

Evaluation of Biometric Identification in Open Systems

Michael Gibbons, Sungsoo Yoon, Sung-Hyuk Cha, and Charles Tappert

Computer Science Department, Pace University
861 Bedford Road, Pleasantville, New York, 10570 USA
mikegibb@us.ibm.com, {scha, ctappert}@pace.edu

Abstract. This paper concerns the generalizability of biometric identification results from small-sized closed systems to larger open systems. Many researchers have claimed high identification accuracies on closed system consisting of a few hundred or thousand members. Here, we consider what happens to these closed identification systems as they are opened to non-members. We claim that these systems do not generalize well as the non-member population increases. To support this claim, we present experimental results on writer and iris biometric databases using Support Vector Machine (SVM) and Nearest Neighbor (NN) classifiers. We find that system security (1-FAR) decreases rapidly for closed systems when they are tested in open-system mode as the number of non members tested increases. We also find that, although systems can be trained for greater closed-system security using SVM rather than NN classifiers, the NN classifiers are better for generalizing to open systems due to their superior capability of rejecting non-members.

1 Introduction

Biometric applications are becoming more common and acceptable in today's society. Technology continues to improve, providing faster processors, smaller sensors and cheaper materials, all of which are contributing to reliable, affordable biometric applications. The most common use of biometrics is for verification. In biometric verification systems, a user is identified by an ID or smart card and is verified by their biometric, i.e., a person's biological or behavior characteristic such as their fingerprint, voice, iris, or signature. This is analogous to a user at an ATM machine using a bank card to identify and a PIN to verify. Another use of biometrics is for identification, which is the focus of this paper. Identification can be applied in a closed system such as employee positive identification for building access, or in an open system such as a national ID system. Positive biometric identification, a 1-to-many problem, is more challenging than verification, a 1-to-1 problem. As stated in [1], "positive identification is perhaps the most ambitious use of biometrics technology."

There have been many promising results reported for closed identification systems. Although high accuracies have been reported in writer, iris and hand geometry studies [4, 10, 12, 15, 17], these accuracies may lead to a false impression of security. One may ask if there really are any situations that correspond to closed worlds[1]. For example, take an employee identification system. Can it be guaranteed that a biomet-

ric of a guest (a non-member) visiting the facility does not match one of an employee (a member)? The answer is no. This paper will investigate the generalizability of biometric identification as it pertains to the security of a system. Our hypothesis is that the accuracies reported for closed systems are relevant only to those systems and do not generalize well to larger, open systems containing non-members.

Since it is impractical to test a true population, we use a reverse approach to support the hypothesis. We will work with a database M of m members, but assume a closed system of m' members, where $m' < m$, and train the system on the m' members. We then have $m - m'$ members to test how well the system holds up when non-members attempt to enter the system. This approach is used on two biometric databases, one consisting of writer data and the other of iris data.

In section 2 of this paper, positive identification and the associated error rates will be explained. In section 3, the biometric databases and the pattern classification techniques used in this paper will be described. In section 4, the statistical experiments to support the hypothesis are described and observations presented. Section 5 concludes with a summary and considerations for future work.

2 Error Rate Types in Biometric Identification

Consider the positive identification model. Positive identification refers to determining that a given individual is in a member database [1]. This is an m -class classification problem – that is, given data from m subjects in a biometric database $M = \{s_1, s_2, \dots, s_m\}$, the problem is to identify an unknown biometric sample d from a person, s_q , where $s_q \in M$. In this model, a classifier can be trained based on all exemplars in M to find decision boundaries, e.g., support vector machine. If a similarity-based classifier such as a nearest neighbor is used, an unknown sample d is compared to each d_i of M . The error rate in this model is simply a number of misclassified instances divided by the testing set size. We claim that a classifier with the lowest error rate is not necessarily the best for security, and that classifier designers might consider the following three types of error rates.

