An Effective Method for Imbalanced Time Series Classification: Hybrid Sampling

Conference Paper · December 2013

DOI: 10.1007/978-3-319-03680-9_38

1 author:

Guohua Liang
University of Technology Sydney

An Effective Method for Imbalanced Time Series Classification: Hybrid Sampling

Guohua Liang

The Centre for Quantum Computation and Intelligent Systems,
Faculty of Engineering and IT, University of Technology, Sydney NSW 2007 Australia
Guohua.Liang@student.uts.edu.au
http://www.qcis.uts.edu.au

Abstract. Most traditional supervised classification learning algorithms are ineffective for highly imbalanced time series classification, which has received considerably less attention than imbalanced data problems in data mining and machine learning research. Bagging is one of the most effective ensemble learning methods, yet it has drawbacks on highly imbalanced data. Sampling methods are considered to be effective to tackle highly imbalanced data problem, but both over-sampling and under-sampling have disadvantages; thus it is unclear which sampling schema will improve the performance of bagging predictor for solving highly imbalanced time series classification problems. This paper has addressed the limitations of existing techniques of the over-sampling and under-sampling, and proposes a new approach, hybrid sampling technique to enhance bagging, for solving these challenging problems. Comparing this new approach with previous approaches, over-sampling, SPO and under-sampling with various learning algorithms on benchmark data-sets, the experimental results demonstrate that this proposed new approach is able to dramatically improve on the performance of previous approaches. Statistical tests, Friedman test and Post-hoc Nemenyi test are used to draw valid conclusions.

Keywords: Hybrid sampling, over-sampling, under-sampling, imbalanced data, time series data, ensemble learning, and classification.

1 Introduction

In data mining research, mining time series data is one of the most challenging problems [1], and the imbalanced data problem is a fundamental classification problem [2]. Most traditional supervised classification learning algorithms are ineffective for highly imbalanced time series classification (HITSC) [3]. Due to its challenging issues of high dimensionality, large scale, and uneven class distribution among different classes, and considering the sequence of the numerical attributes carrying special information as whole instead of individual attributes [4, 5], it has received considerably less attention than imbalanced data problems in data mining and machine learning research. HITSC refers to a situation in which the proportions of the training examples of time series

S. Cranefield and A. Nayak (Eds.): AI 2013, LNCS 8272, pp. 374–385, 2013.

data are varied significantly among different classes. This study mainly focuses on imbalanced binary time series classification (TSC), e.g., the proportion of positive examples that are far fewer than the proportion of negative examples in the training data of the TSC.

Bagging [6] was introduced by Breiman in 1996. Previous research shows that bagging can improve the performance of individual classifiers if base learners are unstable [6–9], but it has a limitation for solving highly imbalanced data problems. Sampling techniques are considered to be one of the most effective ways to tackle highly imbalanced problems, but since both over-sampling and under-sampling techniques have their limitations, it is unclear which sampling schema is able to enhance the performance of bagging. These challenging issues have motivated me to propose a new approach, hybrid-sampling (H-Sampling) techniques, to enhance bagging, for solving HITSC problems.

The proposed new H-sampling approach randomly over-samples the positives and under-samples the negatives to half of the original training size, $\frac{|P|+|N|}{2}$, respectively, to generate a set of balanced bootstrap samples from the original training set. This set of balanced bootstrap samples is used to train a set of classifiers; then each test example is predicted by a set of trained classifiers; lastly, the final prediction of each test example is made by the majority votes of these predictions of the set of trained classifiers. Comparing the performance of this new approach with previous approaches [10, 3, 5],the over-sampling method SPO and under-sampling method with various algorithms on the benchmark data-sets, the experimental results demonstrate that the proposed new approach, H-sampling to enhance bagging, is superior to previous approaches [10, 3, 5], and dramatically improves the performance of previous approaches. Statistical tests, Friedman and post-hoc Nemenyi tests for comparing the performance of multiple learning methods over multiple benchmark data-sets are applied to draw valid conclusions.

