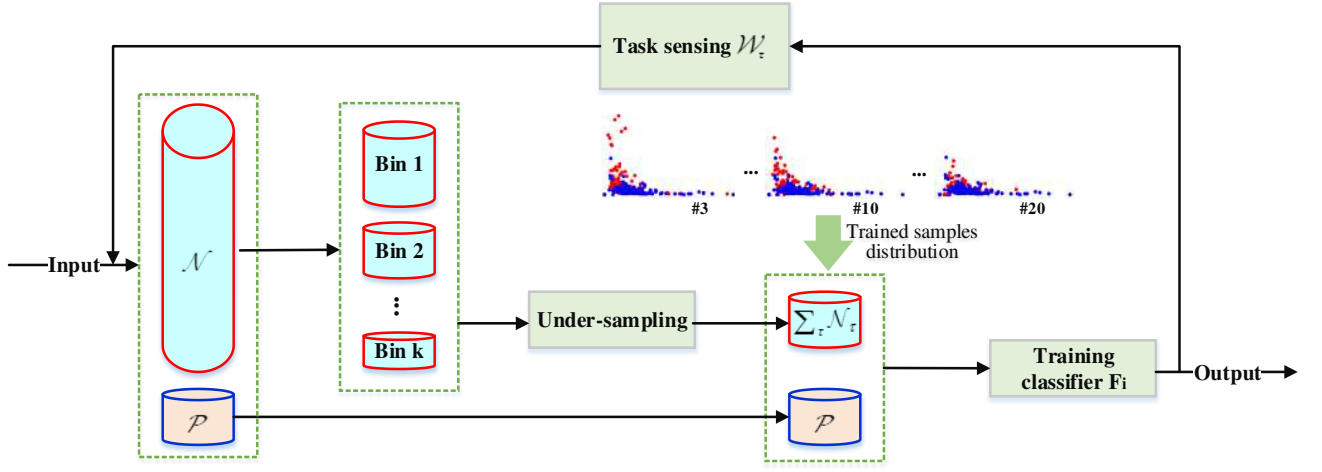


Fig. 1. The proposed heteroscedastic-embedded ensemble framework.

data-distribution weighting and classifier learning is simulated as human-beings learning process of “easy to difficult” (see scatter diagram of trained samples shown in Fig.1). We first learn the samples that can be classified easily. Then, we use the knowledge extracted from these samples to help the model learn complex measurement scale, such as massive region, borderline region, or outliers. In short, our work makes the following main contributions:

- (i) We propose *Classification Hardness*, a learning loss function that is determined by the classification loss combined with heteroscedastic.
- (ii) We propose *Task Sensing*, an adaptive ensemble strategy that includes samples selection and classifiers training.
- (iii) There is no need of any distance calculation and prior feature processing, the proposed method has low computing cost and robustness under missing values.

2. Preliminaries

Symbol definition: The multi-classification tasks can be transformed into binary classification problem. Here, we only consider two classes that are well-known in various practical applications [23-24]. For simplicity, we let the *minority* class with less samples to be positive class and the *majority* class with relatively more samples to be a negative class. $\mathcal{D}$ denoting all the training samples (x, y) , we define imbalance ratio (IR) as the number of majority class N divided by the number of minority class P . Thus, we have:

$$IR = n_{majority} / n_{minority} = |N| / |P| \quad (1)$$

$$s.t. P = \{(x, y) | y = 1\}, N = \{(x, y) | y = 0\}$$

Heteroscedastic uncertainty: The quantifying uncertainty [21-22] is widely used in machine learning methods to mitigate the learning bias. The uncertainty is roughly divided into two main types, namely epistemic uncertainty and aleatoric heteroscedastic uncertainty. The epistemic uncertainty accounts for the uncertainty in the model’s parameters, which is significantly eliminated given enough training data. This is often referred to as model uncertainty. Due to the small number of minority samples, it is impossible to improve the model’s ability in terms of epistemic uncertainty. Contrary, the aleatoric heteroscedastic uncertainty models the noise inherent in the observations, which can be significantly removed by modeling all the explanatory variables with increasing classification precision. Due to data-dependence property of heteroscedastic uncertainty, few inputs potentially can have more noisy outputs than others. Thus, for binary classification task of imbalanced datasets, the ensemble classifier model embedded with heteroscedastic uncertainty is much more