Consider an unknown biometric sample d from a person, s_q , where $s_q \notin M$. If this biometric data of a non member enters directly to the above model, can it be classified correctly? If the classifier has no reject capability, the unknown will be classified into one of the decision regions or as the closest matching subject. However, if the classifier has a reject capability, the number of classes in M becomes $m + 1$, i.e., m member classes + 1 reject class. Therefore, if the questioned instance is in a reject area in SVM or the closest match is outside the nearest neighbor thresholds, the unknown will be classified as none of the members. This study investigates the reject capability of two classifiers: support vector machine and nearest neighbor.

In the later scenario with members and non-members, there are three kinds of error. A ‘false reject’, FR, error occurs when a classifier identifies an unknown biomet-

ric sample d from a person, s_q , where $s_q \in M$, as a reject. The other errors are ‘false accepts’, FA, of which there are two types – those that can occur between members of the system, FA (1), and those that can occur as non-members enter the system, FA (2).

FA(1) occurs when a classifier identifies an unknown biometric sample d from a person, s_q to s_i where $s_q, s_i \in M$ and $s_q \neq s_i$. FA(2) occurs when a classifier identifies an unknown biometric sample d from a person, s_q to s_i , where $s_q \notin M$ and $s_i \in M$.

The frequencies at which the false accepts and false rejects occur are known as the False Accept Rate (FAR) and the False Reject Rate (FRR), respectively. These two error rates are used to determine the two key performance measurements of a biometric system: convenience and security [1]:

$$\begin{aligned} \text{Convenience} &= 1 - \text{FRR} \\ \text{Security} &= 1 - \text{FAR} \end{aligned} \tag{1}$$

In this paper, we will pay close attention to the Security measurement as we test our hypothesis.

3 Biometric databases and Classifiers

Two biometric databases are used to support our claims in this study: the writer and iris biometric databases. Although there are many classifiers to choose from in the field of pattern classification, we used two pattern classification techniques: Support Vector Machines (SVM) and Nearest Neighbor.

3.1 Databases

In a previous study, Cha et al. [3] studied the individuality of handwriting using a database of handwriting samples from 841 subjects’ representative of the United States population. Each subject copied a source document three times. Each document was digitized and features were extracted at the document, word, and character level. For the purposes of this study, we used the same database but focus only on the document features: entropy, threshold, number of black pixels, number of exterior contours, number of interior contours, slant, height, horizontal slope, vertical slope, negative slope, and positive slope. A detailed explanation of these features can be found in [4].

From the iris biometric image database [9], we selected 10 left bare eye samples of 52 subjects. In comparison to the writer database, the iris database has many fewer subjects, but a much larger number of samples per subject. This will allow for more samples to be trained.

After the images are acquired, they are segmented to provide a normalized rectangular sample of the iris. Features are extracted using 2-D multi-level wavelet transforms. For this experiment, 3 levels are used producing a total of 12 parts. The 12 parts produce 12 feature vectors consisting of the coefficients from the wavelet transform. The mean and variance of each vector are obtained to produce a total of 24 features for each sample. See [16] for more information on the 2-D wavelet transforms used.

3.2 Classifiers

In recent years, the SVM classifier has gained considerable popularity among the possible classifiers. The objective of the SVM classifier is to separate data with a maximal margin, which tends to result in a better generalization of the data. Generalization helps with the common classification problem of over-fitting.

The points that lie on the planes that separate the data are the support vectors. Finding the support vectors requires solving the following optimization problem (details of this method can be found in [2, 13]):

$$\begin{aligned} \min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \\ \text{subject to: } y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i > 0 \end{aligned} \quad (2)$$

The geometric representation of the SVM is easily visualized when the data falls into the linear separable and linear non-separable cases. However, real world data tends to fall into the non-linear separable case. For more information on the different cases, please refer to [11] which devotes a chapter to SVMs. To solve the non-linear separable problem, the SVM relies on pre-processing the data to represent patterns in a higher dimension than the original data set. The functions that provide the mapping to higher dimensions are known as phi functions or kernels. Common kernels include Radial Basis Function (RBF), linear, polynomial, and sigmoid. The RBF kernel is used in this study and additional information on this kernel follows in section 4.

