Studies of Biometric Fusion
Appendix D

Fusion by Logistic Regression

William Fellner,[1] Brad Ulery,[1] Austin Hicklin,[1] Craig Watson[2]
[1] Mitretek Systems
[2] National Institute of Standards and Technology

20 July 2006

Abstract

This report discusses the use of logistic regression as a technique for developing optimal classification rules for multi-instance and multi-modal biometric fusion, using fingerprint and face data.

Contents

1 Introduction

This paper explores the use of *logistic regression* as a technique for developing optimal classification rules. The purpose of the paper is at least partly expository. The logistic technique is by no means new, but appears to be underutilized in the biometrics community. This may be due, in part, to the association of regression with classifiers that are linear in the raw scores. That this need not be the case is shown conceptually and by example.

The paper is organized as follows. The relationship between logistic regression and the likelihood ratio is developed. The relationship is illustrated for single scores, using the data described above. Then alternatives to fusion of scores are discussed. Finally, fusion using logistic regression is illustrated in the context of relating TAR to the number of pairs of images fused.

2 Key Statistical Concepts

The early part of the paper contains a number of short sections to highlight the statistical concepts that are being brought together.

2.1 Likelihood

Consider the population of mates, and let X be the score of a random case (pair of images) from that population. The scores will be taken to have a probability density $f(x|T)$, where T is used to denote a *true* mate. For a score from that population whose value is x, $f(x|T)$ is the *likelihood* of the score. Similarly, the density of the scores from the population of non-mates will be denoted by $f(x|F)$.

2.2 Likelihood Ratio and the Neyman-Pearson Fundamental Lemma

The Neyman-Pearson Lemma specifies that the optimal classification of a case, based on the score x, is to classify the case as a mate if the *likelihood ratio* $f(x|T)/f(x|F)$ is sufficiently large. The optimum cut-off point for the ratio is determined by the relative sizes of the two populations (i.e. the *prior probabilities*), as well as the relative costs of misclassification. The ratio itself is independent of these.

2.3 Maximum Likelihood Estimation

In general, the numerator and denominator of the likelihood ratio are unknown, and have to be estimated from a training sample, in which not only the scores are known, but the correct classification as well.

One approach to estimation is to suppose that $f(x|T)$ is known up to a vector of parameters θ. Then, the *maximum likelihood* estimate of θ maximizes the joint likelihood of the sample of mates:

$$L(\theta) = \prod_{mates} f(x|T),$$

where f depends on θ as well as x.

As an example, f might be the Gaussian density, with θ representing the mean and standard deviation.

Exactly the same approach would apply to non-mates, except that the functional form of f and the vector of parameters would be different.

2.4 The Kernel Method of Estimation

One problem with the maximum likelihood approach is that expressing f as a known function up to a small set of unknown parameter values isn't always realistic. The data are often not consistent with a simple model like the Gaussian density. An alternative approach is to estimate f as the sum of *kernel* functions, one for each item in the sample, centered on the score for that item. Letting y denote the value of a score in the sample of mates, and x the argument of f, the estimate of $f(x|T)$ takes the form

$$\hat{f}(x|T) = c^{-1} \sum_{y \in \text{mates}} k\left(\frac{x-y}{c}\right).$$

The choice of kernel is not usually critical, and the Gaussian kernel is often used. The width constant c is selected to give a reasonably smooth result. Of course, the estimation of $f(x|F)$ would be constructed similarly, possibly with a different value for c.

2.5 Logistic Regression

It is not necessary to estimate the densities of the scores of mates and non-mates separately. Only the ratio of the two is required. This allows the use of simpler techniques. The *logistic regression* method estimates the logarithm of the ratio of the densities as

$$\log \frac{f(x|T)}{f(x|F)} = \theta_0 + \theta_1 h_1(x) + \theta_2 h_2(x) + \Lambda \ , \qquad (1)$$

where the *carriers h* are known functions of x. Often the carriers will be powers of x, but it will also be useful to include one or more carriers that capture whether or not x is equal to (or greater than, or less than) a particular value. An example is

$$h(x) = \begin{cases} 1(x=0) \\ 0(x \neq 0) \end{cases}. \qquad (2)$$

As before, the θ are a vector of unknown parameter values, to be estimated from the training data by the principal of maximum likelihood.

2.6 Prior and Posterior Probabilities

To see how maximum likelihood estimation is constructed in the context of logistic regression, consider a single population in which some members are mates, the rest non-mates. For this population, the following is an identity

$$\frac{f(x|T)}{f(x|F)} \equiv \frac{P(F)}{P(T)}\frac{P(T|x)}{P(F|x)} \equiv \frac{1-P(T)}{P(T)}\frac{P(T|x)}{1-P(T|x)}. \quad (3)$$

Here $P(T)$ is the *prior probability* that a member of the combined population is a mate, and $P(T|x)$ is the *posterior probability* given that the selected member has score x. Analogous definitions apply to the non-mates.

Combining Equations (1) and (3) and taking logarithms gives

$$\log\frac{P(T|x)}{1-P(T|x)} = q(x;\theta) = \theta_0' + \theta_1 h_1(x) + \theta_2 h_2(x) + \Lambda \,, \qquad (4)$$

where θ_0' differs from θ_0 in Equation (1) by the quantity $\log\frac{1-P(T)}{P(T)}$. The other values of θ are unchanged. The left-hand side of Equation (4) is commonly called the *log odds* – in this case, that the item with score x is a mate.

If one regards the sample pairs as having been drawn from a single population with prior probability of a mate $P(T)$, then the maximum likelihood estimate of the vector θ maximizes the quantity

$$L(\theta) = \prod_{\text{mates}} P(T|x) \prod_{\text{non-mates}} \big(1-P(T|x)\big). \quad (5)$$

or, equivalently,

$$\log L(\theta) = \sum_{\text{mates}} \log P(T|x) + \sum_{\text{non-mates}} \log\big(1-P(T|x)\big). \qquad (6)$$

The maximizing values of θ are obtained by numerical optimization.