The key contributions of this paper are as follows. (1) This paper addresses the limitations of the existing over-sampling and under-sampling techniques, and proposes a new approach, H-sampling technique to enhance bagging, for improving the performance of prediction models to solve the HITSC problems. (2) Empirically comparing the performance of this new approach with previous approaches on the benchmark data-sets, the experimental results demonstrate that the new approach, H-Sampling integrating the unstable base learner, decision trees $J48$ with bagging, is effective for solving the HITSC problems and is dramatically superior to previous approaches: the over-sampling method, SPO and the under-sampling method with KNN.

The paper is organized as follows. Section 2 presents an outline of the proposed new approach. Section 3 shows related work. Section 4 presents the evaluation measures. Sections 5 and 6 provide the experimental setting and experimental analysis. Section 7 concludes this work.

Algorithm 1. H-Sampling Bagging

Input:
D, original training set, containing $|P|$ positive
and $|N|$ negative instances;
a learning scheme (algorithm, e.g., $J48$);

Output: A composite model, C^*.

Method:

for $i = 1$ **to** k **do**
Create balanced bootstrap samples of
size $|D_i|$ sub-sets, $|D_i| = |P_i| + |N_i|$ where
P_i and N_i are randomly drawn with replacement
from original training set, P and N, respectively:
$|P_i| = |N_i| = \frac{(|P|+|N|)}{2}$ and;
end
return *a set of bootstrap samples* D_i *(containing* k *bootstrap samples);*

Train each base classifier model C_i from D_i;

To use the composite model, C^* for a test set T on an instance x where its true class label is y:
$$C^*(x) = \arg\max_y \sum_i \delta\left(C_i(x) = y\right)$$
Delta function $\delta(\mathbf{.}) = 1$ if argument is true, else 0.

2 Hybrid Sampling Approach

Algorithm 1 outlines the proposed new approach, H-sampling integrating unstable learner decision trees $J48$ with bagging. This new approach is different from previous approaches [10, 3, 5] because H-sampling reduces the disadvantage of under-sampling, loosing to much important information for training, and the disadvantages of over-sampling, over-fitting, high computational cost and longer training time. This new approach, H-sampling, randomly selects the positives and the negatives to the balanced point at half of the original training size, $\frac{|P|+|N|}{2}$. For example, the positives are randomly selected with replacement from the entire positive class to the size of the balanced point; the negatives are randomly selected with replacement from the negative class of original training set to the size of the balanced point.

For the proposed prediction model, suppose the size of an ensemble is k, a set of classifiers C_i (for i=1 to k) is built from a set of balanced bootstrap samples D_i; each new test example is classified by a set of classifiers C_i, and the final prediction is made by majority votes to aggregate the predictions of the set of classifiers C_i by using a delta function $\delta(\mathbf{.}) = 1$ if the prediction of C_i is a true class label, else the delta function $\delta(\mathbf{.}) = 0$.

Majority votes, aggregating the set of predicted class labels, use the delta function to vote for a class and the class label obtaining the highest number of votes is considered as the output of the final prediction.

3 Related Work

This paper proposes a new approach, H-sampling integrating unstable learner decision trees $J48$ with bagging for solving HITSC problems. This new approach is different to previous approaches [10, 3, 5] because both over-sampling and under-sampling techniques have disadvantages. This new approach not only reduces the limitations of over-sampling and under-sampling techniques, but also enhances bagging to effectively improve the performance of the previous approaches for solving HITSC problems.

The main disadvantages of over-sampling are that over-sampling dramatically increases the computational cost of training and training time, and may cause over-fitting, even though it maintains the important information for training, because additional large number of new positive examples with high dimensional features are generated to balance the training set for HITSC [3]. The main disadvantages of under-sampling may lose important and useful information for training and may degrade the performance of the prediction models, even though it significantly reduces the computational cost of training, because only a proportion of the majority class examples are selected to train prediction models.