The other classifier we consider is the Nearest Neighbor classifier, which computes distances from a test subject d to each member d_i of the database, and classifies the test subject as the subject that has the closest distance. The distances can be computed using various methods such city-block distance or Euclidean distance.

A reject threshold can be introduced into the Nearest Neighbor classification. If the distance between test subject d and its' nearest neighbor d_i is within the threshold, the classification is that of the closest member. However, if the distance is greater than the threshold, the subject is rejected and classified as a non-member. In this study, we used a reject threshold.

4 Statistical Experiments

Our hypothesis is that biometric identification on closed systems does not generalize well to larger, open systems containing non-members. In order to investigate this hypothesis, experiments were conducted on subset database $M' \subset M$ from both the writer and iris databases described in section 3.

4.1 Experiment Setup

For each of the databases, training sets were created. Training sets for the writer data consisted of $m' = 50, 100, 200$ and 400 members. Training sets for the iris data consisted of $m' = 5, 15, 25$ and 35 members. These sets included all instances per member, i.e., 3 per member for writer and 10 per member for iris.

For the first part of the experiment, an SVM was trained on the members. Parameter tuning, or SVM optimization, was performed prior to training. The first parameter tuned is the penalty parameter C from equation (2), and depending on the kernel used, there are additional parameters to tune. For this experiment we used an RBF kernel of the form:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \gamma > 0 \quad (3)$$

The γ parameter in equation (3) is the only kernel parameter requiring tuning. A grid-search method as defined in [8] was used to optimize these two parameters.

Tuning the parameters gives 100% accuracy on each of the training sets. Therefore, we have 0% FAR and FRR, or equivalently, 100% security and convenience. The next step is to test non-members to determine the true security of the trained SVM. For each training set we created a combined evaluation set consisting of the trained members plus an increasing number of non-members. The evaluation sets for the 50-writer trained SVM consisted of 50, 100, 200, 400, 700 and 841 subjects, where the first 50 subjects are the members and the remaining subjects are non-members. Similarly, the evaluation sets for the 25-iris trained SVM consisted of 25, 35, 45 and 52 subjects, where the first 25 subjects are the members and remaining subjects are non-members.

In the second part of the experiment, the Nearest Neighbor classifier was used. For this classifier, threshold tuning was required. The threshold has to be large enough to allow identification for the known members, but small enough not to allow non-members to be classified as members. As the threshold increases, the FAR increases and FRR decreases.

The Receiver Operating Characteristic (ROC) curve for the Nearest Neighbor classifier is presented in figure 1. The ROC curve is a plot of FAR against FRR for various thresholds. When designing an identification system, there is a trade off between the convenience (FRR) and security (FAR) of the system. For this experiment, we have chosen an operating threshold that is close to equal error rate, but leaning towards a higher security system.

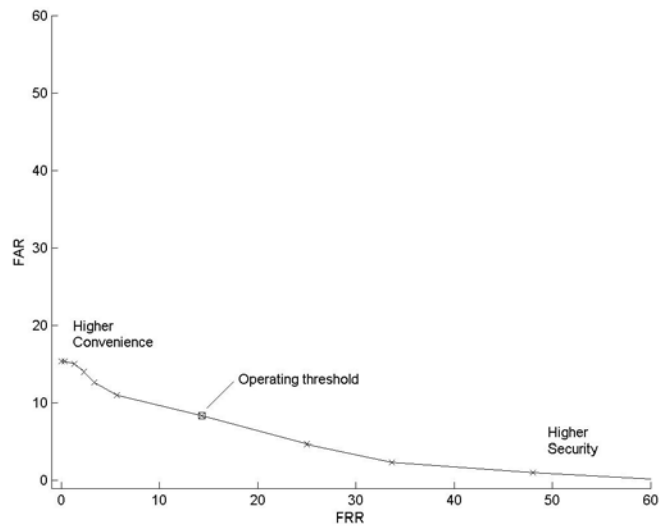


Fig. 1. The ROC curve for the Nearest Neighbor classification of 100 members. The operating threshold was chosen close to equal error rate, but favoring FAR or security