The relative proportions of mates and non-mates in the combined sample have no direct effect on the estimates of θ, except for the overall constant term θ_0'. This term includes the logarithm of the ratio of the prior probabilities, which is estimated from the corresponding ratio in the sample.

Generally, the proportions of mates and non-mates in the population will be quite different from the relative numbers of mates and non-mates in the combined sample. In the examples considered here, we used the weighted likelihood

$$L_W = \left\{\prod_{\text{mates}} P(T|x)\right\}^n \left\{\prod_{\text{non-mates}} \big(1-P(T|x)\big)\right\}^m , \qquad (7)$$

where n is the number of non-mates in the training samples, and m is the number of mates in the training sample. This weighting is equivalent to specifying the prior probability of a mate to be 0.5, so that θ_0 in Equation (1) and θ_0' in Equation (4) are the same quantity.

The only effect of this choice is to facilitate comparison of the ratio of the densities to the logistic regression model, by plotting each as a function of x. If the two approaches gave the same result, then the curves would coincide. (Otherwise they would merely be parallel.)

Logistic regression has some advantages over the direct estimation of the densities of the mates and non-mates. Assuming the correctness of Equation (1), which depends on the selection of carriers, both are derived from the Neyman-Pearson lemma. A second advantage is parsimony. The classifier can be described with a short list of parameter values.

Verlinde et al [Verlinde-00] also cite these reasons for strongly preferring logistic regression to the other techniques they consider.

As noted earlier, the densities for the populations of mates and non-mates are not needed individually. However, logistic regression is not the only approach to modeling the ratio of the densities. Griffin [Griffin-05] models the log(FRR) as a polynomial in log(FAR), then inverting the polynomial as part of the steps to obtaining an estimate of the density ratio. Since polynomials are not generally monotonic, he takes steps to ensure invertibility, such as controlling the degree of the polynomial and restricting its range.

Other approaches are best discussed in the context of fusion (see Section 4).

Equation (3) can be solved for $P(T|x)$, giving

$$P(T|x) = \frac{e^{q(x;\theta)}}{1+e^{q(x;\theta)}}. \qquad (8)$$

The right-hand side of Equation (8), as a function of q, is commonly called the *logistic function*. Hence the name of the regression technique.

Note that for both density ratio and for logistic regression, prior probabilities have no effect on the ROC curve. Thus they only determine the optimum threshold. Determining the optimum threshold requires, in addition to the priors, the relative costs of the two types of misclassification. These are outside the scope of our work.

2.7 Spikes

The distributions of scores, for most of the matchers, include a small number of specific values that occur frequently. Often these are the upper and lower bounds of the scores. For example, a matcher might produce scores within the range 0 to 1. For various reasons, such as the discreteness of the value reported, or in order to save computation time, a clear non-mate might be reported as an exact 0, a clear mate reported as an exact 1. Thus the training data will include a large number of 0's and 1's.

As suggested earlier, under logistic regression, spikes can be modeled using carriers such as Equation (2). For the present data, such a carrier for the upper bound is not needed, since the upper bound value rarely occurs among the known non-mates. The coefficient for such a term would not be estimable. On the other hand, lower bound values occur frequently among both the mates and the non-mates, so terms for these need to be included.

For the kernel method of density estimation, spikes need to be excluded from the data. When forming the ratio of the densities, for spike values, this is the ratio of the proportion among the mates to the proportion among the non-mates. For non-spike values, the ratio of the densities must be adjusted for the relative proportions of non-spike values among the mates and the non-mates.

3 Single-Score Examples

The NDBF06 dataset (described in Part III) consists of 186,867 paired sets of fingerprint and face images. Of these, 64,867 pairs were known mates, 122,000 were known non-mates. There were 3 face matchers (A, B, C) and 3 fingerprint matchers (H, I, Q). For the fingerprint matchers, the left index finger was selected arbitrarily.

The goal here was to make a visual comparison between the log odds obtained from the logarithm of the density ratio (using the kernel method), and the log odds obtained from logistic regression. The carriers used for the latter include the scores raised to various powers, plus one or two carriers in the form of Equation (2) to represent the spikes (though never the upper bound, for the reason given earlier).

The software package JMP[1] is used as a source for the kernel method, and for logistic regression. For the kernel method, JMP's default kernel width is used. Spikes are appropriately handled, as discussed in Section 2.7..

Figure 1 through Figure 6 give the results for Matchers A, B, C, H, I, and Q, respectively. Each figure shows the estimated log of the density ratio, and one or two logistic regression fits. Each of the fits includes the required spike terms. For the face matchers, one of the fits is always a 2nd-order polynomial. A second fit is included, if the addition of higher-order terms appears to improve the fit. For fingerprints, the first fit is always a straight line.

The vertical line in each plot locates the score corresponding to FAR=0.0001, based on the higher-order logistic regression fit. The corresponding value of TAR is also shown.

It appears that the logistic regression tracks the log of the density ratio more closely for the face matchers than for the finger matchers. This happens because fingerprint matchers are far more capable of separating mates from non-mates, as is apparent from the reported TAR values. Therefore, there is less overlap between the data from the mates and the non-mates. It is the overlapping portion of the data that provide the information for estimating both the density ratio and the parameters of the logistic regression.

For similar reasons, the log of the density ratio estimates, themselves, are visibly more variable at the extreme score values, where either the mates or the non-mates are sparse, and the log odds correspondingly difficult to estimate. As compensation, it is necessary to smooth these areas, as was done in other parts of this report, but not here.

For the face matchers, it appears that some further improvement might be obtained by adding a term that captures whether the score is greater than a specified threshold, even though that threshold is not a spike, per se. For example, for Matcher A, one might include

---

[1] JMP is "statistical discovery software" from SAS: http://www.jmp.com/

$$h(x) = \begin{cases} 1(x > 0.73) \\ 0(x \leq 0.73) \end{cases}, \quad (9)$$

since the density ratio rises so rapidly after that value. Such a possibility will not be pursued here.