stable, has a better classification performance, and is more robust to noise. In heteroscedastic regression [25], if θ denotes the model parameters and $f(x)$ denotes the mapping function, then the model’s log likelihood function is:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma(x_i)^2} \|y_i - f(x_i)\|^2 + \frac{1}{2} \log \sigma(x_i)^2 \quad (2)$$

where, N represents the total number of data sample pairs, and σ_i^2 denotes the internal variance of the data samples. Due to the strong characterization of data instability with heteroscedastic model, it is useful in cases where some parts of the observation space have higher noise level than others. Especially, the first term of the model can be similarly changed as L_1 loss or cross entropy, which can be interpreted as the learning loss attenuation with data dependency. For classification tasks, the model predicts a vector which is squashed through a softmax function [19].

3. Proposed method

Inspired by the observation that heteroscedastic uncertainty in classification tasks can be interpreted as the learning loss attenuation with data dependency, we aim to design an adaptive under-sampling strategy in which the learning process focuses on informative samples. Therefore, we propose two novel concepts about classification hardness and task sensing. Moreover, we develop a heteroscedastic-embedded training model for achieving the goal.

3.1. Classification hardness and task sensing

Classification hardness: We use the symbol H to denote the classification hardness function, where H is determined by the classification loss combined with heteroscedastic. For a given sample instance x_i and labeled output y_i , the classification hardness is mathematically defined as:

$$H(x, y) := \frac{1}{D} \frac{1}{2\hat{\sigma}_i^2} \sum_{i=1}^N \|y_i - \hat{y}_i\|_{L_p} + \frac{1}{2} \log \sigma(x_i)^2 \quad (3)$$

where, $\hat{\sigma}_i^2$ denotes the predicted variance for classification results, implicitly from the loss function L_p , such as absolute error, cross entropy, squared error, etc. The first term in this function represents the classification loss obtained by the trained classifier output errors over the model parameters. This term is interpreted as the learning loss attenuation, and it can be any ‘decomposable’ classification loss function combined with inter-class variance. On the other hand, the second term of classification hardness serves as a regularization term, which prevents the classifier from predicting infinite uncertainty for the current training data.

Furthermore, we design a task sensing strategy to adaptively ensemble this classification hardness.

Task sensing: The concept of task sensing includes a dual sampling strategy about sample selection and training task. We use $F(x)$ to denote a trained classifier. Now, the task sensing function of sample (x, y) with respect to F is given by the function $S_s = (H(x, y); F)$. This has two advantages under the imbalance classification scenario. First, it adaptively senses the ensemble classification task between the data under-sampling and classifier's capacity. The heteroscedastic-embedded classifiers can be viewed as learning the relative sample weights of each bin (x, y) over the current classifier F_i . This allows us to learn the informative samples in each input batch of the dataset. Second, it senses the skewed multi-noise data distribution and classification task difficulty. The dataset's distribution structure is influenced by imbalance ratio or multi-noises and varies with the data itself. Thus, by embedding the data-dependent heteroscedastic into ensemble model, the HEE framework with task sensing can gradually build a strong ensemble classifier. Next, we describe the heteroscedastic-embedded training procedure in detail.

3.2. Heteroscedastic embedding

Similar to the harmonizing fashion used in batch training of neural networks [15,27], we split all the majority data samples into k bins according to their task sensing values. The t -th bin B_τ is defined as:

$$B_\tau = \{(x, y) \mid \frac{\tau-1}{k} \leq S_s(H; F) < \frac{\tau}{k}\} \quad (4)$$

Without losing generality, the sample sensing function is normalized as $S \in [0,1]$ and each bin indicates specific sample sensing level. Please note that the trivial samples vary in each iteration. In addition, the ensemble classifier gradually fits the sample training data. This indicates that their population will grow during the training process and the learning bias would not be balanced. Thus, there is a need to gradually slow down the learning

weights for less informative trivial samples. In the method proposed on this work, we introduce the logit factor to harmonize under-sampling, whereas the decreasing level is controlled by a logit growth form of iteration number. Using Eq. (4), the sampling weight of t -th bin is expressed as:

$$W_\tau = \frac{1}{\phi_\tau + \log(i^k)}, \forall \tau=1, \dots, k \quad (5)$$

where,

$$\phi_\tau = \sum_{s \in B_\tau} \frac{S_s(H; F_i)}{|B_\tau|}, \forall \tau=1, \dots, k \quad (6)$$

