Journal of Theoretical and Applied Information Technology
10th February 2015. Vol.72 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NEW APPROACH WITH ENSEMBLE METHOD TO ADDRESS CLASS IMBALANCE PROBLEM

[1]SEYYEDALI FATTAHI, [2]ZALINDA OTHMAN, [3]ZULAIHA ALI OTHMAN

[1,2,3]Affiliation , Data mining and Optimization Research Group, Centre for Artificial Intelligence, Faculty of Information Science and Technology, University KEBANGSAAN MALAYSIA,UKM BANGI, 43600 Selangor,

E-mail: [1]sa.fattahi@yahoo.com, seyyedalifattahi@siswa.ukm.edu.my, [2]zalinda@ukm.edu.my, [3]zao@ukm.edu.my

ABSTRACT

An attractive research in recent years is solving class imbalance problem in imbalanced dataset. The class is imbalanced when the number of one class (majority) is more than another one (minority). The classification of this imbalanced class causes imbalanced distribution and poor predictive classification accuracy. This paper introduces a new ensemble –based method for imbalanced data set classification using Synthetic Minority Over-sampling Technique (SMOTE) and Rotation Forest algorithm to address class imbalance problem. Rotation Forest applied as ensemble classifier combines with well-known re-sampling method (SMOTE). It constructs classifiers with obtaining features by rotating subspaces of the original dataset. The advantages of Rotation Forest rather than other ensemble methods (Boosting, Bagging, Random Subspace) is that same information held as original data sets and no information lost in data sets which used to construct classifiers. Experimental results reveal the effectiveness of SMOTE and Rotation Forest performance at data level in overall accuracy, Cohen's kappa Coefficient, False Negative rate, AUC, and RMSE compared to other related classification ensemble methods (SMOTE-Boost, SMOTE-Bagging, SMOTE-random subspace) on twenty KEEL repository imbalanced datasets (binary dataset not multi-class) which selected randomly from different ratios by implementing Java-based WEKA and STATISTICA software. SMOTE implemented for training data by values of N=100, 200, 300, and 400. Kappa-Error diagram is plotted to analysis the behavior of ensemble methods. The experimental results clarify the validness of proposed ensemble classifier.

Keywords: *SMOTE, Rotation Forest, Random Subspace, Bagging, Boosting*

1 INTRODUCTION

In binary classification, researchers have typically ignored the dataset balancing and instead assumed that the training sets are balanced. Some surveys [1, 2] have indicated that balancing the class distribution of datasets provides better classification performance than using imbalanced datasets (IDs). However, some researchers [3, 4] have argued that the imbalanced dataset problem (IDP) has only a slight effect on classification performance. The IDP occurs when training and test datasets include both negative and positive instances. For example, if the dataset contains 100 positive instances and 10,000 negative instances, then the use of classification algorithms will exhibit underperformance and over-fitting problems due to the IDs. This problem is also called the class imbalance problem (CIP) because there are significantly unequal quantities of positive and negative examples in the dataset being classified [5, 6]. The rare instances (minority) contain more significant information or concepts than the more common instances (majority). To overcome this problem, [5] generated a new dataset called the SMOTED dataset. They generated dataset samples by determining the differences between samples (feature vector) and their related nearest neighbors, multiplied these dissimilarities by a random number between 0 and 1, and created a new feature vector by adding these numbers. At this time, the synthetic examples are generated, and the minority class is over-sampled by introducing synthetic samples alongside the line sections between two specific features. These new samples are joined by any or all of the k-nearest neighbors (k minority class nearest neighbors).

A significant aspect of our experiments in relation to SMOTE is the opportunity for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

combining over-sampling methods with other classifiers as ensemble method. As mentioned by [5], SMOTE has upward compatibility with classifiers. This advantage allows authors and researchers to utilize this technique for over-sampling and then combining with other classifiers. This work proposes a new ensemble method based on SMOTE and Rotation Forest. Rotation Forest is a strong ensemble-based classifier [7] and is described in the methodology section.

1.1 Similar works and related methods

Previous studies have aimed to improve the accuracy of classifiers in a similar manner, and we have compared these studies with our own work. The first such ensemble method is SMOTE-Boosting. [8] combined SMOTE with standard boosting and analyzed the degree to which SMOTE-Boosting can improve the classification accuracy with boosting to avoid misclassification. Standard boosting creates equal weights for all misclassified samples [9] but not hold original information about data. SMOTE-Boosting enhances the sampling weights in the minority class. The achievements of SMOTE-Boosting are high classifier recall and precision.

Another ensemble classifier is a machine learning classifier called BAGGING (Bootstrap AGGregatING) [10]. SMOTE-Bagging [11] is the ensemble classifier that generate samples by bootstrapping (automatic) samples in the original dataset and forcing each sample to create an essential member [12, 13]. [13] compared SMOTE-Boosting and SMOTE-Bagging and noted that ensemble methods, such as boosting and bagging, are useful for improving the accuracy of classifiers by combining pre-processing techniques. [14] proposed an ensemble classifier called SMOTE-Random subspace. This method creates a randomly selected feature subset instead of modifying the training samples (Adaboost) using the training data.

