ISBN 978-93-86878-08-3
12th International Conference on Recent Trends in Engineering and Technology (RTET-2018)
Jan. 10-11, 2018 Bali (Indonesia)

Ensemble Based Learning with Stacking, Boosting and Bagging for Unimodal Biometric Identification System

Supreetha Gowda H D[1], Hemantha Kumar G[1] and Mohammad Imran [2]

[1]Department of Computer Science, University of Mysore, MYSORE-560 007, India

[2]College of Computer Sciences & Information Technology, King Faisal University, AL-HASA, Saudi Arabia

Abstract: *In this paper, we propose Unimodal Biometric Identification System Based on Ensemble machine learning techniques such as Stacking, Boosting and Bagging. The article framework gives the comparative analysis between Ensemble techniques adopting different classifiers (Geometric based classifiers-KNN, SVM, NN and Decision tree based- RF, RPART) and the best feature extraction algorithms such as Texture based-LPQ, Appearance based- ICA1. The experimentation is carried out on physiological biometric traits adopting the standard benchmark databases such as AR facial database and Poly-U fingerprint database. The objective of the paper is to understand whether concentrating on discriminating feature extraction algorithms or performing extensive computation on ensemble techniques with different classification models would contribute greatly in performance of the system.*

Keywords: *Ensemble; Stacking; Boosting; Bagging; Random Forest; Recursive Partition; Decision trees.*

1. Introduction

Building an automated and reliable person identification system in exploring dystopian futures is very much allied with uberveillance and in tracking general populace unequivocally. The biometric data is diverse, it is impossible to generalize the common feature extraction algorithm for all kinds of biometric data such as 1D data (eg: signature, voice), 2D data (face, iris, palmprint) and 3D data(geometrical features of hand, 3D face). The efficient Biometric system utilizes the physiological and behavioral characteristics of users and establishes or confirms the identity by detecting, analyzing dominant features, stores the generic templates obtained from each class and finally announces the claimed identity as genuine or imposter by comparing the extracted patterns from the query image with the template already stored in the database with certain tolerant thresholds. When an individual claim to be a specific user, the biometric system performs (1:1) comparison in judging his identity is called verification, whereas identification (1:N) is the task of determining unknown identity. Once the permanent identifiers i.e the biometric traits or modalities is stealed, it is stolen for lifetime which cannot be recovered. As a result the cyber criminals bear in effort to steal data from the centralized databases. The mobile devices which sits in the jacket pocket manages to monitor the users activity by identifying his locations and updating the main server, owner operating the legacy biometric systems has the record of all the employees and becomes the effective arbiter owner for the identities recorded, mobile technologies especially connected with social media is the central place for culprit activity, phishing attacks getting smarter. Addressing these critical issues is always a open challenge in the field of biometrics today which may grow more sophisticated in near future. One can think of designing a robust and reliable multimodal biometric authentication system with the complimentary techniques such as liveness detection, speech detection with liveness measure etc. Fusion can be done either by pre classification (sensor level and feature level) or post classification (score level and decision level) levels. Feature level fusion is well practiced level than the other fusion levels, as it contains rich raw data and spoofing is literally impossible for a hacker to steal all the templates of the modalities employed. Score level fusion is relatively easy to perform as it does not deal with feature extraction or matching modules. In Decision level fusion (AND, OR) the AND rule is more effective than OR, has the hacker needs both the scoring percentage to trespass the system deliberately. Performing the discriminant analysis aiming at dimensionality reduction by the statistical means of separation between classes is done by appearance based methods such as PCA(Principal Component Analysis), LDA (Fisher Discriminant Analysis), ICA (Independent Component Analysis). Texture based methods such as LBP (Linear Binary Pattern) and its variants, Local Phase Quantization(LPQ) and Gabor also kernel based methods are quiet approachable ideas for non-linear data classification. In the multibiometric systems cancellable transformations will increase the accuracy of ensemble techniques with proper feature selection and selection of base classifier also matters, but research has to carried out in general domain rather than being specific in selection of base classifiers. Classifier ensembles are well recognized in today's pattern classification system, as said by Wolpert's 'no free lunch theorem', indicates that no single classifier is capable of attaining robust and reliable output, which could be done by segregating the decisions from the ensemble classifiers which can be implemented parallel, multithread architecture into a single framework. Extreme Learning Machine's (ELM) due to its efficacy and less complexity, it is widely used in neural networks.