In earlier research, a structure-preserving over-sampling (SPO) [10] method with support vector machines (SVM) was proposed for solving HITSC problems; it achieves better results than other over-sampling methods and state-of-the-art methods in TSC, based on a comparison of the average values of two evaluation measures, F_{value} and Geometric mean (G_{mean}), without statistical analysis to support this conclusion. The study compared SPO with over-sampling methods, which include repeating (REP), SMOTE [11] (SMO), Borderline_SMOTE [12] (BoS), ADASYN [13] (ADA), and DataBoost [14] (DB); and with state-of-the-art methods in TSC, which include Easy Ensemble [15] (Easy), BalanceCascade [15] (Bal), One nearest neighbor classifier using Euclidean distance [16] (1NN), and One nearest neighbor classifier using dynamic time warping distance [17] (1NN_DW).

Our other work [3] proposed an under-sampling technique integrated with SVM, which is more efficient than other more complicated approaches, such as SPO with SVM for HITSC. However, it is unclear whether the under-sampling method with various supervised learning algorithms is more effective than the over-sampling method, SPO, and the under-sampling technique integrated with SVM for HITSC.

Our previous work [5] conducted an empirical evaluation of the performance of over-sampling methods (e.g., the complex SPO [10]) and under-sampling with various supervised learning algorithms selected from Weka [18], such as Sequential Minimal Optimization (SMO) of SVM, decision trees ($J48$), Random Tree (RTree), K Nearest Neighbor (KNN) with default parameter setting K=1, and Multi-layer Proceptron (MLP). The experimental results indicate that the under-sampling technique with KNN achieves better results than the existing complicated SPO method for ITSC.

3.1 Statistical Tests

Friedman and post-hoc Nemenyi tests are applied to compare the performance of the multiple learning methods on multiple data-sets, where it is inappropriate to compare their average value, because the average values are susceptible to outliers [19, 5]. Therefore, average rank is preferred for evaluating the performance of multiple learning methods. This work therefore performs statistical tests to evaluate the performance of the multiple learning methods on multiple data-sets. The Friedman test is utilized to obtain the average rank of the performance of the multiple learning methods on multiple data-sets; the post-hoc Nemenyi test is utilized to check whether there is a statistically significant difference between the learning methods at a 95% confidence interval.

4 Evaluation Metrics

The estimated overall accuracy is an ineffective evaluation measure for the imbalanced classification task [5, 20–22], so two evaluation measures are used for this study: F_{value} and G_{mean}.

Table 1 presents a confusion matrix for a binary classification problem; the columns represent the predicted class, and the rows represent the actual class. The evaluation measures are derived from the confusion matrix as follows:

Table 1. Confusion matrix for a binary classification problem

	Predicted Positives	Predicted Negatives
Actual Positives (P)	True Positive (TP)	False Negative (FN)
Actual Negatives (N)	False Positive(FP)	True Negative (TN)

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

$$TNR = \frac{TN}{TN + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$F_{value} = \frac{2recall * precision}{recall + precision} \tag{5}$$

$$G_{mean} = \sqrt{TPR * TNR} \tag{6}$$

5 Experimental Setup

Java platform is used to implement the new approach, H-sampling technique integrated unstable learner, decision trees J48 [18] with bagging, and to investigate the performance of the new approach and previous approaches. 31 bootstrap samples are used in the ensemble. A 10-trial 10-fold cross-validation evaluation is performed for this study. The Friedman test is used for the calculation of average rank.

Table 2. Time series data-sets [5]

Data-sets		Data Information					Class Information	
Index	Name	TS Length	Instances (P^+&N^-)	P^+	N^-	Ratio P^+/N^-	Previous Class	Altered class
1	Adiac	176	781	23	758	0.0303	37	2
2	S-Leaf	128	1125	75	1050	0.0714	15	2
3	Wafer	152	7164	762	6402	0.0119	2	2
4	FaceAll	131	2250	112	2138	0.0524	14	2
5	Yoga	426	3300	1530	1770	0.8644	2	2

5.1 Data-Sets

Table 2 shows a summary of the characteristics of the five time series data-sets from the public UCR time series repository [23], which were used as the benchmark data-sets of previous work [10, 3, 5].

6 Experimental Results Analysis

This section contains two sub-sections: 6.1 comparison of the performance of over-sampling, under-sampling with various algorithms and H-sampling methods on HITSC; and 6.2 comparison of the performance of other learning methods, SPO, under-sampling with various algorithms, and H-sampling methods for HITSC.