To prove and clarify the validness of proposed new method in compare with exist ensemble methods, we applied our methods on twenty binary datasets of KEEL repository datasets [15, 16] and compare performance at data level in precision, recall, Cohen's kappa Coefficient ,False Negative rate ,AUC , and RMSE.

1.2 Motivation and Justification

Regarding the motivation and justification of this work, we asked five significant questions:

a) What is the main definition of the class imbalance problem in machine learning?

b) Why is class imbalance a significant problem in machine learning?

c) How we can improve the binary classification accuracy of a classifier?

d) Why the ensemble methods have efficient performance for tackle class imbalanced problem?

e) Why Rotation Forest algorithm can be better base classifier in ensemble method?

The main definition of the class imbalance problem in machine learning is the problem in one classification type (binary classification) when the total number of members in one class is substantially different from the number of members in the other class [4, 5, 17, 18]. This problem is particularly severe at various practical boundaries, including anomaly detection [19], fraud detection [20], oil spilling detection [21, 22], and medical diagnosis [23, 24].

In machine learning approaches, the class imbalance problem is divided into two groups: the data level and algorithmic level [4, 25]. The data level includes over-sampling, under-sampling, and hybrid sampling, and the algorithmic level includes cost adjustment, the decision threshold, and probabilistic estimation [23, 26]. We focus on the data level because the motivation of our work is to improve the accuracy of classifiers at the data level. The accuracy improvement depends on the classifier. Based on ensemble learning, in this work, we combined the SOME and Rotation Forest, as ensemble techniques have flexibility with learning algorithms [13] to improve overall accuracy.

[27, 28] categorized reasons of ensemble constructing into three fundamental reasons are Statistical, Computational and Representational. Ensemble methods can reduce the bias and variance of learning algorithm and have efficient performance in tackling class imbalanced problem [13, 27-34]. Between well-known ensemble methods (Rotation forest, Bagging, Boosting, and Random subspace) we chose rotation forest as base classifier in our ensemble methods based on below advantages [7, 35-39]

•Build accurate and diverse classifier

•Fix and hold information about training data through constructing classifier (no data lost-the key benefit)

•Principle Component Analysis (PCA) is utilized to extract the model and transform the data (PCA increase building accurate and diverse classifier through feature extraction)

Journal of Theoretical and Applied Information Technology
10th February 2015. Vol.72 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

•Model of base classifier: decision tree (the main idea for calling "forest")

The main limitation of the introduced method is a noisy setting. SMOTE-RO-FO was considered on noise-free datasets as SMOTE-Boost; despite SMOTE-Bagging and SMOTE-RS being robust to address noisy datasets..

2 METHODOLOGY

We proposed a SMOTE-Rotation Forest method, called SMOTE-RO-FO, for ID classification. This approach is an ensemble method with one over-sampling algorithm (SMOTE) [40] and one ensemble-learning algorithm (Rotation Forest). Rotation Forest is a classifier based on feature extraction which transforms the dataset with preserving all information about original data [7]. Principal Component Analysis (PCA) through the performance transforms the data by:

•Subset of the samples (instances)

•Subset of the classes (two classes)

•Subset of the features

The proposed methodology for addressing class imbalanced problem divided into three steps:

Step 1: Let give X: the objects in the training dataset

X = [X1, X2… Xn] ^{T}: Data point with n features: (N × n) matrix as equation 1

$$X = \begin{bmatrix} x_1^1 & x_2^1 & .... & x_n^1 \\ : & : & & : \\ : & & : & : \\ x_1^N & x_2^N & .... & x_n^N \end{bmatrix} \quad (1)$$

Y = [Y1, Y2, … , YN]^T : class labels with C imbalanced classes: (N × 1) matrix

Cm: Number of minority classes where

L: Number of classifiers in the ensemble

F: the feature set

Step 2: prepare rotation matrix

For i = 1, 2… L (construct the training set for classifier Di by creating N synthetic examples from minority class Cm using the SMOTE algorithm)

For j = 1, 2… k (split F into k subsets where each $M = n/k$ has features)

$X_{i,j}$: Dataset X for the features in $F_{i,j}$ ($X_{1,1}$: dataset X for the features in $F_{1,1}$) – Figure 1 shows the feature sets