In Eq. (5) and Eq. (6), when the logit factor is fixed at i -th classifier, ϕ_τ gradually decreases with the classification loss attenuation decline, and the sampling weights become larger. The degree of classification difficulty and the informative samples in each bin are easily distinguished by ranking them using the proposed hardness function. Therefore, we gradually increase the weight of the sampling in the later iterations. This allows our method to focus on these informative samples and greatly improve the learning procedure. On the basis of forementioned dual sense of sampling and classification, the HHE method harmoniously keeps a reasonable fraction of samples and effectively prevents the proposed framework from overfitting.

In order to demonstrate the aforementioned strategy of classification hardness and task sensing, we used the real-world mammography dataset [31] to illustrate the realizing process. As presented in Fig. 2 (a, b, c), the task sensing strategy explicitly guides the classification tasks as loss attenuation, as well as implicitly induce the sampling process. Besides, as shown in Fig. 1 (scatter diagram of trained samples distribution), this "easy to difficult" sampling process designed by embedding the heteroscedastic into ensemble learning reduces the instability caused by the noise or outliers as well as signifies the importance of these informative borderline samples.

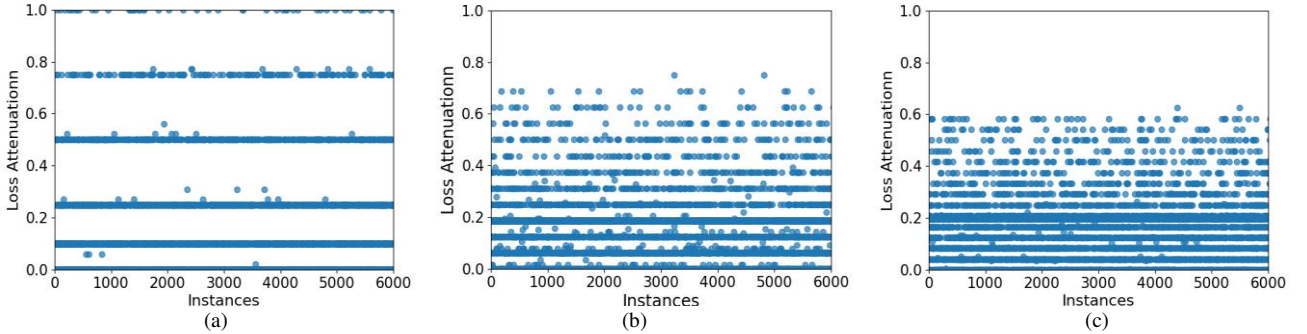


Fig. 2. The task sensing process. (a), (b), (c) denote the loss attention of classification task sensing during the training process corresponding to the ensemble classifier F_3 , F_{13} , and F_{20} , respectively.

3.3. Algorithm Formalization

According to the forementioned derivation about task sensing and heteroscedastic-embedded ensemble, the proposed algorithm formalization details are presented in Algorithm 1. Please note that by updating the classification hardness function in each iteration (line 4-5), we select the most beneficial data samples for the current ensemble classifier training. Then, the logit factor (line 7) harmonizes the under-sampling weights. The logit factor is zero in the first iteration and gradually begin to increase in the last iteration.

Algorithm 1: Heteroscedastic-embedded Ensemble

Input: Training data D ; classification hardness $H(x, y)$, base classifier f and its number n , bins k

Initialization: training f_0 with N_0' and P

- 1: **for** $i=1, \dots, n$ **do**
 - 2: Current ensemble classifier $F_i(x) = \sum_{j=1}^{i-1} f_j(x)$
 - 3: **for** $B_\tau, \tau=1, \dots, k$ **do**
 - 4: Solve W_τ by Eq.(5) and Eq.(6)
 - 5: Under-sampling with $N_\tau \leftarrow P \cdot W_\tau / \sum_m W_m$
 - 6: **end for**
 - 7: Train f_i using new subset $\{P, \sum_\tau N_\tau\}$
 - 8: **end for**
-