In addition to the spike at 0, Matcher Q produces a spike at –1. Since the matchers are "black boxes", the reason for the extra spike is not known. Possibly, the –1 score is meant to communicate that the matcher knows that it cannot make a reliable classification.

It is notoriously perilous to extrapolate polynomial fits outside the range of the data used to obtain these fits. Following Griffin [Griffin-05] one should set boundaries so that any value below the low boundary is classified as a non-mate, and any value above the high boundary is classified as a mate.

One thing that should be apparent from the above discussion is that the form of the logistic model used for logistic regression should not be done blindly. In the examples, the models were deliberately selected to match the density ratios from the kernel method. Simpler approaches, such as histograms of the scores of mates and non-mates will often be more accessible. Inevitably, some trial and error with different models will also be required.

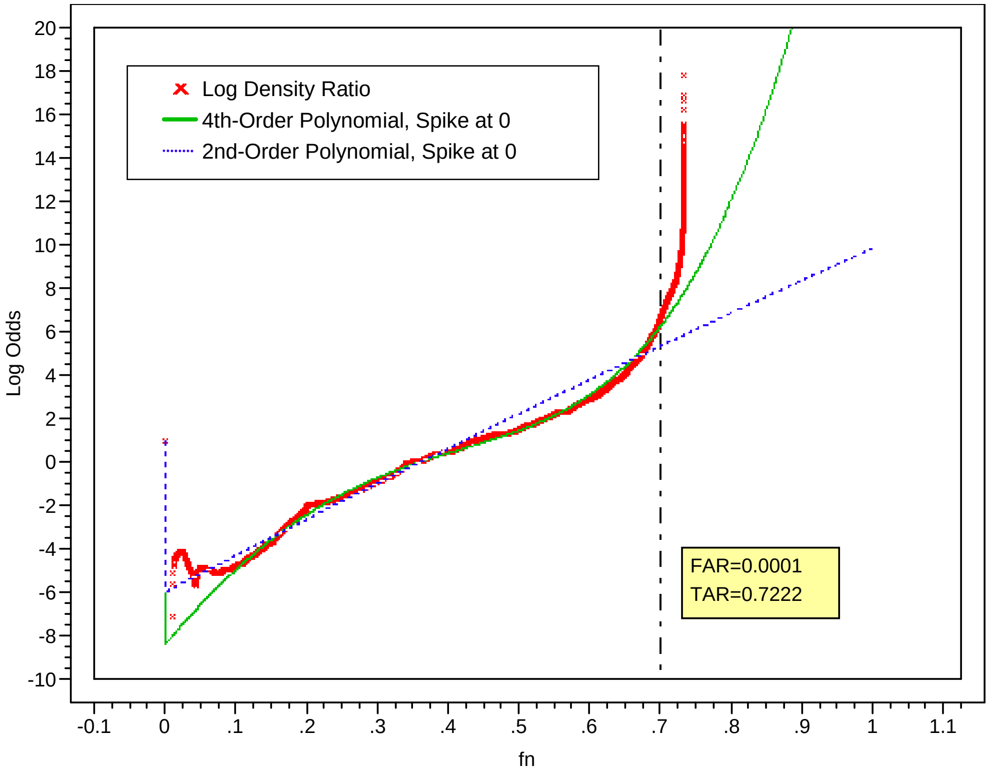

Figure 1: Matcher A: Log Odds (of a Mate) by Face Score (fn)

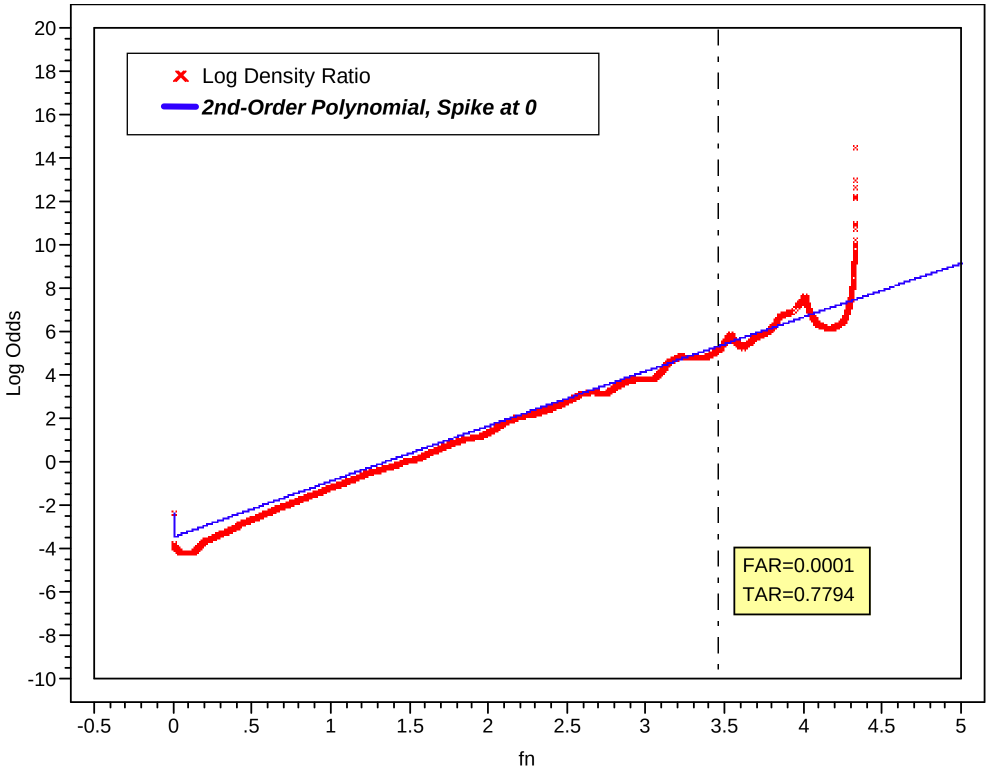

Figure 2: Matcher B: Log Odds (of a Mate) by Face Score (fn)