The rest of the paper structure as follows: section II provides literature reviews, on different biometric modality based system. Section III emphasis on different methods used in this work. Experimental results and its analysis explored in section IV and finally section V drawn the conclusion.

2. Literature Review

Masarat et al [2]. proposed a fuzzy combination from multiple decision classifiers employing KDD 99 data set. Adopting roulette wheel algorithm for feature selection, so as to extract unique feature set from each of the decision trees. On fusion of the decisions from the classifiers, the accuracy was met upto 93% from 15 features selected. Govindarajan and Chandrasekaran [3] experimented with hybrid classifiers-radial basis function (RBF) neural network and SVM and combined the opinions of two diverse classifiers proving ensemble technique performance are better than individual classifiers. Implementing Bagging and Boosting generalized techniques they used Best First Search for feature selection on NSL-KDD data set with the 85.17%. classification accuracy. Gu et al [4] proposed developed an Intrusion Detection system from ensemble technique based on averaging the weights where the weights were generated from multi-objective genetic algorithm in reducing FAR and FRR, they used SVM classifiers and dimensionality reduction techniques such as PCA and ICA .Issues like dealing with novel intrusions, minimizing false alarm rate by the right selection of classifiers for ensembles, choosing the best base classifier and proper dominant feature extraction still require detailed research. Di Wen et.al [5] proposed face spoof detection algorithm and addressed various facial distortion features like blurriness, color moments, color diversity and specular reflection, the four different features are merged in obtaining a 121-dimensional IDA feature vector. In identifying different spoofing attacks the multiple SVM classifier is trained and applied on publicly available face spoof database Idiap REPLAY-ATTACK and CASIA FASD and the proposed method outer performs state-of-the-art methods in spoof detection. Arash Rikhtegar et.al [6] proposed a face recognition system, they employed Genetic algorithm in figuring out the optimum structure from Convolution Neural Network which extracts the dominant features in addressing illumination, occlusions and noisy conditions. The effective predictor SVM is used for classification with the error correction concepts, the final layer of CNN is ensemble with SVM in boosting the system performance. Experiments conducted on ORL and the Yale databases showed encouraging results.

Research by Dietterich and Miranda Dos Santos [7] showed by the experimental and theoretical analysis that classification error is low in Ensemble techniques having multiple base classifiers whose overall accuracy is more than 50% and those are always superior to single classifiers prediction.There are two kinds of approaches in ensemble structures namely Homogeneous (classifiers created with same technique, Eg: Boosting and Bagging) and Heterogeneous ensembles (uses multiple diverse classifiers, Eg: Stacking and Voting). Active researcher Chen et al [8] in the ensemble techniques posed several questions like, Which base components suits the ensemble model, Which is the best base classifier and how to decide it, How to segregate the decisions of the base classifiers? Folino et al [9] employed Boosting group of algorithms for intrusion detection, they used genetic programming in creating decision trees. AdaBoost.M2 was adopted in ensemble with those classifiers. KDD 99 dataset was employed and the results shown to be encouraging with Boosting algorithms. Syarif et al.[10] using both Homogeneous (Bagging and Boosting) and Heterogeneous (Stacking) ensemble approaches worked on intrusion detection and achieved 99% classification rate on NSL-KDD data set. They adopted four classification algorithms namely, Naïve Bayes, J48 (decision trees), JRip (rule induction), and IBK (nearest neighbor). However, identifying new kinds of intrusion could achieve only 60% of accuracy. Anne Magaly de.

3. Methods and Materials

3.1. Local phase quantization

LPQ extracts the phase information in the frequency domain and it is insensitive to blur. Initially the image is decorrelated and maximum information is preserved as the samples are statistically independent. STFT is applied on a small region L X L called $\boldsymbol{P_s}$ at every pixel S of an image, $\boldsymbol{F}$ ($\boldsymbol{S}$) is applied using 4 2D frequencies,