Output: ensemble classifiers $F(x) = \sum_{m=1}^n f_m(x)$

4. Experimental evaluation

4.1. Datasets and Training Set

Evaluation criteria: We use F1-score, G-mean, i.e., harmonic/geometric mean of precision and recall [22]-[23], and area under precision-recall curve (AUCPRC) [25] for evaluating the performance of the proposed binary classification scenario. The abbreviation in the criteria formula is: True Positive (TP) False Negative (FN) Negative False Positive (FP) True Negative (TN).

$$\begin{aligned} \text{-Recall} &= \frac{TP}{TP + FN} \\ \text{-Precision} &= \frac{TP}{TP + FP} \\ \text{-F1} &= 2 \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \\ \text{-G\_mean} &= \sqrt{\text{Recall} \cdot \text{Precision}} \end{aligned}$$

Dataset: The effectiveness of the proposed method is investigated using a series of imbalanced synthetic datasets as presented in [30]. We used two high levels of underlying class distribution overlapping with $IR=199$ and 299 in our experiments. Besides, we use five highly class-imbalanced real-world datasets, including credit fraud [28,29] which has only 492 frauds out of 284,807 transactions ($IR=578:1$), payment simulation with 8,213 frauds out of 6,362,620 transactions ($IR=773:1$), forest with 20,931 out of 5,749,132 record pairs ($IR=42:1$) available in the UCI repository [30,32] and mammography [31] with 260 out of 11186 data samples and $IR=42:1$.

Comparison methods: In this experiment, we compare HEE with 6 representative algorithms from the state-of-art ICDL methods presented in [33]. This includes UBagging [2], S-Boost [11], S-Bagging [12], SPE [13], MESA [15], and RUSBoost [16]. In order to ensure the fair comparison, all of these methods are deployed with decision trees [27] as the base classifier. Meanwhile, we use 60% of the total data for training, 20% of the total data for validation, and the remaining 20% is used for testing the classifier.

Parameters Setting: In our implementation, we use absolute error presented in Eq. (3) as the classification hardness loss. The setting of k -bins number is determined according to the number of samples in the dataset. In this work, we use k -bin equal to 10. As presented in Fig. 3, the accuracy of all methods when the number of base classifiers increases with the number of iterations. However, when the number of base classifiers is 20, the classification performance almost reaches the maximum. Thus, we choose 25 base classifiers for all the methods compared in this work. It is evident from the results that the proposed method

obtains consistent performance in terms of AUCPRC and is stable for different number of base classifiers.

4.2. Results analysis

Generalized performance comparison: Table 1 provides a comparison of results for 2 synthetic datasets and 5 real-world datasets. The results show that the proposed method outperforms other canonical approaches among the three evaluation indexes and also in terms of stability. Moreover, we observe that all the tested methods perform well on the less-overlapping dataset, i.e., 1st row. Due to the random under-sampling caused by some important information loss, the performance of resampling methods like UBagging and S-Boost is slightly degraded as compared to other methods. As overlapping intensifies, i.e., 2nd row, the boosting-based methods, i.e., S-Boost and RUSBoost, perform poorly in terms of classification because of the training process being biased by the noise samples. The ensemble strategy, such as the S-Bagging, SPE and MESA are less influenced by noise, however, their performance is still lower than HEE. Notably, the classification performance of all the methods is reduced for real-world datasets, but HEE remains stable and relatively gives best classification results. This is because the proposed method adaptively selects informative training subsets for improving the prediction performance, while being robust to noise/outliers. In such highly class-imbalanced and multi-noise datasets, the minority class is poorly represented and lacks a clear structure. Thus, the resampling and boost-based ensemble methods may increase the weight of noise. For instance, S-Boost, S-Bagging may deteriorate the classification performance. Similarly, by ignoring the data-dependent uncertainty, the adaptive ensemble techniques like SPE and MESA are usually unable to understand what a model does not making them less robust to real-world noisy data. Contrary, the proposed method has the ability to capture the inherent noise by using the heteroscedastic embedding, which senses the skewed multi-noise data distribution and classification task difficulty. Thus, the outliers and noise does not significantly affect the generalization ability of the proposed model. The comparison results show that the proposed HEE method is a simple but effective method, which has the ability to cope with overlapped, noisy, and highly class-imbalanced datasets. Moreover, the proposed method outperforms many state-of-the-art imbalanced learning approaches.