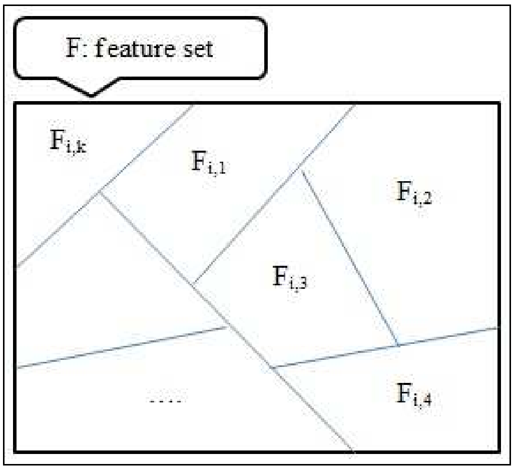

Figure 1: Feature Sets Splitting

Step 3: eliminate random subset of classes from $F_{i,j}$

Select a bootstrap sample from $X_{i,j}$ to obtain new samples $X'_{i,j}$

Run PCA on $X'_{i,j}$ using only M features

Arrange the principle components for all j to obtain rotation matrix

Rearrange the rows of R_i (as equation 2) to match the order of features in F and obtain R_i^a

Build classifier D_i using XR_i^a as a training set

Journal of Theoretical and Applied Information Technology
10th February 2015. Vol.72 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$R_1 = \begin{bmatrix} a_{1,1}^{(1)}, a_{1,1}^{(2)}, \dots, a_{1,1}^{(M_1)} & 0 & \dots & 0 \\ 0 & a_{1,2}^{(1)}, a_{1,2}^{(2)}, \dots, a_{1,2}^{(M_2)} & \dots & 0 \\ \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & a_{1,k}^{(1)}, a_{1,k}^{(2)}, \dots, a_{1,k}^{(M_k)} \end{bmatrix} \quad (2)$$

In step 1 we insert training data as ($N \times$ n) matrix and set the classes of this training data as $(N \times 1)$ matrix with Cm minority class. In step 2, we create rotation matrix which is necessary for rotation forest algorithm. Before preparing rotation matrix, the synthetic datasets which is constructed by SMOTE, inserted and then split into f feature sets (figure 1). For eliminating random subset of classes through splitting, at the next step we select bootstrap samples which are created by split data and obtain new bootstrapped samples. After this, Principle Component analysis is run to transform data and increase constructing accurate classifier by arranging and rearranging PCA. Now we built ensemble classifier.

[35] adapted rotation forest algorithm for solving regression problem. In this work, we ignored the regression problem and focus on class imbalanced problem.

3 EVALUATION MEASUREMENTS

3.1 Confusion Matrix

The confusion matrix is a straightforward technique for evaluating classifier performance in the binary classification (Table 1). It records the resulting correctly and incorrectly recognized examples of each class. The first column provides the class labels (Negative, Positive) of samples, and the first row of the table provides the class label predictions (true rate, false rate).

Table 1: Confusion Matrix

<table>
<tr><td colspan="2"></td><td colspan="2">Predicted class</td></tr>
<tr><td colspan="2">confusion matrix</td><td>(Positive)</td><td>(Negative)</td></tr>
<tr><td rowspan="2">Actual class</td><td>Positive class</td><td>TP</td><td>FN</td></tr>
<tr><td>Negative class</td><td>FP</td><td>TN</td></tr>
</table>

The confusion matrix reports the validity of the classification models. The quality and validity measures of the classification are built on the confusion matrix.

Table 1 presents the four important categories. The first category is True Negative (TN), i.e., negative examples that are classified correctly as negative. The second category is True Positive (TP), i.e., positive examples that are classified correctly as positive. The third category is False Negative (FN), i.e., positive examples that are classified incorrectly as negative. The fourth category is False Positive (FP), i.e., negative examples that are classified incorrectly as positive [23, 26, 27]. These rates are given by Equations (3)-(6).

TN = True Negative rate (Percentages of negative belongings that classified correctly and belonging to negative class) = specificity

TP = True Positive rate (Percentages of positive belongings that classified correctly and belonging to positive class) = recall = sensitivity

FN = False negative rate (Percentages of positive belongings that misclassified and belonging to negative class) => it will show which classifier is better to choose (FN with low percentage value is considered).

FP = False positive rate (Percentages of negative belongings that misclassified and belonging to positive class) = fall out

$$TN_{rate} = \frac{TN}{TN + FP} \quad (3) \qquad TP_{rate} = \frac{TP}{TP + FN} \quad (4)$$

$$FN_{rate} = \frac{FN}{FN + TP} \quad (5) \qquad FP_{rate} = \frac{FP}{FP + TN} \quad (6)$$

3.2 Classifier Accuracy Evaluation

The accuracy evaluation is a prepared metric for ID that is suitable for majority class prediction but not for minority class prediction. Thus, more practical evaluation metrics, such as precision, recall, F-value, and area under the curve (AUC), are required. The accuracy of the classifier is determined by Equation (7).

$$Classifier\ \ Accuracy = \frac{\mathrm{TP + TN}}{\mathrm{TP + FN + FP + TN}} \quad (7)$$

Precision, Recall, F-value defined as equations (8)-(10)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\text{Precision} = \frac{TP}{TP + FP} \quad \textbf{(8)}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} = TP_{rate} \quad \textbf{(9)}$$

$$F - value = 2 \times \frac{precision \times recall}{precision + recall} \quad \textbf{(10)}$$

The F-value (F-score) defined by Equation (10) is a metric for comparing the precision and recall. The tradeoff of recall and precision is a significant factor in the F-value measurement. Precision is in conflict with recall, and the precision decreases when both TP and FP increase (see equation (8)). In this case, F-value is used to define the goodness of the classifier or learning algorithm (Precision, recall and F-score did not indicate in this paper).

