Second-Level Partition for Estimating FAR Confidence Intervals in Biometric Systems

Rongfeng Li[1,2] , Darun Tang[1,2], Wenxin Li[1], and David Zhang[2]

[1] Key Laboratory of Machine Perception, Peking University, Beijing 100871, China
[2] Department of Computing, The Hong Kong Polytechnic University, Hong Kong
{rongfeng,tangcatcat,lwx}@pku.edu.cn, csdzhang@polyu.edu.hk

Abstract. Most biometric authentication algorithms make use of a similarity score that defines how similar two templates are according to a threshold and the accuracy of the results are expressed in terms of a False Reject Rate (FRR) or False Accept Rate (FAR) that is estimated using the training data set. A confidence interval is assigned to any claim of accuracy with 90% being commonly assumed for biometric-based authentication systems. However, these confidence intervals may not be as accurate as is presumed. In this paper, we report the results of experiments measuring the performance of the widely-used subset bootstrap approach to estimating the confidence interval of FAR. We find that the coverage of the FAR confidence intervals estimated by the subset bootstrap approach is reduced by the dependence between two similarities when they come from two individual pairs shared with a common individual. This is because subset bootstrap requires the independence of different subsets. To deal with this, we present a second-level partition to the similarity score set between different individuals, producing what we call a subset false accept rate (SFAR) bootstrap estimation. The experimental results show that the proposed procedures greatly increase the coverage of the FAR confidence intervals.

Keywords: Biometric, performance evaluation, bootstrap, confidence interval.

1 Introduction

A biometric authentication system verifies the identity of an individual by using biometric features such as palmprints, fingerprints, the face, or iris. Systems do this by matching template samples of an individual's features and accepting or rejecting them. Most biometric authentication algorithms make use of a similarity score that defines how similar two templates are. The score makes use of a threshold but the setting of that threshold still allows systems to erroneously either accept or reject a particular match. The accuracy of a biometric authentication system, expressed in terms of these two error rates, the False Reject Rate (FRR) and False Accept Rate (FAR) [1][2][4][5][7], is estimated using the training data set and a confidence interval is assigned to any claim of accuracy with a confidence interval of 90% being commonly assumed for biometric-based authentication systems. We use the statistics coverage to evaluate how much real-world FRR/FAR can be cover with the estimated FRR/FAR confidence intervals.

X. Jiang and N. Petkov (Eds.): CAIP 2009, LNCS 5702, pp. 58–65, 2009.

Most methods for estimating error rate confidence intervals are either parametric or non-parametric. The parametric methods are based on the probability distribution of the similarity score between two patterns. Various such methods [1][10][11][12][13] assume the score distributions conform to an i.i.d. Gaussian distribution but in practical applications the distribution of the score is often unknown. In contrast, non-parametric methods do not require knowledge of distribution of the similarity score. The most widely used of the non-parametric methods is bootstrap, which has been shown to be robust when handling unknown similarity score distributions [6]. A thorough description of bootstrap methods can be found in [3]. Bolle et al. [6][8] applied bootstrap to estimating FRR and FAR confidence intervals and found it superior to the parametric methods. The bootstrap method still requires the i.i.d. score distribution while the dependence between similarities which come from a same pair of individual will lead to the unconformity of i.i.d.. Subset bootstrap [9] partition the dataset into subset by the individual pairs so that it can handle this kind of dependency between similarities scores of the same individual pairs. Two-level bootstrap [14] improved on subset bootstrap in that it allowed faster convergence. However, there is another kind of dependence between two similarities when they come from two individual pairs which share a common individual. For example we have A,B,C three persons. The similarity between A, B and the similarity between A,C are dependent because (A,B) and (A,C) share the common person A. This kind of dependence can not be avoided by both of the subset bootstrap approach and the two-level bootstrap. This reduces the coverage of FAR confidence interval, for the subset bootstrap required the independence of different subsets as we know in [9].