Robustness under Missing Values: In order to demonstrate the robustness and practicality of the proposed method for the imbalanced data classification task, we perform simulations on the dataset with missing values. Similar to method presented in [13] for coping the missing values, we randomly select values from all the features from the original datasets. Then, we replace them with zeros. We performed a series of seven experiments on the payment simulation dataset, when 0% / 20% / 40% / 60% values are missing. As shown in Table 2, the proposed method demonstrates robust performance for different levels of missing values, while other methods perform poorly, when the number of missing records are too high.

Sensitivity to Hyper-parameters: We implement parameter sensitivity training on synthetic dataset with $IR=199$ and real-world mammography dataset. As shown in Fig.4, when the number of bins is 5, the performance of the proposed method using cross entropy (CE), squared error (SE), and absolute error (AE) in Eq.(3) becomes stable. It is noteworthy that the classification hardness function using AE has the best performance. This is because the loss attenuation appropriately reflects the matching degree of the classifier results over data bins distribution. Meanwhile, k determines how detailed the task sense is, thus setting a small k , e.g., $k < 4$, may lead to poor performance.

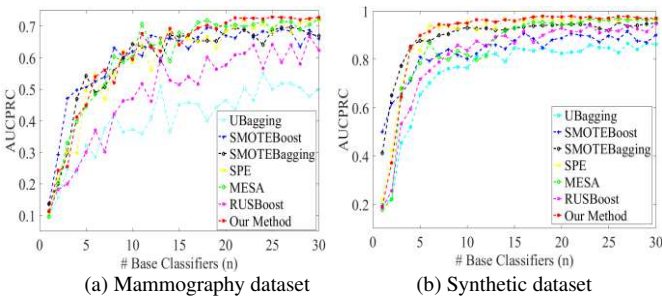


Fig. 3. The performance of 7 methods on two datasets with the number of base classifiers (n).

Table 1

Performance of 7 methods on synthetic and real-word datasets

Dataset	Metric	UBagging	S-Boost	S-Bagging	SPE	MESA	RUSBoost	Ours
Synthetic IR=199:1	AUCPRC	0.825±0.012	0.901±0.011	0.950±0.017	0.973±0.002	0.980±0.007	0.947±0.014	0.979±0.001
	F1	0.851±0.011	0.862±0.011	0.916±0.013	0.930±0.002	0.930±0.005	0.897±0.015	0.932±0.001
	GM	0.851±0.012	0.862±0.011	0.914±0.014	0.930±0.002	0.940±0.009	0.894±0.015	0.933±0.001
Synthetic IR=299:1	AUCPRC	0.814±0.023	0.860±0.014	0.932±0.015	0.957±0.009	0.929±0.012	0.929±0.020	0.965±0.003
	F1	0.839±0.022	0.857±0.013	0.909±0.015	0.918±0.007	0.903±0.013	0.890±0.021	0.927±0.004
	GM	0.840±0.022	0.858±0.014	0.909±0.015	0.919±0.007	0.904±0.013	0.891±0.021	0.927±0.004
Credit fraud IR=578:1	AUCPRC	0.302±0.018	0.801±0.012	0.741±0.013	0.790±0.020	0.864±0.008	0.649±0.022	0.873±0.008
	F1	0.305±0.014	0.826±0.010	0.720±0.012	0.845±0.018	0.850±0.012	0.771±0.018	0.857±0.012
	GM	0.351±0.015	0.826±0.010	0.726±0.009	0.849±0.008	0.853±0.014	0.771±0.019	0.858±0.014
Mammo IR=42:1	AUCPRC	0.477±0.020	0.619±0.015	0.584±0.021	0.685±0.011	0.706±0.011	0.600±0.023	0.727±0.010
	F1	0.562±0.017	0.623±0.013	0.542±0.022	0.705±0.010	0.686±0.010	0.631±0.024	0.685±0.011
	GM	0.568±0.016	0.624±0.013	0.548±0.022	0.702±0.010	0.691±0.010	0.633±0.025	0.700±0.010
Protein IR=111:1	AUCPRC	0.602±0.013	0.729±0.018	0.713±0.012	0.818±0.012	0.807±0.018	0.682±0.023	0.820±0.007
	F1	0.691±0.014	0.713±0.015	0.724±0.010	0.814±0.010	0.802±0.015	0.702±0.020	0.817±0.007
	GM	0.687±0.015	0.715±0.014	0.718±0.011	0.813±0.011	0.803±0.015	0.706±0.018	0.814±0.005
Forest IR=211:1	AUCPRC	0.643±0.022	0.873±0.012	0.899±0.007	0.912±0.009	0.901±0.013	0.723±0.021	0.927±0.004
	F1	0.744±0.016	0.834±0.010	0.856±0.009	0.853±0.010	0.848±0.011	0.769±0.022	0.869±0.003
	GM	0.747±0.016	0.835±0.011	0.856±0.009	0.853±0.010	0.855±0.011	0.763±0.022	0.869±0.003
Payment IR=773:1	AUCPRC	0.673±0.015	0.888±0.020	0.913±0.013	0.938±0.003	0.945±0.001	0.801±0.018	0.959±0.003
	F1	0.701±0.013	0.890±0.014	0.890±0.009	0.892±0.004	0.886±0.001	0.790±0.017	0.901±0.002
	GM	0.670±0.017	0.841±0.015	0.871±0.009	0.889±0.003	0.886±0.001	0.811±0.014	0.903±0.001