3.3 Kappa Statistic

Kappa statistic or Cohen's Kappa coefficient (CKC) which introduced by [41] is a statistical metrics to assess measuring agreement [42-45], intera-rater reliability [46-48]. In a simple way, the equation of Kappa statistic indicates by Equation (12).

$$\kappa = \frac{\text{observed agreement - chance agreement}}{\text{1 - chance agreement}} \quad (12)$$

If metric agrees by chance purely, there is not really agreement at all despite, only agreement by expected by chance consider as a true agreement. Kappa is a metric of true agreement. There is a proportion for every observed and chance agreement which proportion of any possible beyond chance agreement as an achieved beyond chance agreement [48].

The magnitude range of kappa is diverse usually between 0 and 1. The value 1 indicates perfect agreement and represents the raters agreeing in own classification of every case. The value 0 indicates agreement not better than which expected by chance. In clinical and medical context rarely these ranges are from -1 to 1. The negative range means if two raters are considered, although with more than two raters it will be consider as possible as minimum higher value [49]. We used CKC diagram (Figures 2-3) to analysis of our classifier ensembles [50].

3.4 Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)

MAE measures the tradeoffs between predictors (unknown outputs) and outcomes (known outputs) as the average value of the errors. MAE ignored their direction. In other words, MAE is a metric which measures differences between predicted results and observed results. Despite some researchers [51, 52] argued the advantages of MAE in assessing average model performance but other authors [53-55] used RMSE as standard metric in model error.

RMSE measures as MAE but has advantages in large error. RMSE has strong affect in large errors (penalize large errors). Both of these metrics have range between 0 to infinity. Result value which is closer to 0 is perfect values. We utilized both of MAE and RMSE as metric tools for performance of classifiers to represent the variety between them. Equations 12, 13 are interpreted these metrics. Which e shows error rate and n indicates number of samples.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|e_i\right| \quad (12) \qquad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \quad (13)$$

4 EXPERIMENTS AND RESULTS

The performance of SMOTE-RO FO is examined and compared with the three well-known ensemble methods illustrated by table 5. The experimental results for the twenty IDs illustrated in figures 4-7 for the average accuracy of classifiers, average AUC for the receiver operating characteristic (ROC) [56, 57] average CKC, and average of RMSE.

AUCs for the applied methods indicate the SMOTE-RO-RO has the highest accuracy among the classifiers considered. The AUC is a measurement tool for determining the power of a test [58, 59]. The AUC has a statistical meaning for the Wilcoxon test of ranks [60]. The CKC and AUC are the main evaluation method in this work, defined by Equations (12) and (14) because in this work we focus to improve accuracy with low error rate. The area under the ROC indicates which model performs better on average. In other words, a larger AUC is desired.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \qquad (14)$$

4.1 Datasets

All twenty datasets are downloaded from KEEL repository dataset [15, 16] and have not any missing value. Imbalanced Ratio (IR) of these datasets is between 1.82 and 8.79.

4.2 Methods and Options

Tables 2 - 4 show all parameters about setting, methods and options are applied in WEKA [61-64].

Table 2: parameters specification for SMOTE

Parameters
Class value: 0 (Auto – detected the non-empty minority class)
Nearest neighbours: 5 (The number of nearest neighbour to use)
Percentage: 100, 200, 300, 400, and 500 (The percentage of SMOTE instances to create)
Random seed: 1 (Seed used for random sampling)

Table 3: Parameters specification for C4.5

Parameters
Prune: true (Pruning is performed)
Confidence factor: 0.25 (The confidence factor used for pruning – small values incur more pruning)
MinNumObj: 2 (Minimum number of instanced per leaf)
UseLaplace: false

Table 4: Parameters specification for Rotation Forest, AdabosstM1, Bagging and Random Subspace

Rotation Forest
Number of features in subset: 3 (Rest is by default in WEKA)
Number of iteration: 10
AdaboostM1
Number of iteration: 10 (Rest is by default in WEKA)
Bagging
Bag size percentage: 100
Number of iteration: 10 (Rest is by default in WEKA)
Random Subspace
Sub Space size: 0.5
Number of iteration: 10 (Rest is by default in WEKA)

4.3 Kappa-Error Diagram

Figures 2-3 show the kappa error in compare with RMSE for four ensemble methods, which applied for twenty imbalanced datasets. In Kappa-error, the small value of kappa shows the more divers. The results indicate that SMOTE-RO FO is not more diverse as other ensembles but is most accurate classifier. Figure 2 illustrate the best domain for ensemble classifiers with most accurate value and less error value. SMOTE –RO FO is similar to SMOTE Bagging but more accurate. Figure 3 shows 3D view of figure 2 and clearly indicates SMOTE-RO FO output is in the best domain with highly accurate values.