In this work, we first give brief introduction to the subset bootstrap method We then propose a second-level partition to the individual pairs which separate the dependent individual pairs. This then allows us to estimate the FAR confidence intervals using what we refer to as Subset FAR (SFAR). Finally, we compared the proposed methods against subset bootstrap, using Recognition Algorithm Testing Engine (RATE) [15], which is an online testing system for the biometric algorithms, and found that the proposed procedures greatly increase the coverage of the FAR confidence intervals.

The rest of this paper is organized as follows: Section 2 formulates the problem of establishing a suitable confidence interval of the FAR and provides a formulation for the problem of subset bootstrap estimation [6]. Section 3 presents a second-level partition to the individual pairs and proposes the use of Subset FAR (SFAR) to estimate the FAR confidence intervals. Section 4 provides the experimental results of second-level partition and SFAR approach for FAR confidence interval estimation. Section 5 offers our Conclusion.

2 Subset Bootstrap Estimation

2.1 Preliminaries

Most biometric authentication algorithms define a similarity score s which is used in deciding how similar two templates are. A threshold t is specified to decide the rejection and acceptance which will cause the False Reject Rate (FRR) and False Accept Rate

(FAR). FRR and FAR are expressions of statistical random variables for each threshold t. Real-world applications use the training data set to estimate these random variables and thereby describe the accuracy of the biometric authentication system. As 100% accuracy of the FRR and FAR is impossible, claims of accuracy are set within a (1-α)100% confidence interval. The (1-α)100% confidence level FRR confidence interval [FRR_u(t), FRR_d(t)] and FAR confidence interval [FAR_u (t), FAR_d(t)] are defined as follows:

$$P(FRR_u(t)<FRR(t)<FRR_d(t))>1-\alpha \tag{1}$$

$$P(FAR_u(t)<FAR(t)<FAR_d(t))>1-\alpha \tag{2}$$

α is the probability that the real-world FRR/FAR are outside the estimated interval.

2.2 Subset Bootstrap Estimation

Subset bootstrap estimation [9] can be formulated as follows.

Suppose we have N individuals and we acquire d sample templates from each individual, giving Nd templates. These generate $Nd(d-1)/2$ self-similar match scores (a comparison of an individual's own templates) and $N(N-1)d^2/2$ mismatch scores between the templates of different individuals.

We denotes the collection of mismatch scores between different individuals by S and partition S into N(N-1)/2 subset according to different individual index pairs. That is

$$S = S[1] \cup S[2] \cup ... \cup S[M], \qquad M = N(N\text{-}1)/2 \tag{3}$$

The Subset Bootstrap estimation of the FAR confidence interval in is described as following steps:

Step1) Divide the mismatch scores set S into M subset S[i], i=1,2,…,M according different individual index pairs so that M=N(N-1)/2

Step2) Do B times (k=1 to B):

i) Generate random integer array $r_1, r_2, ..., r_M$ with replacement from {1,2,…,M}

ii) Generate the bootstrap resample set $S_k = \bigcup_{i=1}^{M} S[r_i]$

iii) Calculate the $FAR_k(t)$ using the equations:

$$FAR_k(t) = \frac{1}{d^2 N(N-1)/2} \sum_{s \in S_k} I\{s < t\} \tag{4}$$

iv) Sort the B bootstrap estimates $FAR_k(t)$ k=1,2,…,B by

$$FAR_1^*(t) \le FAR_2^*(t) \le ... \le FAR_B^*(t)$$

Step 3) Eliminate the bottom α/2 and the top α/2 of the B bootstrap estimations. The margin of the leftover estimations gives the (1-α)100% confidence intervals. That is $[FAR^*_{\lceil(\alpha/2)B\rceil}(t), FAR^*_{\lfloor(1-\alpha/2)B\rfloor}(t)]$

3 Second-Level Partition and SFAR Estimation

The subset bootstrap FAR confidence interval estimation partition the mismatch score set to N(N-1)/2 subsets and assuming they are independent. However, if we denote that S[i,j] is the similarity score set of two different individual i,j. for three different individuals i, j, k, we cannot assume S[i,j] and S[j,k] are independent because they have a common individual j.