Table 2

Robustness performance of 7 methods on missing values

Model	UBagging	S-Boost	S-Bagging	SPE	MESA	RUSBoost	Ours
0%	0.379±0.041	0.761±0.013	0.783±0.009	0.786±0.011	0.786±0.017	0.493±0.071	0.807±0.003
20%	0.301±0.033	0.665±0.017	0.691±0.021	0.705±0.013	0.705±0.030	0.310±0.052	0.748±0.010
40%	0.192±0.012	0.401±0.025	0.434±0.018	0.586±0.022	0.586±0.051	0.287±0.021	0.602±0.019
60%	0.110±0.007	0.272±0.029	0.192±0.029	0.390±0.039	0.390±0.075	0.094±0.013	0.445±0.033

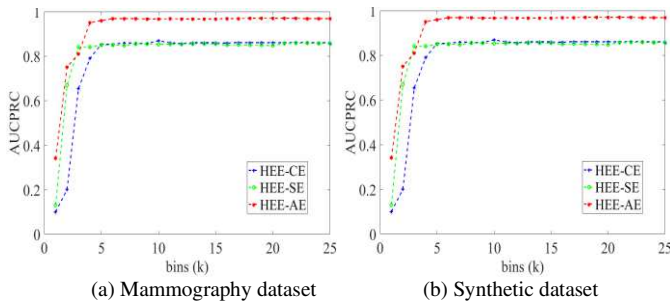


Fig.4 Sensitivity to hyper-parameters of our method on two datasets with the number of bins (k).

Therefore, the results in Fig. 4 also indicate that the proposed method is robust to different selection of k .

5. Conclusion

To overcome the difficulties of imbalance classification arriving from highly imbalanced, large-scale, overlapping between classes and defective datasets, we proposed the task sensing to integrate both the classification hardness and data-dependent heteroscedastic into a novel heteroscedastic-embedded ensemble (HEE) framework. This strategy gradually builds a strong ensemble classifier by harmonizing the data resampling and model-training. As compared with the other methods, extensive simulations on both synthetic and real-world, low-quality, and skewed dataset classification tasks demonstrate the effectiveness, robustness, and stability of the proposed method.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported primarily by National Natural Science Foundation of China (Grant No. 62161051). Yunnan University of Finance and Economics Talent Project (No. 2021D17, 2022D18). We particularly thank UCI and Kaggle for the Dataset and the constructive suggestions of reviewers.

References

- [1] G. O. Young, "Synthetic structure of industrial plastics," in *Plastics*, 2nd ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, pp. 15–64, 1964.
- [2] C. Guo, A. B. Haixiang, et al., "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220-239, 2017.
- [3] J. H. Xue, Z. Ma, M. Roveri and N. Japkowicz, "Guest editorial special issue on recent advances in theory, methodology, and applications of imbalanced learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2688-2690, 2020.
- [4] P. Skryjowski, and B. Krawczyk, "Influence of minority class instance types on SMOTE imbalanced data oversampling," *Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications*, 2017.
- [5] H. Junjie, Y. Haiqin, R. Michael, et al., "Online Nonlinear AUC Maximization for Imbalanced Data Sets," *IEEE Transactions on Neural Networks & Learning Systems*, 2017.
- [6] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning," *Proceedings of the 2005 international conference on Advances in Intelligent Computing*, Volume Part I, 2005.