4.2 Results and Analysis

In the positive identification model we consider two errors: false accepts and false rejects. Since the SVM was able to train the members to 100% accuracy, we eliminate the false accepts and false rejects for members. The remaining tests are non-members and therefore can only produce false accepts. The false accepts correlate to the security measurement of the system, a measure of extreme importance to the system. In figure 2, the security results are shown for the writer data.

As hypothesized, for each curve, as the number of non-members increases, the security monotonically decreases (or equivalently, the FAR monotonically increases). It might also be noted that the final rates to which the security curves decrease appear to converge – that is, to approach asymptotes. To ensure that this is not an artifact of the particular handwriting data used, we obtained similar experiment results on the iris data as presented in figure 3. The iris data in figure 3 follows the same pattern as the writer data in figure 2, although convergence is not as evident for these data.

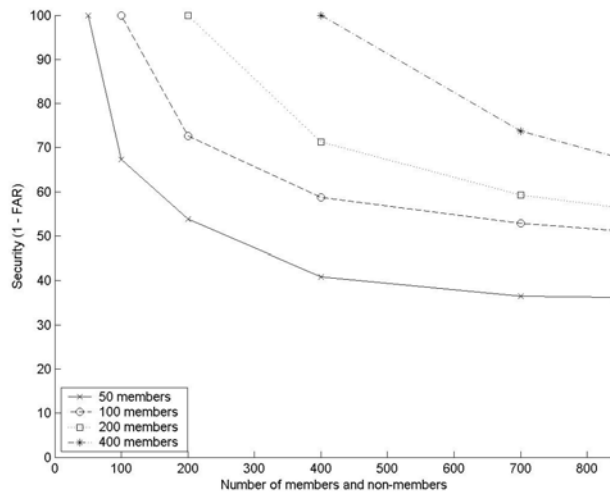


Fig. 2. Security results for writer data using SVM as non-members are introduced to the system

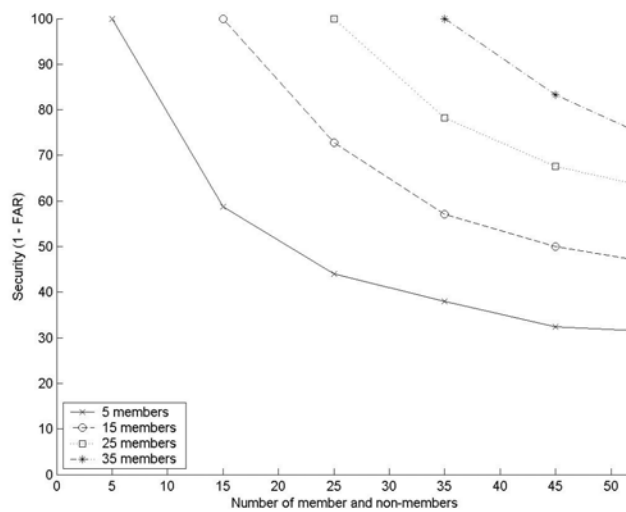


Fig. 3. Security results for iris data using SVM as non-members are introduced to the system

Next, we present the results for the Nearest Neighbor classifier. As can be seen in figure 4, the same pattern emerges, although in this experiment we did not obtain 0% FAR for the members. When using the Nearest Neighbor approach, a one-versus-all method was used to obtain the accuracy for the closed environment.