In this section we will present a second-level partition to S so that we can handle the second-level subsets separately to avoid the dependence.

Suppose m is a integer which can divide N(N-1)/2 exactly , K=N(N-1)/2m.

The second-level partition is as follows:

$$S^{(k)} = S[i_1^{(k)}, j_1^{(k)}] \cup S[i_1^{(k)}, j_1^{(k)}] \cup \ldots \cup S[i_m^{(k)}, j_m^{(k)}] \tag{5}$$

$$S = \bigcup_{k=1}^{K} S^{(k)} \tag{6}$$

Here, k=1,2,…,K.

For each k=1,2,…, K, if the index $i_1^{(k)}, j_1^{(k)}, i_2^{(k)}, j_2^{(k)}, \ldots, i_m^{(k)}, j_m^{(k)}$ are exactly 2m different integers, we call the partition (6) an independent partition. We indicate Subset False Accept Rate(SFAR) for each subset $S^{(k)}$:

$$SFAR^{(k)}(t) = P(s > t \mid s \in S^{(k)}) \tag{7}$$

Notice that

$$FAR(t) = P(s > t \mid s \in S\ ) = \sum_{k=1}^{K} P(S > t \mid s \in S^{(k)}) P(s \in S^{(k)} \mid s \in S) \quad \text{Thus,}$$

c (8)

Then, we can estimate the FAR confidence interval by the K SFAR confidence intervals.

For each k=1,2,…, K, the subset bootstrap estimation in Section 2 can work independently according to the partition (3). Thus, we have K SFAR bootstrap 90% confidence intervals: $[SFAR^{(k)}{}_d(t), SFAR^{(k)}{}_u(t)]$, k=1,2,…,K

Let

$$FAR_d(t) = \frac{1}{K} \sum_{k=1}^{K} SFAR^{(k)}{}_d(t) \tag{9}$$

$$FAR_u(t) = \frac{1}{K}\sum_{k=1}^{K} SFAR^{(k)}{}_u(t) \tag{10}$$

Then $[FAR_d(t), FAR_u(t)]$ is the 90% confidence interval of FAR(t).

The steps of the second-level partition and SFAR estimation can be illustrated in Fig. 1.

When implementing the second-level partition (6), the immediate question is how to partition it independently such that: $i_1^{(k)}, j_1^{(k)}, i_2^{(k)}, j_2^{(k)}, \ldots, i_m^{(k)}, j_m^{(k)}$ are 2m different integers. One way is to use the following partition:

1. if N is an odd number: $S^{(k)} = \bigcup_{\substack{i+j\equiv k-1 (\bmod N)\\ 1\le i\le j\le N}} S[i,j]$, m=(N-1)/2

2. if N is an even number: $S^{(k)} = \left(\bigcup_{\substack{i+j\equiv k (\bmod N\text{-}1)\\ 1\le i< j\le N}} S[i,j]\right) \cup \left(\bigcup_{\substack{2i\equiv k (\bmod N\text{-}1)\\ 1\le i\le N}} S[i,N]\right)$, m=N/2

With this method, we can avoid the dependence between two similarities when they come from two individual pairs which share a common individual. The coverage is higher than 90% when the training data size is above 100, as shown in Section 4.