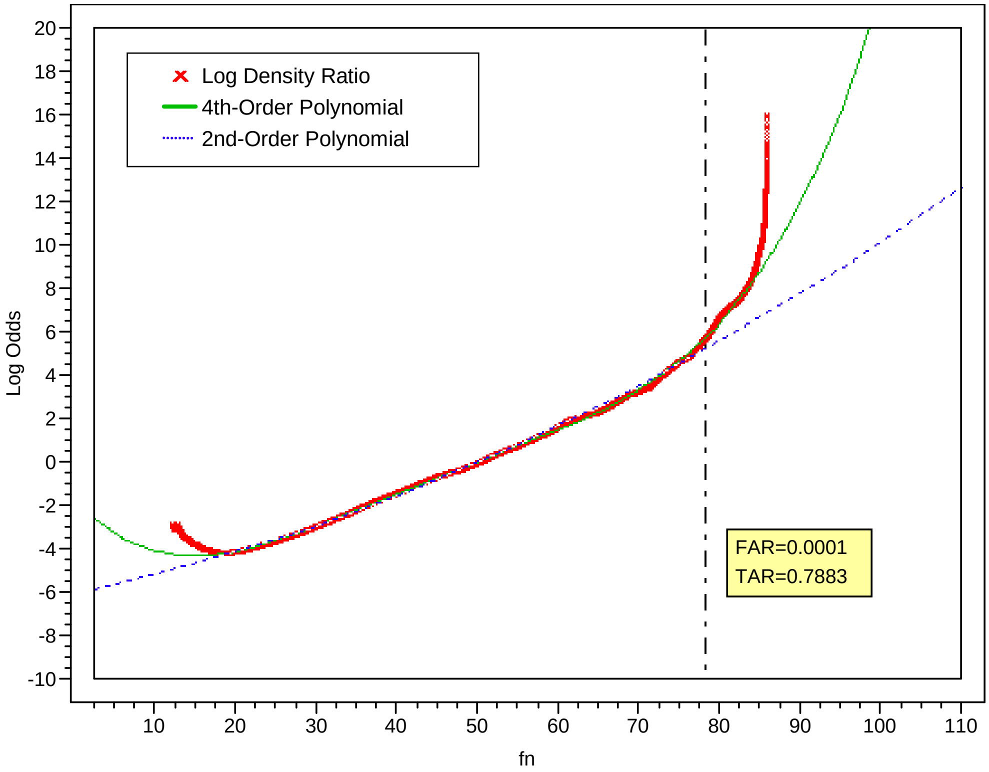

Figure 3: Matcher C: Log Odds (of a Mate) by Face Score (fn)

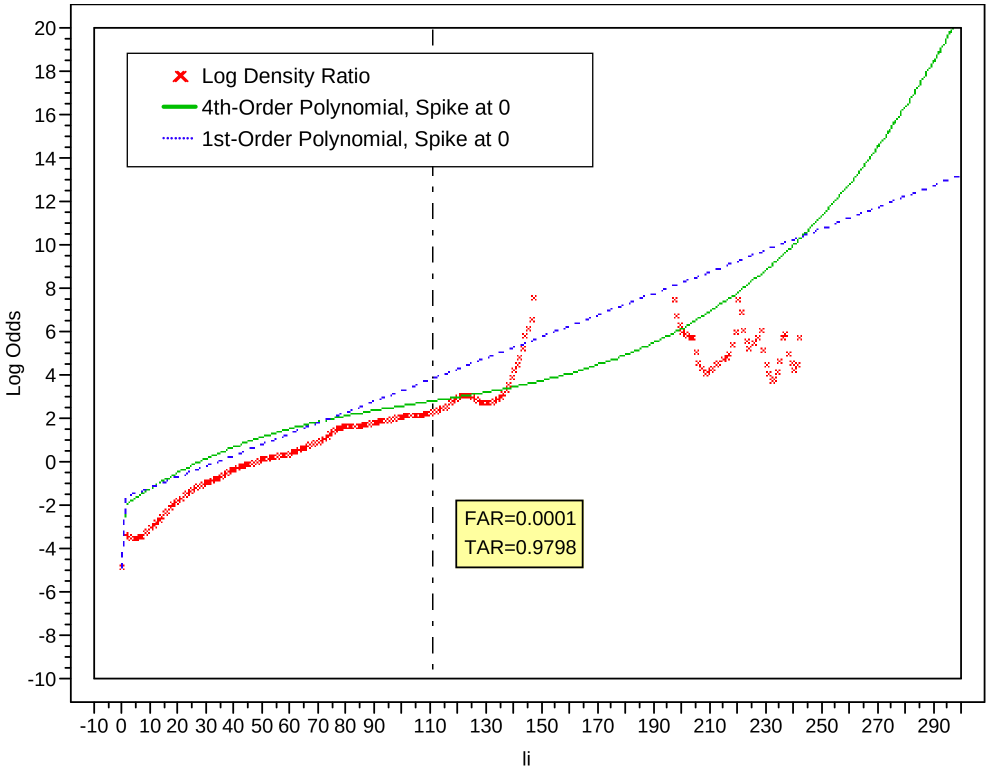

Figure 4: Matcher H (Left Index Finger): Log Odds (of a Mate) by Finger Score (li)