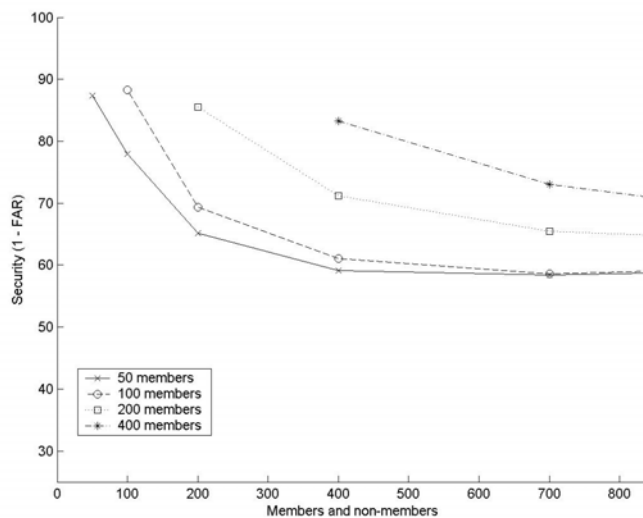


Fig. 4. Security results for writer data using Nearest Neighbor as non-members are introduced to the system.

Last, we present a comparison of the results from the two classifiers used in this experiment. Figure 5 illustrates the security performance for 100 members of the writer database. Notice, although Nearest Neighbor does not perform as well on the closed environment, it eventually meets and surpasses the performance of the SVM as non-members enter the system.

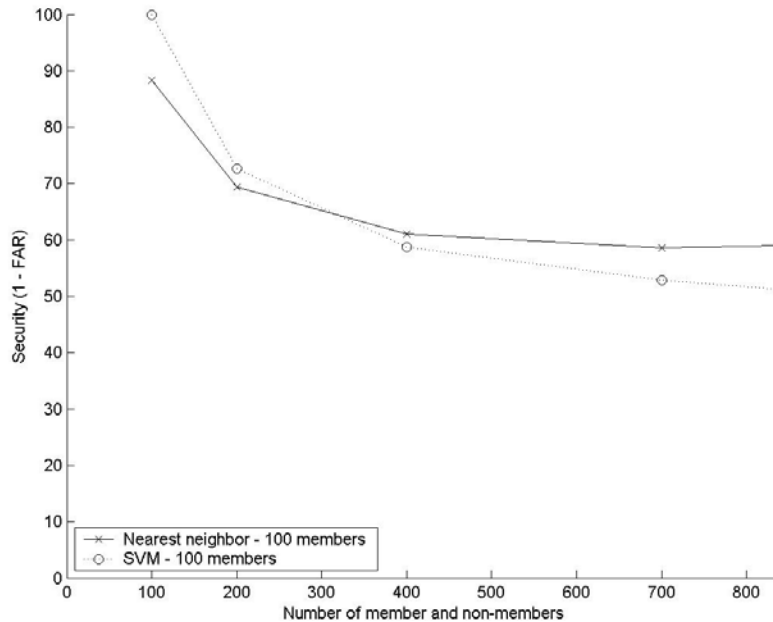


Fig. 5. A comparison of the performance of the Nearest Neighbor and SVM classifiers on the writer data consisting of 100 members

5 Conclusions

In this paper, we found that system security (1-FAR) decreases rapidly for closed systems when they are tested in open-system mode as the number of non members tested increases. Thus, the high accuracy rates often obtained for closed biometric identification problems do not appear to generalize well to the open system problem. This is important because we believe that no system can be guaranteed to remain closed. This hypothesis was validated by experiments on both writer and iris biometric databases.

We also found that, although systems can be trained for greater closed-system security using SVM rather than NN classifiers, the NN systems are better for generalizing to open systems through their capability of rejecting non members. Thus, it appears that the reject thresholds of NN classifiers do a better job of rejecting non members than the reject regions of SVM classifiers.

As commented by a reviewer of this paper, most complex biometrics systems use more complex classifiers. Given that performance on closed systems is not sufficient for applications where security is essential, we feel even the most complex classifiers should be tested in an open environment. For a more in depth analysis of various classifiers on open environments, refer to [6].