$I_1 = [a, 0]^T$, $I_2 = [0, a]^T, I_3 = [a, a]^T, I_4 = [a, -a]^T$ and $F_S = \left[Re\{F_S^C\}, Im\,\{F_S^C\}\right]^T$, $Re\{\cdot\}$ and $Im\{\cdot\}$ are real and imaginary parts of a pixel. Then in the next step, the co-efficients of F_S are decorrelated by taking co-variance matrix C, co-variance between 2 positions S_i and S_j of P_s and defined by $\sigma_{i,j} = \rho^{\|S_i - S_j\|}$ where ρ is the correlation co-effiicient. LPQ image is obtained by quantizing decorrelated co-efficient G_S , $LPQ_{image} = \sum_{h=1}^{8} qh\ 2^{h-1}$, qh is binary quantizer on h^{th} component $q_h = \begin{cases} 1 & \text{if } qh \geq 0 \\ 0 & \text{otherwise} \end{cases}$

3.2. Independent Component Analysis

Let **S** be the random vector and its elements are mixtures of $s_1, s_2 s_3 \dots \dots . s_n$ and **P** be also a random vector with elements $p_1, p_2 p_3 \dots \dots p_n$. The matrix **A** is given with the elements $\boldsymbol{a_{ij}}$ and by vector matrix notation, the mixing model is written as $\boldsymbol{S = AP} \dots \dots \dots (\mathbf{1})$**.** The above equation can also be re-written as $\boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{a_i p_i} \dots \dots \dots . (\mathbf{2})$**.** Equation (1) is called statistical ICA. The components $\boldsymbol{p_i}$ are statistically independent and must have non Gaussian distribution. On estimating the matrix **A**, inverse **'W'** can be computed in obtaining independent components by $\boldsymbol{P = WS}$**.** ICA is closely related to Blind Source Separation, where 'Source' is the original signal, 'Blind' indicates we know only less information about it.

3.3. K Nearest Neighbor

KNN is basically non parametric in nature and simple, also called as lazy learning algorithm. KNN classifier is a well-known classification algorithm that incorporates different data types by fixing up K with training data observations. KNN can be used for both classification (discrete decision making) and regression (continuous decision making). Various distance measures such as Euclidean Distance, Manhattan Distance and Minkowski Distance could be used for classification.

3.4. Support Vector Machine

SVM is a supervised learning which can be employed for both classification and regression technique, which was introduced by Vapnik, Boser, Guyon. SVM works for smaller dataset and it is said to be efficient as it works on subset of data. SVM finds the hyper plane in dividing two classes by maximizing the gap between categories.

3.5. Neural Network model

Neural network model hooks many neurons and the output of one neuron could be influenced by other neurons. Input from the input layer is fed to each node of the hidden layer and the model is trained, usually there are multiple hidden layers processing the input received from the previous layers before reaching the output neuron. In Feed forward neural network, the signal passes in single direction passing through many hidden layers.

3.6. Stacking

Stacking with probability distribution was proposed by Ting and written in 1999. Stacking is one of the ensemble machine learning technique which introduces the meta learner concept and builds model by learning and combines the predictions given by base models in a best way. Combining multiple classifiers decisions from different learning algorithms $L_1, L_2, L_3 \ldots\ldots L_N$ on single dataset, given by feature space $S_i = (x_{i,} y_i)$. In stacking, we have 2 tiers of classifiers. In first phase the primary base classifiers are learnt with certain feature space, the behavior of each classifiers is learnt at this level and the undesired behavior obtained by improper training is addressed by meta level classifiers.

3.7. Boosting

Boosting was proposed by Schapire in 1990 by boosting the weak classifier to the strongest classifier by iteratively training with the misclassified instances. Boosting generates the ensemble of classifiers. Essentially Boosting generates three classifiers, where the first classifier is given with the subset of the complete data and trained up, then the second classifier is given with the subset of data given to the first classifier, where half of the data were in agreement and the data which is misclassified is trained by the second classifier. Finally, the data which is both in disagreement with the first and second classifier is given to the third classifier in obtaining high classification rate. In 1997, Freund and Schapire proposed generalized boosting ensemble technique as Adaptive Boosting or 'AdaBoost' for short. AdaBoost produces set of hypothesis and then uses weight majority of the classes determined by hypothesis generated in combining the decisions. By this way, the preceding classifiers which is not properly learnt, the successive classifiers learns with the misclassified data and emerges to be an strong classifier by Boosting technique complementing the previous ones.