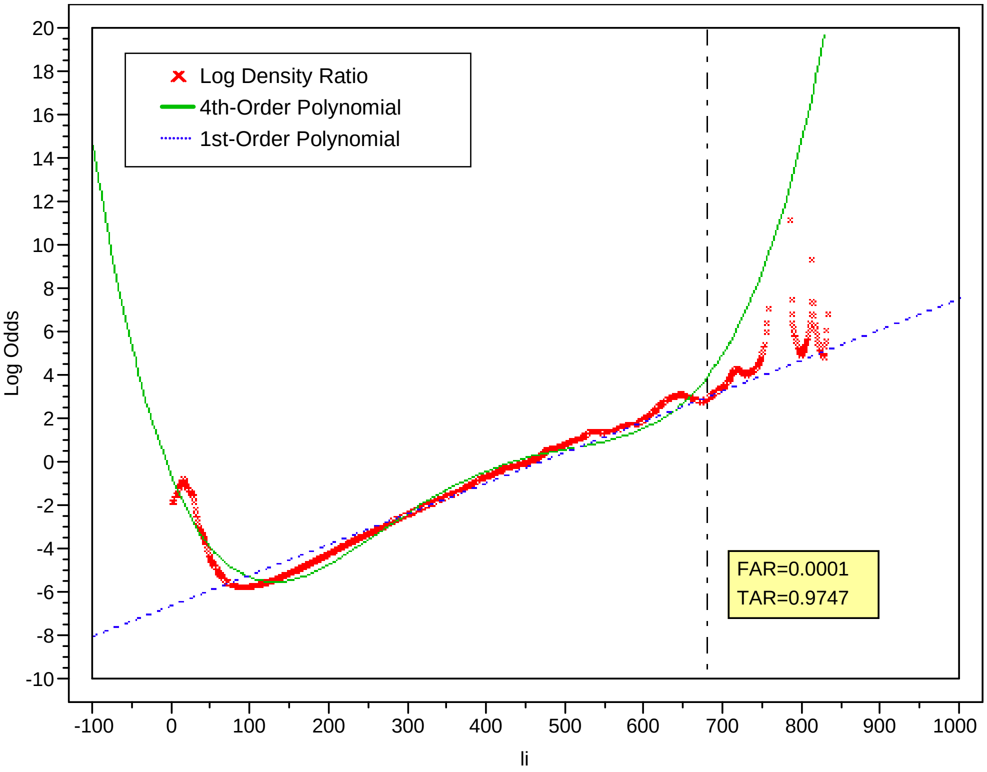

Figure 5: Matcher I (Left Index Finger): Log Odds (of a Mate) by Finger Score (li)

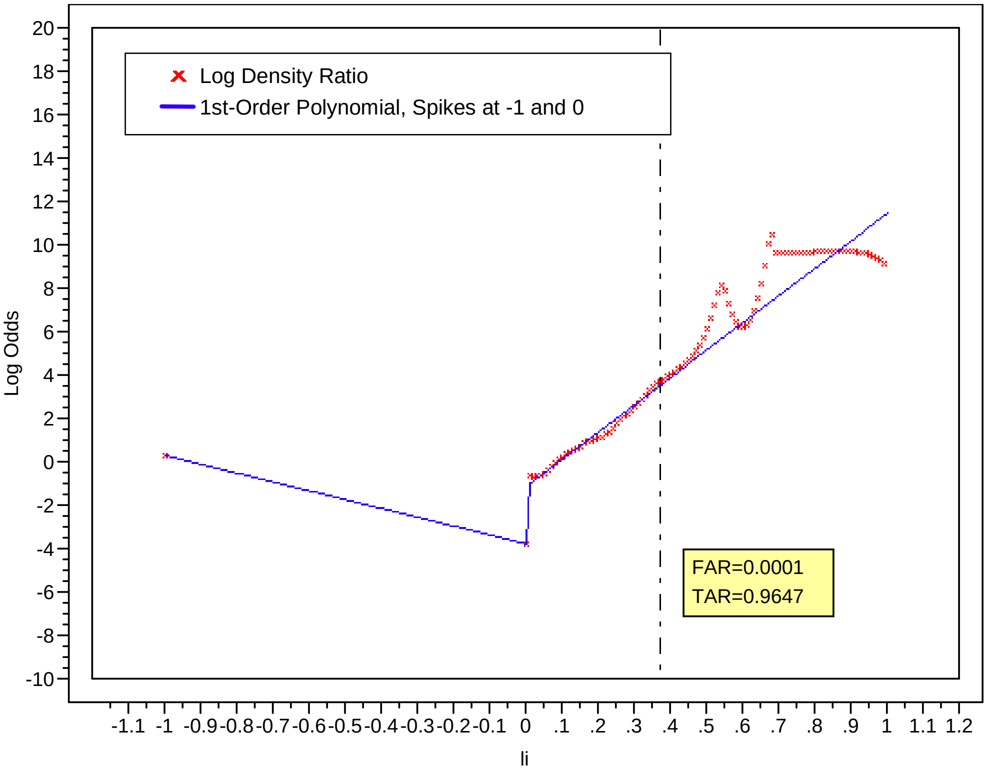

Figure 6: Matcher Q (Left Index Finger): Log Odds by Finger Score (li)

4 Fusion

Fusion refers to the combination of two or more scores to obtain a classifier with a higher level of performance. One of the simplest is to let the single-score classifiers “vote”, with rules that define the number of “mate” votes needed for the “mate” classification. For example, with two scores, the rule might assign “mate” if either of the two single-score classifiers assigns “mate”. A simple extension would be to apply weights to the votes. For example, based on the results reviewed above, one might assign higher weights to finger matchers, lower weights to face matchers.

The Neyman-Pearson lemma continues to hold if x is a vector of scores. However, multivariate implementations of the kernel method are not readily available. More fundamentally, as the number of scores to be fused increases, the data become more sparse, and the uncertainty in the density ratio correspondingly greater.

If one assumes that the scores to be fused are statistically independent of each other, then the single-score densities can be multiplied together, and the optimum classification procedure is to set a threshold on the ratio of the products of the densities (or, equivalently, the product of the ratios).

Under mild dependencies, the optimum classifier will typically not be a major improvement over one based on the assumption of independence. Nonetheless, the nature of these dependencies can be interesting in their own right.

For example, Figure 7 shows the correlation matrix for the 10 raw finger scores from Matcher I, separately for the populations of mates and non-mates. The fingers are shown in anatomical order, left little (ll) to right little (rl). Colors are used to make patterns easier to see. None of the correlations are negative. Among the mates, the correlations are stronger than among the non-mates, but the patterns are similar. One also sees that corresponding fingers on left and right hands tend to be more strongly correlated, as do the four fingers collected in each slap. Even more intriguing is the fact that adjacent fingers tend to be more correlated with each other than non-adjacent fingers.