In summary, we demonstrated that the generalization capability of closed biometric systems in open environments is poor, and that the significantly larger error rates should be taken into account when designing biometric systems for positive identification. For increased security, for example, multi-modal biometrics might be considered [14].

5.1 Future work

When designing an identification system, there is a trade off between the convenience and security of the system. Most systems would choose security over convenience. However, in our implementation of SVM for the writer data, we imply choosing convenience over security (guarantee 0 false rejects). In our study on writer data, there just are not enough samples per member to put into the testing set. It would therefore be beneficial to run further experiments against larger biometric databases.

We think that it may be advantageous to develop open identification systems by using a verification model. Also, it would be beneficial to explore the features for the given biometrics since security results could improve with features that better identify a member. Additional forms of biometrics, such as fingerprint, face, voice, hand geometry or a combination of biometrics might also be tested to further test the hypothesis.

References

1. Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K., Senior, A.W.: Guide to Biometrics. Springer (2004)
2. Burges, C.: A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*. 2:121-167 (1998)
3. Cha, S.-H., Srihari, S.N.: Writer Identification: Statistical Analysis and Dichotomizer. Proceedings International Workshop on Structural and Syntactical Pattern Recognition (SSPR 2000), Alicante, Spain. pp. 123- 132 (2000)
4. Cha, S.-H., Srihari, S.N.: Handwritten Document Image Database Construction and Retrieval System. Proceedings of SPIE Document Recognition and Retrieval VIII, Vol. 4307 San Jose (2001)
5. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons (2001)
6. Gibbons, M.: On Evaluating Open Biometric Identification Systems. Master's Thesis, CSIS Department, Pace University (2005)
7. Grother, P., Phillips, P.J.: Models of Large Population Recognition Performance. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) Vol. 2. Washington, D.C. pp. 68-77 (2004)
8. Hsu, C.-W., Chang, C.-C., Len, C.-J.: A Practical Guide to Support Vector Classification.
9. Kee, G., Byun, Y., Lee, K., Lee, Y.: Improved Techniques for an Iris Recognition System with High Performance. *Lecture Notes Artificial Intelligence* (2001)
10. Krichen, E., Mellakh, M.A., Garcia-Salicetti, S., Dorizzi, B.: Iris Identification Using Wavelet Packets. *Pattern Recognition, 17th International Conference on (ICPR'04)* Vol. 4. pp. 335-338 (2004)
11. Kung, S.Y., Mak, M.W., Lin, S.H.: Biometric Authentication: A Machine Learning Approach. Pearson Education Company (2004)

12. Ma, Y., Pollick, F., Hewitt, W.T.: Using B-Spline Curves for Hand Recognition. Pattern Recognition, 17th International Conference on (ICPR'04) Vol. 3. Cambridge UK. pp. 274-277 (2004)
13. Osuna, E., Freund, R., Girosi, F.: Support Vector Machines: Training and Applications. MIT Artificial Intelligence Laboratory and Center for Biological and Computational Learning Department of Brain and Cognitive Sciences. A.I. Memo No 1602, C.B.C.L. Paper No 144 (1997)
14. Prabhakar, S., Pankanti, S., Jain, A.K.: Biometric Recognition: Security & Privacy Concerns. IEEE Security and Privacy Magazine. Vol. 1, No. 2. pp. 33-42 (2003)
15. Schlapbach, A., Bunke, H.: Off-line Handwriting Identification Using HMM Based Recognizers. Pattern Recognition, 17th International Conference on (ICPR'04) Vol. 2. Cambridge UK. pp. 654-658 (2004)
16. Woodford, B.J., Deng, D., Benwell, G.L.: A Wavelet-based Neuro-fuzzy System for Data Mining Small Image Sets.
17. Zhang, D., Kong, W.-K., You, J., Wong, M.: Online Palmprint Identification. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 25, No. 9. pp. 1041-1050 (2003)