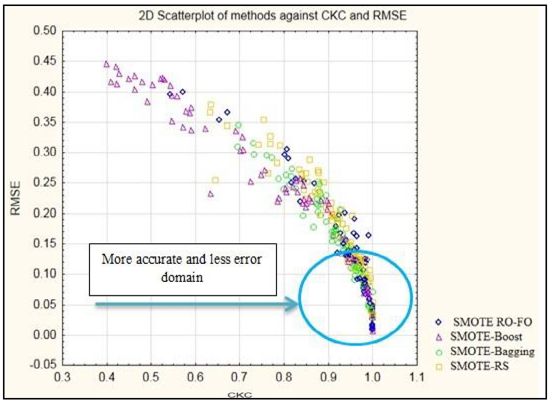

Figure 2: Scatterplot of RMSE and CKC

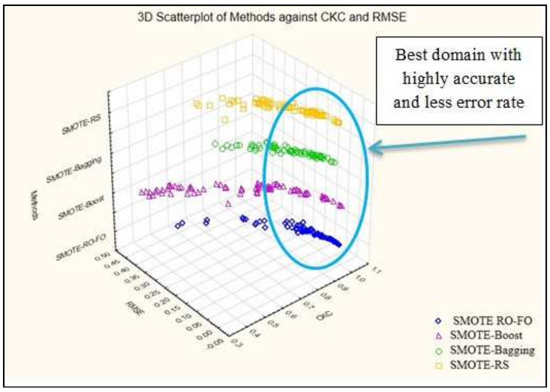

Figure 3: 3D scatterplot of Methods, RMSE and CKC

4.4 Results

Figure 4 shows average of overall accuracy of ensemble methods performance in N= 100%, 200%, 300%, and 400% of sample size. SMOTE-RO FO has more accurate values in compare with other ensemble methods. Totally, the acceptable and efficient accuracies for SMOTE-RO FO obtained in N= 200%, 300, and 400%. Figures 5-6 plotted AUC and CKC output for twenty datasets. We compared these metrics together for introduced ensemble method to show performance is better on average with large AUC and closer CKC magnitude to 1. Figure 6 shows the average values for CKC which this value of CKC for SMOTE-RO FO is closer to 1. Figure 7 also indicate that result values of RMSE is closer to 0 (perfect output) for SMOTE RO-FO performance.

Journal of Theoretical and Applied Information Technology
10th February 2015. Vol.72 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

These average magnitudes present low error and high agreement of SMOTE-RO FO performance and outputs. We showed four evaluation metrics by average values: Accuracy, AUC, CKC, and RMSE. The evaluation results indicated that our introduced method significantly outperformed the other ensemble methods. Also experimental results confirmed the effectiveness of our approaches and achieved higher accuracy for both minority and majority classes than other techniques.