6.1 Comparison of the Performance of Over-sampling, Under-sampling, and Hybrid-sampling Methods

Table 3 presents a comparison of the performance of this new approach, H-sampling to enhance bagging, with previous approaches, over-sampling methods and the under-sampling with various algorithms based on the F_{value} and G_{mean} measures. The experimental results indicate that this new approach, H-sampling to enhance bagging, achieves the best performance with F_{value} across all over-sampling methods and the under-sampling with various algorithms on average value and average rank of F_{value}. This new approach achieves the highest average value 0.962 with smallest standard deviation (STD) 0.031 and the best average rank 1.4, respectively, which are the best results across all methods; while KNN

Table 3. Comparison of the performance of over-sampling, under-sampling methods with different learning algorithms, and H-sampling to enhance bagging based on the evaluation metrics F_{value} and G_{mean}

Metrics	Data-set	Results from Previous Research [10]						Results from previous Work [5]					This Work
		Over-sampling Methods						Under-sampling					H-sampling
	Name	REP	SMO	BoS	ADA	DB	SPO	SVM	J48	RTree	KNN	MLP	H-Bagging
F_{value}	Adiac	0.375	0.783	0.783	0.783	0.136	0.963	0.967	0.883	0.903	0.918	0.947	**0.975**
	S-Leaf	0.761	0.764	0.764	0.759	0.796	0.796	0.841	0.820	0.849	0.836	0.786	**0.932**
	Wafer	0.962	0.968	0.968	0.967	0.977	0.982	0.891	0.929	0.956	**0.999**	0.933	0.980
	FaceAll	0.935	0.935	0.935	0.935	0.890	0.936	0.957	0.876	0.863	0.909	0.919	**0.995**
	Yoga	0.710	0.729	0.721	0.727	0.689	0.702	0.744	0.771	0.811	0.807	0.780	**0.926**
	AverageValue	0.740	0.836	0.834	0.834	0.698	0.876	0.880	0.856	0.876	0.894	0.873	**0.962**
	STD	0.236	0.108	0.110	0.109	0.332	0.122	0.110	0.061	0.055	0.075	0.083	**0.031**
	AverageRank	8.90	6.90	7.30	7.70	8.70	4.50	7.40	7.80	6.40	4.40	6.60	**1.40**
	CD	7.45											
G_{mean}	Adiac	0.480	0.831	0.831	0.831	0.748	**0.999**	0.957	0.910	0.920	0.958	0.975	0.989
	S-Leaf	0.800	0.861	0.861	0.849	0.898	0.898	0.902	0.809	0.812	0.887	0.856	**0.976**
	Wafer	0.965	0.969	0.970	0.970	0.980	0.984	0.903	0.907	0.956	**0.998**	0.937	0.988
	FaceAll	0.950	0.950	0.950	0.950	0.948	0.957	0.966	0.870	0.860	0.929	0.925	**0.997**
	Yoga	0.741	0.756	0.750	0.755	0.724	0.735	0.630	0.807	0.803	0.808	0.774	**0.976**
	AverageValue	0.787	**0.783**	0.872	0.871	**0.860**	**0.915**	0.872	0.861	0.870	0.916	0.893	**0.985**
	STD	0.197	0.088	0.090	0.089	0.117	0.108	0.138	0.051	0.067	0.073	0.079	**0.009**
	AverageRank	9.30	6.80	6.90	7.20	7.50	4.10	6.60	8.60	8.20	4.20	7.20	**1.40**
	CD	7.45											

with the under-sampling method achieves the average value 0.894 with STD 0.075 and average rank 4.40, respectively, which is the second best across all methods on F_{value}.

On average value and average rank of the G_{mean} measure, this new approach, H-sampling to enhance bagging achieves the highest average value 0.985 with smallest STD 0.009 and lowest average rank 1.40, respectively, which is the best across all the compared methods; while, the SPO over-sampling method achieves average value 0.915 with STD 0.108 and average rank 4.1, respectively, which is the second best across all the compared methods on average rank of the G_{mean} measure, whereas KNN with the under-sampling method achieves average value 0.916 with STD 0.073 and average rank 3.4, respectively, which is the second best across all the compared methods on average of the G_{mean} measure. The results highlighted in red indicate the correction of the previous work [10, 5].