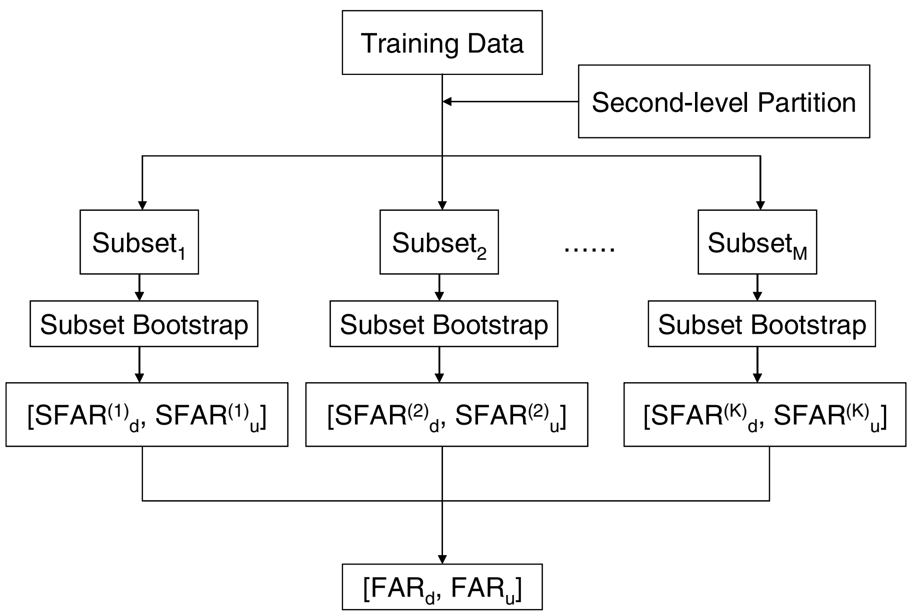

Fig. 1. Steps of the second-level partition and SFAR estimation

4 Experimental Results

Recognition Algorithm Testing Engine (RATE) [15] is an online performance evaluation system for pattern recognition algorithms. It was developed by the AI-lab of Peking University [16] and offers three different palmprint databases DATA1, DATA2, DATA3, allowing the users to submit their biometric algorithm and see the testing results. DATA1 were acquired from the AI-lab [16] and DATA2 and DATA3 were acquired from Biometric Centre of the Hong Kong Polytechnic University [17]. Images of the palmprint in these databases, which are the central parts extracted from the original images, are 128×128 and 200dpi. Table 1 shows the number of individuals and templates in the databases.

Table 1. Number of individuals and templates in the databases

	DATA0	DATA1	DATA2
N	51	213	261
d	15	4	10

N denotes the number of individuals
d denotes the number of templates for each individual

We submit 4 different palmprint recognition algorithms “fft”, ”surface”, ”texture” and “wavelet” developed by Li [18][19][20]. We choose the 40% sample size from the dataset as a training set to estimate the confidence interval. It is repeated 100 times. To calculate the coverage, we use the method described in Section 2, choosing 80% sample size as a testing set and including the training set. Tables 2 shows the coverage of the FAR confidence interval using the subset bootstrap method described in Section 2.

Table 2. Coverage of the FAR confidence interval using the subset bootstrap

	DATA1	DATA2	DATA3
fft	69%	18%	7%
surface	48%	11%	6%
texture	24%	51%	1%
wavelet	49%	21%	15%

We can see the coverage decreases while the sample size increases. This is probably because the dependence between two similarities when they come from two individual pairs which share a common individual increasing along with the sample size.

Tables 3 shows the results for FAR confidence interval estimated when using second-level partition and SFAR estimation, as described in Section 3.

The coverage of the FAR confidence interval estimated using SFAR bootstrap is much higher than when using subset bootstrap. The coverage was more than 90% on DATA3, which has a training set larger than 100.

Table 3. Coverage of the FAR confidence interval using second-level partition and SFAR estimation

	DATA1	DATA2	DATA3
fft	21%	60%	99%
surface	100%	77%	92%
texture	92%	99%	100%
wavelet	84%	96%	90%

5 Conclusion

In this paper, we introduced a second-level partition to the mismatch similarity score set and present a SFAR bootstrap confidence interval estimation that allows us to avoid dependence between two similarities when they come from two individual pairs which shared a common individual. Experimental results show that applying the second-level partition greatly improves the coverage of confidence intervals when compared with conventional subset bootstrap estimation.