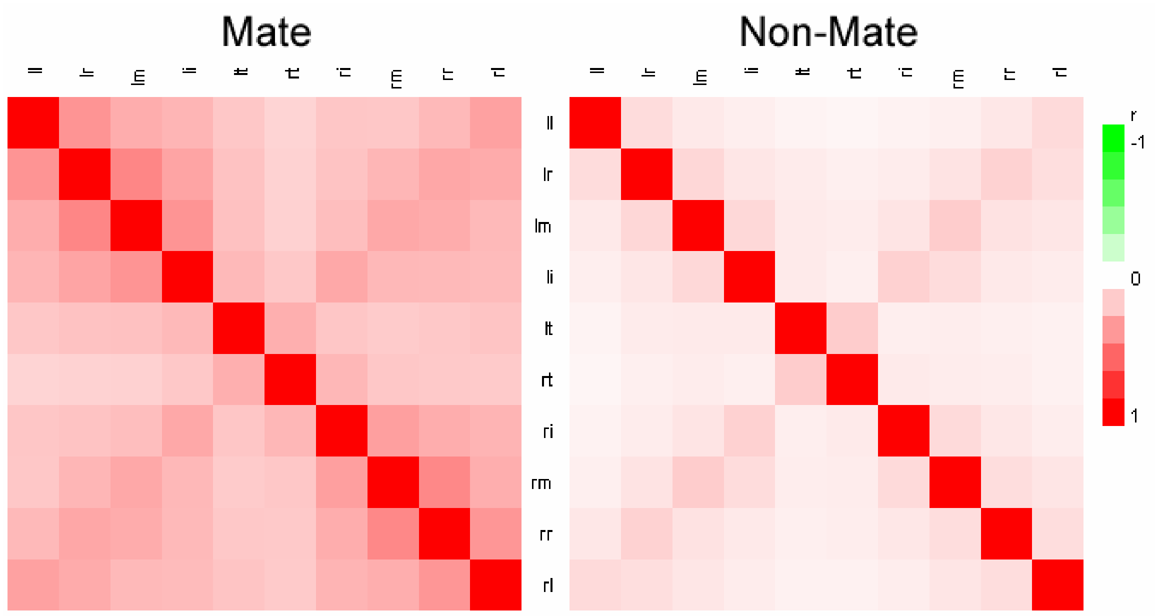

Figure 7: Color map of correlations for Matcher I

Tables 1 and 2 show the corresponding numerical correlations.

	ll	**lr**	**lm**	**li**	**lt**	**rt**	**ri**	**rm**	**rr**	**rl**
ll	1.0000	0.4161	0.3184	0.2877	0.2175	0.1672	0.2218	0.2194	0.2710	0.3677
lr	0.4161	1.0000	0.4738	0.3538	0.2392	0.1741	0.2336	0.2854	0.3424	0.3258
lm	0.3184	0.4738	1.0000	0.4192	0.2416	0.1792	0.2530	0.3409	0.3239	0.2727
li	0.2877	0.3538	0.4192	1.0000	0.2708	0.2121	0.3386	0.2746	0.2727	0.2638
lt	0.2175	0.2392	0.2416	0.2708	1.0000	0.3135	0.2198	0.2028	0.2133	0.2287
rt	0.1672	0.1741	0.1792	0.2121	0.3135	1.0000	0.2800	0.2186	0.2083	0.2065
ri	0.2218	0.2336	0.2530	0.3386	0.2198	0.2800	1.0000	0.3741	0.3199	0.2915
rm	0.2194	0.2854	0.3409	0.2746	0.2028	0.2186	0.3741	1.0000	0.4642	0.3172
rr	0.2710	0.3424	0.3239	0.2727	0.2133	0.2083	0.3199	0.4642	1.0000	0.4102
rl	0.3677	0.3258	0.2727	0.2638	0.2287	0.2065	0.2915	0.3172	0.4102	1.0000

Table 1: Correlations of Matcher I Scores among the Mates. Values range from 0.1672 to 0.4738 (ignoring the identity diagonal).

	ll	**lr**	**lm**	**li**	**lt**	**rt**	**ri**	**rm**	**rr**	**rl**
ll	1.0000	0.1354	0.0861	0.0659	0.0451	0.0360	0.0472	0.0610	0.0939	0.1434
lr	0.1354	1.0000	0.1563	0.0959	0.0760	0.0625	0.0735	0.1087	0.1756	0.1261
lm	0.0861	0.1563	1.0000	0.1528	0.0819	0.0714	0.1039	0.1979	0.1107	0.0976
li	0.0659	0.0959	0.1528	1.0000	0.0788	0.0624	0.1804	0.1339	0.0842	0.0708
lt	0.0451	0.0760	0.0819	0.0788	1.0000	0.1994	0.0646	0.0694	0.0609	0.0549
rt	0.0360	0.0625	0.0714	0.0624	0.1994	1.0000	0.0807	0.0707	0.0670	0.0515
ri	0.0472	0.0735	0.1039	0.1804	0.0646	0.0807	1.0000	0.1438	0.0918	0.0693
rm	0.0610	0.1087	0.1979	0.1339	0.0694	0.0707	0.1438	1.0000	0.1331	0.0983
rr	0.0939	0.1756	0.1107	0.0842	0.0609	0.0670	0.0918	0.1331	1.0000	0.1333
rl	0.1434	0.1261	0.0976	0.0708	0.0549	0.0515	0.0693	0.0983	0.1333	1.0000

Table 2: Correlations of Matcher I Scores among the Non-Mates. Values range from 0.0360 to 0.1994 (ignoring the identity diagonal).

There are several reasons why such correlations might occur. The prints for a single subject's hands are physically similar because they share the same genes and the similar daily patterns of use. A set of fingerprints collected from that subject are even more similar because they share the same collection environment, including the same operator and device.