Figs 1 and 2 present a comparison of this new approach, H-sampling to enhance bagging, with previous approaches, over-sampling and under-sampling with various algorithms, with the Nemenyi test, where the x-axis indicates the ranking order of the sampling methods; the y-axis indicates the average rank of the F_{value} and G_{mean} performance, respectively, and the vertical bars indicate the "Critical Difference". Groups of sampling methods that are no significantly different at a 95% confidence interval are indicated when the vertical bars overlap. Comparing the performance of this new approach with previous approaches, over-sampling [10] and under-sampling with various algorithms [24], based on F_{value} and G_{mean}, H-sampling with bagging has the best average rank on both measures. KNN with the under-sampling method has the second best average rank of F_{value}; while the SPO over-sampling method has the second best average rank of G_{mean}. Statistical tests indicate that there is no statistically significant difference at a 95% confidence interval between over-sampling SPO, under-sampling KNN, and H-sampling with Bagging on the average rank of F_{vlaue} and G_{mean}; however, there is statistically significant difference at a 95% confidence interval between H-sampling and two over-sampling methods, DB and REP on F_{vlaue} measure, and between H-sampling and two over-sampling methods,J48 and REP on G_{mean} measure.

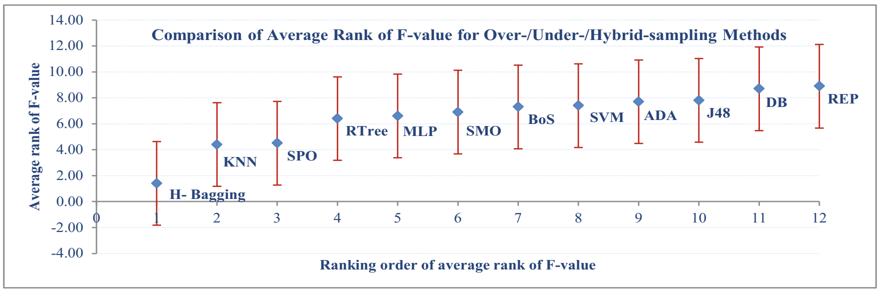

Fig. 1. Comparison of average rank of the F_{value} with the Nemenyi test for the over-sampling, SPO, under-sampling, and H-sampling methods, where the x-axis indicates the ranking order of all the sampling methods with learning algorithms, the y-axis indicates the average rank of the F_{value}, and the vertical bars indicate the "Critical Difference"

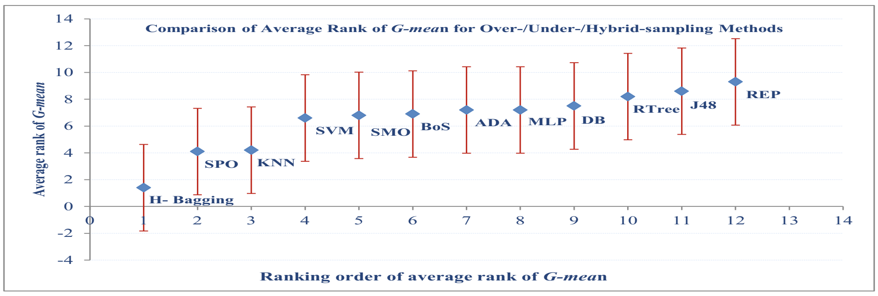

Fig. 2. Comparison of average rank of the G_{mean} with the Nemenyi test for all the over-sampling, under-sampling, and H-Bagging methods, where the x-axis indicates the ranking order of all the sampling methods with learning algorithms, the y-axis indicates the average rank of the G_{mean}, and the vertical bars indicate the "Critical Difference"