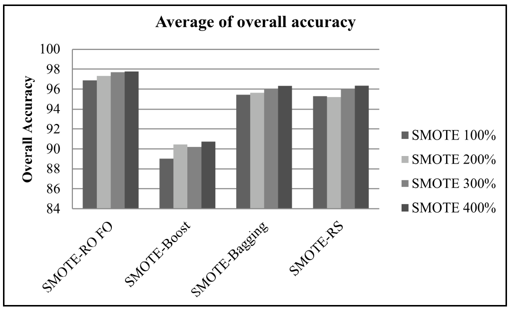

Figure 4: Average of overall accuracy

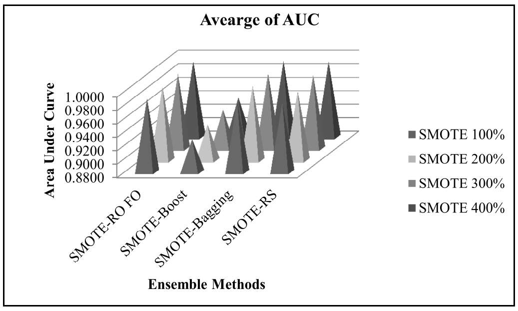

Figure 5: Average of AUC

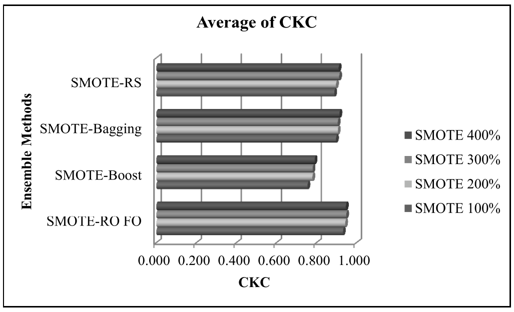

Figure 6: Average of CKC

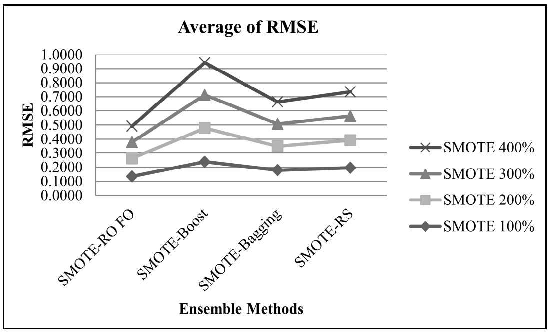

Figure 7: Average of RMSE

5 THE LIMITATION OF USING ACCURACY AND ERROR RATE METRICS

[3] noted that the ROC metric tools consider different classification errors to be equally significant. However, these metrics are suspect when the class distribution is strongly biased toward the majority class. For this reason, we test false negative rate measurement to prove the evaluation results. We compare average of FNrate for our method and other ensemble methods. In this case, if the value of FNrate is closer to 0, it is better to choose the machine learning algorithm as the classifier [17, 65]. Equation (15) based on results in figure 8 indicates average of FNrate for SMOTE-RO FO is less than FNrate for other methods in average. On the other hand, equation (15) and figure 8 indicate that type II error of SMOTE-RO FO is less than other methods.

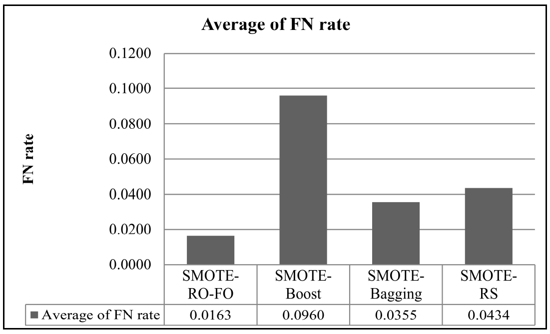

Figure 8: Average of False Negative Rate for ensemble methods

$$FN_{rate}(SMOTE\text{-}RO\text{-}FO) < FN_{rate}(OtherMethods) \quad (15)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6 ABBREVIATIONS AND ACRONYMS

SMOTE: Synthetic Minority Over-sampling Technique; AdaBoost : Adapted Boosting Algorithm; Bagging: Bootstrap aggregating; PCA: Principle Component Analysis; ROC: Receiver Operating Characteristic; K-NN: K – Nearest Neighbor classifier; AUC: Area Under ROC Curve; FN: False Negative rate; CKC: Cohen's Kappa Coefficient; RMSE: Root Mean Square Error; KEEL: Knowledge Extraction based on Evolutionary Learning; WEKA: Waikato Environment for Knowledge Analysis.

7 FUTURE WORK

Potential future directions of this work include (i) extending the introduced method to compare with other over-sampling, under-sampling techniques, support vector machine, and neural network, (ii) evaluating with more evaluation metrics (Precision, Recall, F-Score) in imbalance domain, (iii) implementing SMOTE RO FO to imbalance datasets with IR more than 8.79, and 500% of its original size.

8 SUMMARY OF RESULTS

Table 5 shows the summary of different ensemble methods which performed in CIP for twenty imbalance datasets. H indicates High, L indicates low, VH indicates very high, and VL indicates very low.

Table 5: summary of results

Study	Accuracy	AUC	CKC	RMSE	FN rate
SMOTE-Boost (Chawla, Nitesh V ,2003)	L	L	H	H	H
SMOTE-Bagging(Wang, Shuo, 2009)	H	H	L	L	L
SMOTE-RS (Huang, Hsiao-Yun, 2012)	H	H	L	L	L
SMOTE-RO FO (Proposed method)	VH	VH	VL	VL	VL

.

9 CONCLUSIONS

In this work we analyzed the behavior of four ensemble methods to address class imbalance problem in data level for binary class. Our results show that the novel ensemble method in practice for twenty imbalanced datasets, obtained effective and accurate results. On the other hand, our introduce ensemble method (SMOTE-RO FO) established as ensemble method to obtain high accurate classifier. The SMOTE-RO FO method successfully used the benefits from SMOTE and Rotation Forest algorithm in class imbalanced problem. Which Rotation Forest improved the accuracy of classifier by focusing on class samples and SMOTE algorithm improved the performance of classifier in the minority of class samples. It should be noted that although SMOTE-RO FO is less diverse in general but has most accurate value and less error rate. Acceptable results for SMOTE-RO FO obtained in N= 200%, 300% and 400% of sample size. In compare with performance of other ensemble methods, SMOTE-RO FO is similar to SMOTE-Bagging but more accurate. Totally, in average SMOTE-RO FO can address class imbalance problem with high overall accuracy, better AUC, efficient CKC, and lower FNrate (type II error) in performance. The evaluation results are proven together and justified validness of introduced novel ensemble method.