In fact, our approach takes the FAR as the average of SFARs. But it is doubtful that the average of the SFARs confidence interval will automatically generate the confidence interval of FAR. An issue for future work is to demonstrate the average of the SFAR confidence intervals will generate the FAR confidence interval.

The experiment in this paper is based on palmprint database and with small sample size. Thus another future work is to extend this work to other biometric databases and larger sample sizes.

References

[1] Shen, W., Surette, M., Khanna, R.: Evaluation of Automated Biometrics-Based Identification and Verification Systems. Proceedings of the IEEE 85(9), 1464–1478 (1997)

[2] Golfarelli, M., Maio, D., Maltoni, D.: On the Error-Reject Trade-Off in Biometric Verification Systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7) (July 1997)

[3] Politis, D.N.: Computer-intensive methods in statistical analysis. IEEE Signal Processing Magazine 15(1), 39–55 (1998)

[4] Wayman, J.L.: Error rate equations for the general biometric system. IEEE Robotics & Automation Magazine 6(1), 35–48 (1999)

[5] Phillips, P.J., Martin, A., Wilson, C.L., Przybocki, M.: An Introduction to Evaluating Biometric Systems. Computer 33(2), 56–63 (2000)

[6] Bolle, R.M., Pankanti, S., Ratha, N.K.: Evaluation techniques for biometrics-based authentication systems (FRR). In: 15th International Conference on Pattern Recognition, Proceedings, September 2000, vol. 2(3-7), pp. 831–837 (2000)

[7] Dahel, S.K., Xiao, Q.: Accuracy performance analysis of multimodal biometrics. In: Information Assurance Workshop, IEEE Systems, Man and Cybernetics Society, June 18-20, pp. 170–173 (2003)

[8] Bolle, R.M., Ratha, N.K., Pankanti, S.: IBM Thomas J. Watson Research Center. In: Proceedings of the 17th International Conference on An Evaluation of Error Confidence Interval Estimation Methods, August 23-26, vol. 3, pp. 103–106 (2004)

[9] Bolle, R.M., Ratha, N.K., Pankanti, S.: Error analysis of pattern recognition systems—the subsets bootstrap. Computer Vision and Image Understanding 93(1), 1–33 (2004)

[10] Schmid, N.A., O'Sullivan, J.A.: Performance Prediction Methodology for Biometric Systems Using a Large Deviations Approach. IEEE Transactions on Signal Processing 52(10) (October 20, 2004)

[11] Gamassi, M., Lazzaroni, M., Misino, M., Piuri, V., Sana, D., Scotti, F.: Quality Assessment of Biometric Systems: A Comprehensive Perspective Based on Accuracy and Performance Measurement. IEEE Transactions on Instrumentation and Measurement 54(4) (August 2005)

[12] Snelick, R., Uludag, U., Mink, A., Indovina, M., Jain, A.: Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(3), 450–455 (2005)

[13] Hube, J.P.: Using Biometric Verification to Estimate Identification Performance. In: Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research, pp. 1–6 (2006)

[14] Poh, N., Bengio, S.: Estimating the Confidence Interval of Expected Performance Curve in Biometric Authentication using Joint Bootstrap. In: IEEE International Conference on Acoustics, Speech and Signal Processing, April 15-20, vol. 2, pp. II-137 – II-140 (2007)

[15] `http://rate.pku.edu.cn`

[16] `http://ai.pku.edu.cn`

[17] `http://www.comp.polyu.edu.hk`

[18] Li, W., Zhang, D., Xu, Z.Q.: Palmprint Recognition by Fourier Transform. Journal of Software 5(13), 879–886 (2002)

[19] Li, W., You, J., Zhang, D.: A Texture-based Approach to Palmprint Retrieval for Personal Identification. Accepted by IEEE Trans. on Multimedia (2003)

[20] Kong, W.K., Zhang, D., Li, W.: Palmprint Feature Extraction Using 2-D Gabor Filters. Pattern Recognition 36(10), 2339–2347 (2003)