6.2 Comparison of the Performance Learning Methods, Over-sampling SPO, Under-sampling, and H-sampling Methods

Table 4 presents a comparison of the performance of previous work (learning methods [10] and the under-sampling with various algorithms [5]), and this work H-sampling with bagging based on F_{value} and G_{mean} evaluation measures. The experimental results indicate that H-sampling with bagging achieves the best performance on F_{value} and G_{mean} across all previous approaches, and H-sampling methods on average value of 0.962 and 0.985, and average rank of 1.40 and 1.40, respectively, which is the best average value and average rank of F_{value} and G_{mean} across all previous learning methods [10] and under-sampling method [5]; KNN achieves an average value of 0.894 and 0.916, and an average rank of 3.0 and 2.4, respectively, which is the second best average value and average rank of F_{value} and G_{mean} across all the remaining methods.The results highlighted in red indicate the correction of the previous work [10].

Table 4. Comparison of the performance of learning methods from previous research [10] and learning algorithms with under-sampling [5], and H-sampling from this work based on evaluation metrics: F_{value} and G_{mean}

Metrics	Data-set	Results from Previous Research [10]					Results from Previous Work [5]					This Work
		Learning Methods					Under-sampling					H-sampling
	Name	Easy	Bal.	1NN	1NN_DW	SPO	SVM	J48	RTree	KNN	MLP	H-Bagging
F_{value}	Adiac	0.534	0.348	0.800	0.917	0.963	0.967	0.883	0.903	0.918	0.947	**0.975**
	S-Leaf	0.521	0.578	0.716	0.429	0.796	0.841	0.820	0.849	0.836	0.786	**0.932**
	Wafer	0.795	0.954	0.949	0.857	0.982	0.891	0.929	0.956	**0.999**	0.933	0.980
	FaceAll	0.741	0.625	0.802	0.959	0.936	0.957	0.876	0.863	0.909	0.919	**0.995**
	Yoga	0.356	0.689	0.652	0.710	0.702	0.744	0.771	0.811	0.807	0.780	**0.926**
	AverageValue	0.589	0.639	0.784	**0.774**	0.876	0.880	0.856	0.876	0.894	0.873	**0.962**
	STD	0.179	0.218	0.112	0.215	0.122	0.092	0.061	0.055	0.075	0.083	**0.031**
	AverageRank	10.4	9	8.4	7.2	4.6	4.6	6.6	4.6	3.8	5.4	**1.4**
	CD	7.00										
G_{mean}	Adiac	0.782	0.897	0.875	0.920	**0.999**	0.957	0.910	0.920	0.958	0.975	0.989
	S-Leaf	0.721	0.898	0.798	0.572	0.898	0.902	0.809	0.812	0.887	0.856	**0.976**
	Wafer	0.817	0.970	0.953	0.870	0.984	0.903	0.907	0.956	**0.998**	0.937	0.988
	FaceAll	0.792	0.918	0.983	0.985	0.957	0.966	0.870	0.860	0.929	0.925	**0.997**
	Yoga	0.464	0.688	0.695	0.741	0.735	0.630	0.807	0.803	0.808	0.774	**0.976**
	AverageValue	0.713	0.874	**0.861**	**0.818**	**0.915**	0.872	0.861	0.870	0.916	0.893	**0.985**
	STD	0.145	0.108	0.117	0.164	0.108	0.113	0.051	0.067	0.073	0.079	**0.009**
	AverageRank	10.80	6.50	7.20	7.10	3.50	7.10	7.20	6.50	3.20	5.50	**1.40**
	CD	7.00										

Figs 3 and 4 present a comparison of the performance of previous work (learning methods and the under-sampling method with various algorithms) and this new approach, H-sampling to enhance bagging, using the Nemenyi test, where the x-axis indicates the ranking order of the learning methods and learning algorithms; the y-axis indicates the average rank of F_{value} and G_{mean} performance, respectively, and the vertical bars indicate the “Critical Difference”.