REFERENCES

[1] A. Estabrooks, T. Jo, and N. Japkowicz, "A multiple resampling method for learning from imbalanced data sets," Computational Intelligence, vol. 20, pp. 18-36, 2004.
[2] G. M. Weiss and F. J. Provost, "Learning when training data are costly: the effect of class distribution on tree induction," J. Artif. Intell. Res.(JAIR), vol. 19, pp. 315-354, 2003.
[3] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," ACM SIGKDD Explorations Newsletter, vol. 6, pp. 20-29, 2004.
[4] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," Intelligent data analysis, vol. 6, pp. 429-449, 2002.
[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: special issue on learning from imbalanced data sets," ACM SIGKDD Explorations Newsletter, vol. 6, pp. 1-6, 2004.
[7] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, pp. 1619-1630, 2006.
[8] N. Chawla, A. Lazarevic, L. Hall, and K. Bowyer, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting," in Knowledge Discovery in Databases: PKDD 2003. vol. 2838, N. Lavrač, D. Gamberger, L. Todorovski, and H. Blockeel, Eds., ed: Springer Berlin Heidelberg, 2003, pp. 107-119.
[9] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Resampling or Reweighting: A Comparison of Boosting Implementations," in Tools with Artificial Intelligence, 2008. ICTAI'08. 20th IEEE International Conference on, 2008, pp. 445-451.
[10] J. R. Quinlan, "Bagging, boosting, and C4. 5," 1996, pp. 725-730.
[11] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," in Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on, 2009, pp. 324-331.
[12] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," Machine learning, vol. 36, pp. 105-139, 1999.
[13] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 42, pp. 463-484, 2012.
[14] H.-Y. Huang, Y.-J. Lin, Y.-S. Chen, and H.-Y. Lu, "Imbalanced data classification using random subspace method and SMOTE," in Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on, 2012, pp. 817-820.
[15] J. Alcalá, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, et al., "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," Journal of Multiple-Valued Logic and Soft Computing, 2010.
[16] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, et al., "KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework," Journal of Multiple-Valued Logic & Soft Computing, vol. 17, 2011.
[17] N. Japkowicz, "The class imbalance problem: Significance and strategies," in Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI’2000), 2000, pp. 111-117.
[18] N. V. Chawla, "Data mining for imbalanced datasets: An overview," Data Mining and Knowledge Discovery Handbook, pp. 875-886, 2010.
[19] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Computing Surveys (CSUR), vol. 41, p. 15, 2009.
[20] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: classification of skewed data," ACM SIGKDD Explorations Newsletter, vol. 6, pp. 50-59, 2004.
[21] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," Machine learning, vol. 30, pp. 195-215, 1998.
[22] C. Brekke and A. H. Solberg, "Classifiers and confidence estimation for oil spill detection in ENVISAT ASAR images," Geoscience and Remote Sensing Letters, IEEE, vol. 5, pp. 65-69, 2008.
[23] Y. Yang and G. Ma, "Ensemble-based active learning for class imbalance problem," J. Biomedical Science and Engineering, vol. 3, pp. 1021-1028, 2010.
[24] D.-C. Li, C.-W. Liu, and S. C. Hu, "A learning method for the class imbalance problem with medical data sets," Computers in Biology and Medicine, vol. 40, pp. 509-518, 2010.
[25] Y. Sun, A. K. C. Wong, and S. K. MOHAMED, "Classification of imbalanced data: A review," International Journal of Pattern Recognition and Artificial Intelligence, vol. 23, pp. 687-719, 2009.
[26] R. Barandela, J. S. Sanchez, V. Garcia, and E. Rangel, "Strategies for learning in class imbalance problems," Pattern Recognition, vol. 36, pp. 849-851, Mar 2003.
[27] T. Dietterich, "Ensemble methods in machine learning," Multiple classifier systems, pp. 1-15, 2000.
[28] T. G. Dietterichl, "Ensemble learning," The handbook of brain theory and neural networks, pp. 405-408, 2002.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[29] L. Rokach, "Ensemble methods in supervised learning," in Data mining and knowledge discovery handbook, ed: Springer, 2010, pp. 959-979.
[30] J. J. Liao, C. H. Shih, T. F. Chen, and M. F. Hsu, "An ensemble-based model for two-class imbalanced financial problem," Economic Modelling, vol. 37, pp. 175-183, Feb 2014.
[31] N. Rahman and D. Davis, "Addressing the Class Imbalance Problems in Medical Datasets," International Journal of Machine Learning and Computing, vol. 3, pp. 224-228, 2013.
[32] Y. Yang and G. Ma, "Ensemble-based active learning for class imbalance problem," Engineering, vol. 2, 2010.
[33] L. Shoemaker, "Ensemble Learning With Imbalanced Data," University of South Florida, 2010.
[34] V. Chitraa and A. S. Thanamani, "REVIEW OF ENSEMBLE CLASSIFICATION," 2013.
[35] C. Pardo, J. F. Diez-Pastor, C. García-Osorio, and J. J. Rodríguez, "Rotation Forests for regression," Applied Mathematics and Computation, vol. 219, pp. 9914-9924, 2013.
[36] L. I. Kuncheva and J. J. Rodríguez, "An experimental study on rotation forest ensembles," in Multiple Classifier Systems, ed: Springer, 2007, pp. 459-468.
[37] C.-X. Zhang, J.-S. Zhang, and G.-W. Wang, "An empirical study of using Rotation Forest to improve regressors," Applied Mathematics and Computation, vol. 195, pp. 618-629, 2/1/ 2008.
[38] C.-X. Zhang and J.-S. Zhang, "A variant of Rotation Forest for constructing ensemble classifiers," Pattern Analysis and Applications, vol. 13, pp. 59-77, 2010/02/01 2010.
[39] H. Jiawei and M. Kamber, "Data mining: concepts and techniques," San Francisco, CA, itd: Morgan Kaufmann, vol. 5, 2001.
[40] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
[41] J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, 1960.
[42] J. Carletta, "Assessing agreement on classification tasks: the kappa statistic," Computational linguistics, vol. 22, pp. 249-254, 1996.
[43] B. Di Eugenio and M. Glass, "The kappa statistic: A second look," Computational linguistics, vol. 30, pp. 95-101, 2004.
[44] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," biometrics, vol. 33, pp. 159-174, 1977.
[45] T. O. Kvalseth, "Note on Cohen's kappa," Psychological reports, vol. 65, pp. 223-226, 1989.
[46] J. L. Fleiss, B. Levin, and M. C. Paik, Statistical methods for rates and proportions: John Wiley & Sons, 2013.
[47] K. L. Gwet, "Variance estimation of nominal-scale inter-rater reliability with random selection of raters," Psychometrika, vol. 73, pp. 407-430, 2008.
[48] J. Sim and C. C. Wright, "The kappa statistic in reliability studies: use, interpretation, and sample size requirements," Physical therapy, vol. 85, pp. 257-268, 2005.
[49] T. McGinn, P. C. Wyer, T. B. Newman, S. Keitz, and R. Leipzig, "Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic)," Canadian Medical Association Journal, vol. 171, pp. 1369-1373, 2004.
[50] L. I. Kuncheva, "A bound on kappa-error diagrams for analysis of classifier ensembles," Knowledge and Data Engineering, IEEE Transactions on, vol. 25, pp. 494-501, 2013.
[51] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Climate Research, vol. 30, p. 79, 2005.
[52] T. Chai and R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?," Geoscientific Model Development Discussions, vol. 7, pp. 1525-1534, 2014.
[53] T. Chai, H.-C. Kim, P. Lee, D. Tong, L. Pan, Y. Tang, et al., "Evaluation of the United States National Air Quality Forecast Capability experimental real-time predictions in 2010 using Air Quality System ozone and NO 2 measurements," Geoscientific Model Development, vol. 6, 2013.
[54] S. McKeen, J. Wilczak, G. Grell, I. Djalalova, S. Peckham, E. Y. Hsie, et al., "Assessment of an ensemble of seven real□time ozone forecasts over eastern North America during the summer of 2004," Journal of Geophysical Research: Atmospheres (1984–2012), vol. 110, 2005.
[55] N. Savage, P. Agnew, L. Davis, C. Ordóñez, R. Thorpe, C. Johnson, et al., "Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