- [7] T. Liu, X. Zhu, W. Pedrycz, *et al.*, "A design of information granule-based under-sampling method in imbalanced data classification," *Soft Computing*, vol.24, no.1, 2020.
- [8] F. Ren, P. Cao, W. Li, *et al.*, "Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm," *Computerized Medical Imaging & Graphics*, pp. 55:54, 2017.
- [9] Y. Sun, M. S. Kamel, A. K. C. Wong, *et al.*, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol.40, no.12, pp.3358-3378, 2007.
- [10] S. H. Khan, M. Hayat, M. Bennamoun, *et al.*, "Cost Sensitive Learning of Deep Feature Representations from Imbalanced Data," *IEEE Transactions on Neural Networks & Learning Systems*, vol.29, no.8, pp:3573-3587, 2018.
- [11] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "Smoteboost: Improving prediction of the minority class in boosting," *European conference on principles of data mining and knowledge discovery*. Springer, pp. 107-119, 2003.
- [12] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," 2009 IEEE Symposium on Computational Intelligence and Data Mining, pp. 324-331, 2009.
- [13] Z. Liu, W. Cao, Z. Gao, *et al.*, "Self-paced Ensemble for Highly Imbalanced Massive Data Classification," *2020 IEEE 36th International Conference on Data Engineering*, pp. 841-852, 2020.
- [14] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Progress in Artificial Intelligence*, 5.4, 2016.
- [15] Z. Liu, W. Cao, Z. Gao, *et al.*, "MESA: Boost Ensemble Imbalanced Learning with MEta-Sampler," *Conference on Neural Information Processing Systems*, 2020. [Online]. Available: <https://arxiv.org/abs/2010.08830>
- [16] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, *et.al.*, "Rusboost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 185-197, 2010.
- [17] M. Duan, D. Liu, X. Chen, *et al.* "Self-balancing Federated Learning with Global Imbalanced Data in Mobile Systems," *IEEE Transactions on Parallel and Distributed Systems*, pp.99:1-1, 2020.
- [18] T. Takata, K. Shiraishi, S. Kumagai, *et al.*, "Calculating and estimating second cancer risk from breast radiotherapy using monte carlo code with internal body scatter for each out-of-field organ," *Journal of Applied Clinical Medical Physics*, 2020
- [19] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539-550, 2009.
- [20] R. Barandela, R. M. Valdovinos, and J. S. Sánchez, "New applications of ensembles of classifiers," *Pattern Analysis & Applications*, vol. 6, no. 3, pp. 245-256, 2003.
- [21] C. Englert, P. Galler, P. C. Harris, *et al.*, "Machine learning uncertainties with adversarial neural networks," *European Physical Journal C*, vol.79, no.1, 2019.
- [22] A. Kendall, Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision," *Advances in neural information processing systems*, pp: 5574-5584, 2017.
- [23] Kendall, Y. Gal, R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp: 7482-7491, 2018.
- [24] Dal Pozzolo, G. Boracchi, O. Caelen, *et al.*, "Credit card fraud detection: a realistic modeling and a novel learning strategy," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3784-3797, 2018.
- [25] H. He and Y. Ma, "Imbalanced learning: foundations, algorithms, and applications," *John Wiley & Sons*, 2013.
- [26] Z. Nenadic, "Information Discriminant Analysis: Feature Extraction with an Information-Theoretic Objective," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 29, no. 8, pp. 1394, 2018.
- [27] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81-106, 1986.
- [28] Li, Y. Liu, and X. Wang, "Gradient harmonized single-stage detector," *arXiv preprint arXiv:1811.05181*, 2018.
- [29] D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, *et.al.*, "Credit card fraud detection: a realistic modeling and a novel learning strategy," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3784-3797, 2018.
- [30] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp.1-5, 2017.
- [31] N. V. Chawla, K.W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002
- [32] Dheeru Dua and G. Casey, "UCI machine learning repository," 2017.
- [33] F. Alberto, G. Salvador, G. Mikel, P. Ronaldo C., and K. Bartosz, "Learning from Imbalanced Data Sets", Springer, 2018.