Groups of learning methods and learning algorithms that are no statistically significant difference at a 95% confidence interval are indicated when the vertical bars overlap. Comparing the previous approaches [10, 5] and this approach, H-sampling to enhance bagging, based on F_{value} and G_{mean}, H-sampling with bagging has the best average rank on both measures. KNN with under-sampling method has the second best average rank of F_{value} and G_{mean}. The statistical test results demonstrate that H-sampling with bagging method is statistically significantly better than 1NN, Bal. and Easy on F_{value}, and better than Easy on G_{mean} at a 95% confidence interval; however, there is no statistically significant difference between this new approach, H-sampling with bagging and previous approaches, over-sampling SPO and under-sampling KNN at a 95% confidence interval on both the F_{value} and G_{mean} measures.

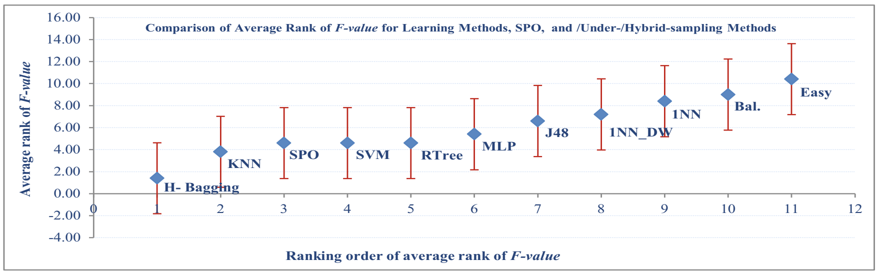

Fig. 3. Comparison of average rank of the F_{value} metric with the Nemenyi test for the learning methods, SPO, under-sampling, and H-sampling methods, where the x-axis indicates the ranking order of all the learning methods and sampling methods with learning algorithms, the y-axis indicates the average rank of F_{value}, and the vertical bars indicate the "Critical Difference"

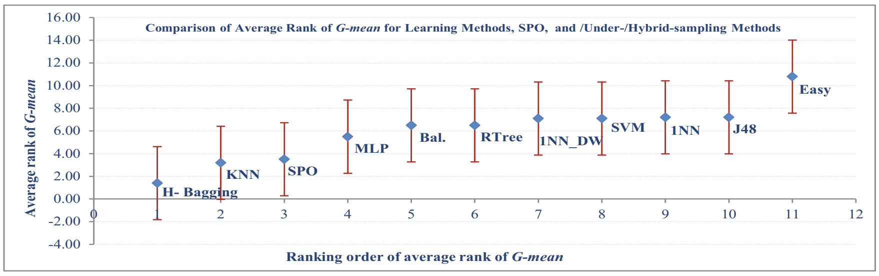

Fig. 4. Comparison of average rank of the G_{mean} metric with the Nemenyi test for the learning methods, SPO, under-sampling, and H-sampling methods, where the x-axis indicates the ranking order of all the learning methods, sampling methods with learning algorithms, the y-axis indicates the average rank of G_{mean}, and the vertical bars indicate the "Critical Difference"

7 Conclusion

This paper has addressed the limitations of existing techniques of over-sampling and under-sampling methods, and proposed a new approach, H-sampling schema to enhance bagging for improving the performance of previous approaches. It has empirically compared this new approach with the previous approaches of over-sampling, SPO and under-sampling with various algorithms based on two evaluation measures, F_{value} and G_{mean} on benchmark data-sets. Statistical tests are used to draw valid conclusions.

This new approach, H-sampling, reduces the computational cost and training time of over-sampling by using fewer positives in training, and increases the capability of under-sampling by using more negatives for training. Bagging is one of most effective ensemble learning methods for improving the performance of the individual classifiers when base learners are unstable, but it also has a limitation on highly imbalanced problems. The new approach integrates H-sampling technique with an unstable base learner $J48$ to enhance bagging for further improving the performance of the previous approaches. The experimental results demonstrate that this new approach H-sampling method to enhance the bagging dramatically improves the performance of the previous approaches, over-sampling SPO and under-sampling with KNN. This new approach achieves the highest average value with the lowest STD and the lowest average rank on both evaluation measures, and it is superior to previous approaches on both evaluation measures. For future work, I would like to investigate the impact of the performance of H-sampling integrating bagging with other base learners: unstable learners and stable learners.