4.1 Fusion with Logistic Regression

In the context of logistic regression, fusion is accomplished by including carriers h for all of the scores being fused. One may also include carriers that involve more than one score, such as products of powers. Of course, as was noted for kernel density estimation, the complexity of the model may increase rapidly with the number of scores fused. Here, complexity refers to the number of carriers, each of which involves a coefficient θ to be estimated.

For the present data, several trial models seemed to suggest that the coefficients for carriers involving multiple scores were almost always statistically insignificant. Thus, such carriers were not used. In that case, the number of coefficients would increase linearly with the number of scores fused. Even so, reducing the maximum power to which a score is raised may be necessary in order to obtain a stable estimate.

Under the assumption of independence, the regression models for the individual scores can simply be added together. There is some advantage to being able to do this. One does not need training data to fuse the scores (some data is still needed to set a threshold based on the ROC curve).

4.2 Logistic Regression Alternatives to Maximum Likelihood

One of the problems with any maximum likelihood procedure, including logistic regression, is that the results may be sensitive to model misspecification or data anomalies. Several authors have proposed alternative techniques that are "robust" to departures from assumptions.

Jain et al [Jain-99c] find the linear combination of scores that maximizes the FRR for a specified FAR. The optimization is done under the assumption that, within the populations of mates and non-mates, the scores are distributed independently, and that Equation (1) can be used to construct the ROC curve, which is therefore continuous.

Pepe et al [Pepe-05] find the linear combination of scores that maximize the area under the empirical ROC curve. If, for a given linear combination, the values for the combined samples are ranked from smallest to largest, then this is equivalent to finding the linear combination that minimizes the sum of the ranks among the non-mates (or maximizing among the mates). There would seem to be a relationship to the method described in Hettmansperger and McKean [Hettmansperger-77] for linear regression, although it is not currently clear how to exploit this relationship.

The procedure would be difficult to apply to large data sets, as the authors have observed, because it is iterative, and each iteration requires that the data be ranked anew. Moreover, the procedure will cycle after it gets close to an optimal solution, so determining convergence is also a problem.

In Equation (5), extremely low scores among the mates, or high scores among the non-mates, will be highly influential because they drive the corresponding logarithms towards (minus) infinity. Bianco and Yohai [Bianco-96] suggest an alternative criterion where the contributions from individual scores are effectively bounded. Croux and Haesbroeck [Croux-03] made some key modifications to their method. As modified by Croux and Haesbroeck, and simplifying somewhat, the Bianco-Yohai estimate minimizes

$$\sum_{\text{mates}} \rho(-\log P(T|x)) + \sum_{\text{non-mates}} \rho(-\log(1-P(T|x))), \qquad (10)$$

where

$$\rho(t) = \begin{cases} te^{-\sqrt{d}} & \text{if } t \le d \\ -2e^{-\sqrt{t}}\left(1+\sqrt{t}\right) + e^{-\sqrt{d}}\left(2\left(1+\sqrt{d}\right)+d\right) & \text{otherwise} \end{cases}, \qquad (11)$$

where d is a specified constant. (The authors use $d = 0.5$.) It is easy to see that $\rho(t)$ is bounded. Unlike the Bianco-Yohai original estimate, Croux and Haesbroeck show that this estimator exists whenever the maximum likelihood estimator exists. The authors provide a detailed algorithm for solving the minimization problem.

The Croux-Haesbroeck algorithm is implemented in S-plus, and presumably readily translated into R.

Alternatives to maximum likelihood could be an improvement in the event of model misspecification. For example, the methods might mitigate the distortion caused by spikes not included in the model. The focus in this paper has been to use the data to guide the development of the model, thus reducing the risk of misspecification.

4.3 How Many Fingers are Needed?

To illustrate the use of logistic regression for fusing fingerprint scores, consider the problem of determining the incremental benefit of involving additional fingers in the fusion. A stepwise version of logistic regression is used, in which the finger providing the greatest increase in the likelihood is added to the set of fingers already fused. The process is continued until adding additional fingers provides no significant increase in the likelihood.

In each case, the carriers for an individual finger are the score itself, plus additional carriers for spikes. The spikes are at 0 for Matcher H, and at –1 and 0 for Matcher Q.

The choice to avoid the higher-order polynomial fits used for single-score matches was made, in order to avoid simultaneous estimation of large numbers of parameters, as the number of fingers increased.

The stepwise procedure is difficult to do when the single-finger model includes terms for spikes, since there is generally no easy way to identify to most software packages that such terms are linked together. Fortunately, the model for Matcher I includes no such terms. Accordingly, the stepwise procedure with Matcher I was used to identify sequential sets of fingers. The same sets were used for Matchers H and Q.

The results are shown in Figures 8, 9, and 10 for Matchers H, I, and Q, respectively. Each plot shows FRR plotted against FAR, both using logarithmic scales, for increasing numbers of fingers. The specific fingers are identified. However it is likely that some other combinations of fingers would produce very similar results.

The plots for Matchers H and I suggest that approximately 0.05% of the mates (~30 individuals) in the training set will not be correctly classified, regardless of the number of fingers or the FAR threshold used for the classification. Visual review of the data found 33 subjects for whom the mate associations were incorrect.[2] In other words, for this dataset, FRR below 0.051% cannot generally be achieved unless data integrity errors are excluded; that limit is reached with a variety of 3- or 4-finger combinations.

The plots also reveal some differences among the matchers. For Matcher I, three fingers are sufficient to reach this 0.05% limit, while Matcher H takes 4 fingers but reaches a slightly lower FRR. Matcher Q, on the other hand, does not do as well with the five fingers.