3.8. Bagging

Breiman's bootstrap aggregating, in short 'Bagging' was introduced by Breiman in 1996 [1]. In Bagging, various subsets of data are drawn from the training set which includes the data with replacement from the complete set of training data, then the classifiers are modeled using subset of training data. Some of the bootstrap samples may be cloned and some may be missing, each classifier is built from each bootstrap samples, as a result the outcome of each classifiers may differ. The decisions are done by majority voting obtained from distinct classifiers. Since each instance has the probability of $1 - (1 - 1/m)^m$ to be present in the bootstrap sample, so on an average 63.2% unique samples in bootstrap pool, while 36.8% of the original are not selected.

3.9. Recursive partition tree

Recursive partitioning is a tree based classification method applied on multivariate data analysis. Rpart basically generates a decision tree which strives to classify the feature set based on its dependent variables. In first step of rpart divide the training set into different partitioning in the form of decision trees, partition condition will be terminated using purity measures. Once tree reaches to its maximum number of node, the specified number of trees the recursion process will be terminated. Conventionally rpat is an recursive process, which converts the training data in to dynamical tree structures which helps to classify Random forest technique is not sensitive to outliers where accuracy is generated automatically and overfitting is not a problem here. Random forests are constructed from regression trees and RF are used for classification and regression purposes. Different subsets are generated from training data and remaining variables with the trained data depicts the error rate. The predictions generated by each tree is combined either by averaging or voting. Breiman's proposed randomness to the actual growing tree to find best predictor in partitioning of data in each node of the tree in a random way, so that the best way of dissimilar trees could be constructed by exhaustive search. Smaller subset would yield less predictive power and low correlation. On applying decision trees to the overall training dataset it may lead to poor performance due to overfitting of data.data.

3.10. Random Forest

Random forest technique is not sensitive to outliers where accuracy is generated automatically and overfitting is not a problem here. Random forests are constructed from regression trees and RF are used for classification and regression purposes. Different subsets are generated from training data and remaining variables with the trained data depicts the error rate. The predictions generated by each tree is combined either by averaging or voting. Breiman's proposed randomness to the actual growing tree to find best predictor in partitioning of data in each node of the tree in a random way, so that the best way of dissimilar trees could be constructed by exhaustive search. Smaller subset would yield less predictive power and low correlation. On applying decision trees to the overall training dataset it may lead to poor performance due to overfitting of data.

4. Experimental Results

In developing Unimodal Biometric identification system, We have employed AR face database and Poly-U High resolution fingerprint (FP) images and the feature extraction algorithms such as LPQ (Local Phase Quantization) and ICA1 (Independent Component Analysis). We have opted in choosing the various classifiers KNN- a non-parametric approach, SVM- Separating hyperplane classifier, Neural Networks(NN), Recursive Partition-Statistical method for multivariate analysis and Random forest (RF). The accuracy obtained by different classifiers are tabulated in Table-1, table-2, Table-3. One can understand the reliability of the developed unimodal Biometric identification system on inferring the results and it is also a comparison study among the classifiers as well as the adopted feature extraction algorithms.

TABLE I: Accuracy On Stacking Ensemble Technique Applied On Different Classification Models

Algorithms	Face_LPQ	FP_LPQ	FACE_ICA1	FP_ICA1
KNN	91.57	92.67	72.43	83.56
SVM Radial	89.63	92.02	86.65	85.57
NN	71.57	71.57	84.34	71.57
RPART	85.38	86.33	86.90	85.74
RF	88.29	90.04	86.25	86.16

Table-1 gives the results obtained by Stacking Ensemble Technique, KNN classifier with LPQ features is yielding higher rate of accuracy for face and FP modalities with 91.569226 and 92.66812 respectively when compared with the other considered classifiers. SVM and RF with LPQ features is also performing well, whereas NN with LPQ features is underperforming. RP with ICA1 features on face and FP obtained 86.9012 and 85.74437 respectively and is the best performer when compared to other classifiers on ICA features.

TABLE II: Accuracy On Boosting Ensemble Technique Applied On Different Classification Models

Algorithms	Face_LPQ	FP_LPQ	FACE_ICA1	FP_ICA1
KNN	91.50	92.84	71.57	83.43
SVM Radial	89.68	92.01	85.74	85.56
NN	71.57	71.568	83.14	71.59
RPART	85.58	86.12	85.19	85.67
RF	88.21	89.99	85.95	86.11

Table-2 tabulates the results obtained by Boosting Ensemble Technique. Again here, KNN classifier with the LPQ features on face and FP modality has obtained the higher level of accuracy 91.502 and 92.84 respectively

https://doi.org/10.15242/DiRPUB.DIR0118010

and it has proved a reliable classifier again in Boosting after the stacking ensemble technique results on LPQ features. RF with ICA1 on face and FP traits has yielded 85.95and 86.11 respectively and it is the best performer, RPART and SVM is also performing better on ICA1 features.