initial evaluation," Geoscientific Model Development, vol. 6, 2013.
[56] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern recognition, vol. 30, pp. 1145-1159, 1997.
[57] T. Fawcett, "An introduction to ROC analysis," Pattern recognition letters, vol. 27, pp. 861-874, 2006.
[58] D. Powers, "Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation," Journal of Machine Learning Technologies, vol. 2, pp. 37-63, 2011.
[59] S. Chen, H. He, and E. A. Garcia, "RAMOBoost: Ranked minority oversampling in boosting," Neural Networks, IEEE Transactions on, vol. 21, pp. 1624-1642, 2010.
[60] D. J. Hand, Construction and assessment of classification rules vol. 15: Wiley Chichester, 1997.
[61] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," ACM SIGKDD Explorations Newsletter, vol. 11, pp. 10-18, 2009.
[62] G. Holmes, A. Donkin, and I. H. Witten, "Weka: A machine learning workbench," in Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on, 1994, pp. 357-361.
[63] E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. H. Witten, et al., "Weka," in Data Mining and Knowledge Discovery Handbook, ed: Springer, 2005, pp. 1305-1314.
[64] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, et al., "WEKA---Experiences with a Java Open-Source Project," The Journal of Machine Learning Research, vol. 9999, pp. 2533-2541, 2010.
[65] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the class imbalance problem," in Natural Computation, 2008. ICNC'08. Fourth International Conference on, 2008, pp. 192-201.