Acknowledgments. Special thanks to Dr. Tony Bagnall for generously sharing his ARFF format time series data-sets.

References

1. Yang, Q., Wu, X.: 10 challenging problems in data mining research. International Journal of Information Technology & Decision Making 5(4), 597–604 (2006)
2. Hoens, T.R., Qian, Q., Chawla, N.V., Zhou, Z.-H.: Building decision trees for the multi-class imbalance problem. In: Tan, P.-N., Chawla, S., Ho, C.K., Bailey, J. (eds.) PAKDD 2012, Part I. LNCS, vol. 7301, pp. 122–134. Springer, Heidelberg (2012)
3. Liang, G., Zhang, C.: An efficient and simple under-sampling technique for imbalanced time series classification. In: CIKM 2012, pp. 2339–2342 (2012)
4. Hidasi, B., Gáspár-Papanek, C.: ShiftTree: An interpretable model-based approach for time series classification. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) ECML PKDD 2011, Part II. LNCS, vol. 6912, pp. 48–64. Springer, Heidelberg (2011)
5. Liang, G., Zhang, C.: A comparative study of sampling methods and algorithms for imbalanced time series classification. In: Thielscher, M., Zhang, D. (eds.) AI 2012. LNCS, vol. 7691, pp. 637–648. Springer, Heidelberg (2012)
6. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)

7. Quinlan, J.: Bagging, boosting, and c4.5. In: Proceedings of the 13th National Conference on Artificial Intelligence, pp. 725–730 (1996)
8. Bauer, E., Kohavi, R.: An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning 36(1), 105–139 (1999)
9. Dietterich, T.: An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine Learning 40(2), 139–157 (2000)
10. Cao, H., Li, X., Woon, Y., Ng, S.: SPO: Structure preserving oversampling for imbalanced time series classification. In: Proceedings of the IEEE 11th International Conference on Data Mining, ICDM 2011, pp. 1008–1013 (2011)
11. Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16(1), 321–357 (2002)
12. 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)
13. He, H., Bai, Y., Garcia, E., Li, S.: ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: IEEE International Joint Conference on Neural Networks, IJCNN 2008, pp. 1322–1328. IEEE (2008)
14. Guo, H., Viktor, H.L.: Learning from imbalanced data sets with boosting and data generation: The DataBoost-IM approach. ACM SIGKDD Explorations Newsletter 6(1), 30–39 (2004)
15. Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 39(2), 539–550 (2009)
16. Wei, L., Keogh, E.: Semi-supervised time series classification. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 748–753. ACM (2006)
17. Xi, X., Keogh, E., Shelton, C., Wei, L., Ratanamahatana, C.A.: Fast time series classification using numerosity reduction. In: Proceedings of the 23rd International Conference on Machine Learning, ICML 2006, pp. 1033–1040 (2006)
18. Witten, I., Frank, E.: Data Mining: Practical Machine Learning Tool and Techniques. Morgan Kaufmann (2005)
19. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)
20. Liang, G.: An investigation of sensitivity on bagging predictors: An empirical approach. In: 26th AAAI Conference on Artificial Intelligence, pp. 2439–2440 (2012)
21. Liang, G., Zhu, X., Zhang, C.: The effect of varying levels of class distribution on bagging with different algorithms: An empirical study. International Journal of Machine Learning and Cybernetics (2012), http://link.springer.com/article/10.1007%2Fs13042--012--0125--5
22. Liang, G., Zhang, C.: Empirical study of bagging predictors on medical data. In: 9th Australian Data Mining Conference, AusDM 2011, pp. 31–40 (2011)
23. Keogh, E., Zhu, Q., Hu, B., Hao, Y., Xi, X., Wei, L., Ratanamahatana, C.A.: The UCR Time Series Classification/Clustering homepage (2011), http://www.cs.ucr.edu/~eamonn/time_series_data/
24. Liang, G., Zhu, X., Zhang, C.: An empirical study of bagging predictors for imbalanced data with different levels of class distribution. In: Wang, D., Reynolds, M. (eds.) AI 2011. LNCS, vol. 7106, pp. 213–222. Springer, Heidelberg (2011)