TABLE III: Accuracy On Bagging Ensemble Technique Applied On Different Classification Models.

Algorithms	Face_LPQ	FP_LPQ	FACE_ICA1	FP_ICA1
SVM Radial	89.46	91.96	85.68	85.51
RF	88.26	90.18	86.13	86.07

Table-3 tabulates the results obtained from Bagging Ensemble Technique. Here we have conducted experimentation considering only SVM and RF classifiers, because the remaining classifiers considered for experimentation for Stacking and Boosting Techniques showed very poor performance and hence those results are not tabulated. SVM radial on LPQ features for both the modalities has obtained good results. RF on ICA1 features is the best performer compared to SVM.

5. Conclusion

The Extensive Experimental results on ensemble techniques with different classification algorithms shows that there is no significant increase in performance with the results tabulated from all the tables, there is slight increase of accuracy in Boosting when compared with stacking. Ensemble techniques however has higher computational cost and usage of more memory space. Our experimental results convey important guidelines by tabulating comparative results on ensemble machine learning techniques, which classifier's performance is better. We have employed the best feature extraction methods LPQ and ICA1. From the results obtained, we can infer that committing more time on selection of discriminating and the best feature extraction algorithms with dimensionality reduction techniques would yield good performance in overall system recognition or verification rate and also the computation time is reduced rather than relying on more computation consuming techniques such as Stacking, Boosting and Bagging.

6. References

[1] Breiman L. Bagging predictors. Mach Learn 1996;24(2):123–40.
https://doi.org/10.1007/BF00058655
https://doi.org/10.1023/A:1018054314350

[2] Masarat S, Taheri H, Sharifian S. A novel framework, based on fuzzy ensemble of classifiers for intrusion detection systems. In: 2014 4th international conference on computer and knowledge engineering (ICCKE),. IEEE; 2014. p. 165–70.
https://doi.org/10.1109/ICCKE.2014.6993345

[3] Govindarajan M, Chandrasekaran R. Intrusion detection using an ensemble of classification methods. In: World congress on engineering and computer science, vol. 1. 2012. p. 1–6.

[4] Gu Y, Zhou B, Zhao J. PCA-ICA ensembled intrusion detection system by pareto-optimal optimization. Inform Technol J 2008;7:510–15.
https://doi.org/10.3923/itj.2008.510.515

[5] D. Wen, H. Han and A. K. Jain, "Face Spoof Detection With Image Distortion Analysis," in IEEE Transactions on Information Forensics and Security, vol. 10, no. 4, pp. 746-761, April 2015.
https://doi.org/10.1109/TIFS.2015.2400395

[6] Rikhtegar, Arash & Pooyan, Mohammad & Manzuri, M.T.. (2016). GA-OPTIMIZED STRUCTURE OF CNN FOR FACE RECOGNITION APPLICATIONS. IET Computer Vision.
https://doi.org/10.1049/iet-cvi.2015.0037

[7] Dietterich TG. Ensemble methods in machine learning. In: Multiple classifier systems. Springer; 2000. p. 1–15. https://doi.org/10.1007/3-540-45014-9_1

[8] Chen Y,Wong M-L, Li H. Applying ant colony optimization to configuring stacking ensembles for data mining. Exp SystAppl 2014;41(6):2688–702. https://doi.org/10.1016/j.eswa.2013.10.063

[9] Folino G, Pizzuti C, Spezzano G. An ensemble-based evolutionary framework for coping with distributed intrusion detection. Genet Program Evolvable Mach 2010;11(2):131–46. https://doi.org/10.1007/s10710-010-9101-6

[10] Syarif I, Zaluska E, Prugel-Bennett A,Wills G. Application of bagging, boosting and stacking to intrusion detection. In: Machine learning and data mining in pattern recognition. Springer; 2012. p. 593–602. https://doi.org/10.1007/978-3-642-31537-4_46