4.4 Cross Validation

The numbers of non-mates above the FAR=0.0001 score and mates below that score are small. This raises the concern that the high TAR (low FRR) observed in the previous section may be

---

[2] In the NBDF06 dataset, 33 subjects out of 64,867 (0.051%) were found to have some or all of their fingers misidentified, of whom 24 (0.037%) also have their faces misidentified. FRR can pass the 0.051% limit with some finger combinations, but not the 0.037% limit.

overly influenced by the samples, and not adequately reflective of the population from which they were drawn. Particularly, FRR may be biased downward.

Cross-validation can be used to produce a more unbiased estimate of TAR at FAR=0.0001, and, incidentally, to give a confidence interval for the true average TAR for this estimation method. This is illustrated with the results for Matcher I.

Cross-validation requires that the estimation method first be rigorously defined: Stepwise logistic regression is used to fit a model involving the linear terms only for the 10 individual finger scores. The first 5 selected fingers are used to construct the model.

The cross-validation was carried out by assigning each pair of images (probe and gallery) randomly to one of 10 groups. For each of the 10 groups, the model was developed using only the pairs not in the group. The ROC curve from the model was then constructed using only pairs in the group. In effect, 90% of the data are used to develop the model, 10% to evaluate it.

Table 1 shows, for each of the groups, the TAR corresponding to FAR=0.0001, as well as a list of the 5 fingers selected (not in the order selected). There is considerable variation among the individual results, expected since, with only about 6000 mates, very few will fall below the threshold. (In one case, the fit achieved total separation of mates and non-mates.)

With respect to which fingers were selected, there is considerable (but not total) agreement among the 10 groups. This would be expected, however, since there is a lot of overlap among the groups with respect to the data used in model construction.

The TAR mean and standard deviation are shown at the bottom of the table. The mean TAR is 0.99951, showing little bias relative to results based on the full data set. A 95% confidence interval for the true TAR is 0.99951±0.00018.

Group	TAR	Selected Fingers
1	0.9998	rt, rr, rm, lt, li
2	0.9994	rt, rr, lt, lr, li
3	0.9994	rt, rr, lt, lr, li
4	0.9995	rt, rr, lt, lr, li
5	0.9997	rt, rr, rm, lt, lr
6	0.9995	rt, rr, lt, lr, li
7	0.9994	rt, rr, lt, lr, li
8	1.0000	rt, rr, ri, lt, lr
9	0.9992	rt, rr, rm, lt, lr
10	0.9992	rt, rr, rl, ri, lt
Mean	0.99951	
Standard Deviation	0.00026	

Table 1. Cross-Validation Results for matcher I

4.5 How Many Fingers Assuming Independence?

If one can assume that the scores for the 10 fingers are independent within the populations of mates and non-mates, then the regression-derived scores $q(x;\theta)$ for different fingers can be simply added together.

There are two reasons why the ability to easily combine fingers is useful. First, one can create new fusions without the data needed to perform a logistic regression on the carriers for the combined fingers, as was done in Figures 8, 9, and 10. Second, there is no need to reduce the degree of polynomial fitted to single scores in order to keep the number of parameters θ manageable.

Figure 11 shows, for Matcher I, the results of simply adding together 1st-order polynomials separately obtained for the sequential sets of fingers used in Figure 9. Figure 12 does the same, but with 4th-order polynomials. These correspond to the regressions considered for Matcher I with the left index finger in Figure 5. As in Figure 9, FRR is plotted against FAR.

Figures 9, 11, and 12 are almost identical, except at the lowest FAR levels, where the differences seen are with respect to the classification of single items.

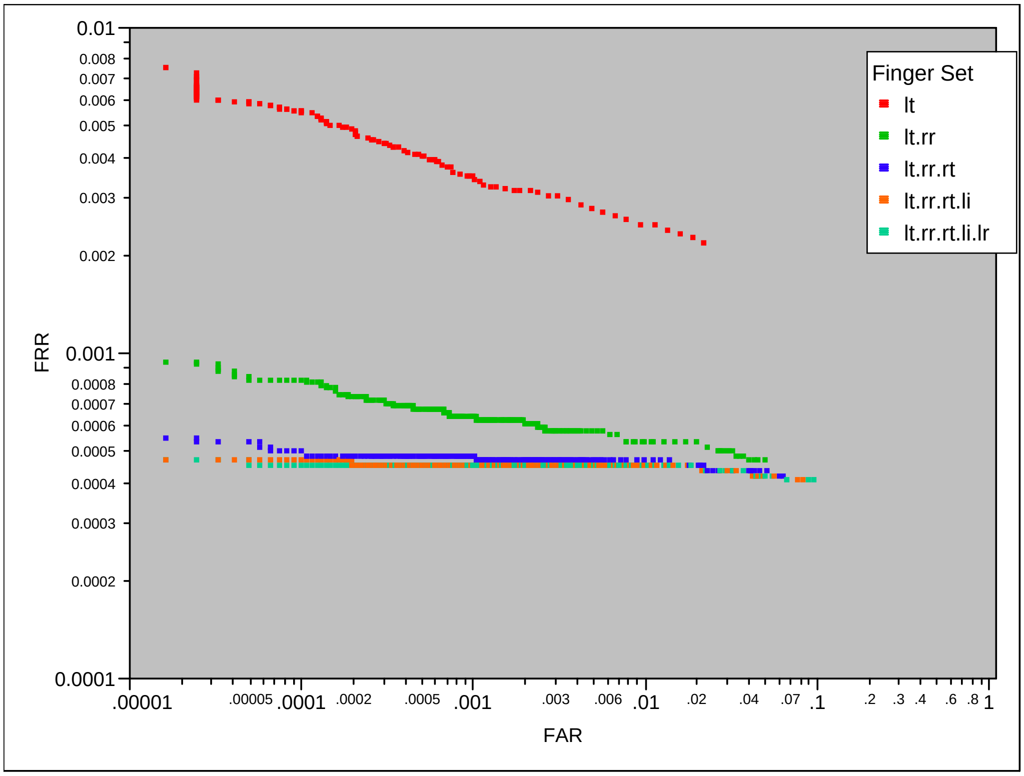

Figure 8: Matcher H: FRR by FAR, by Finger Set

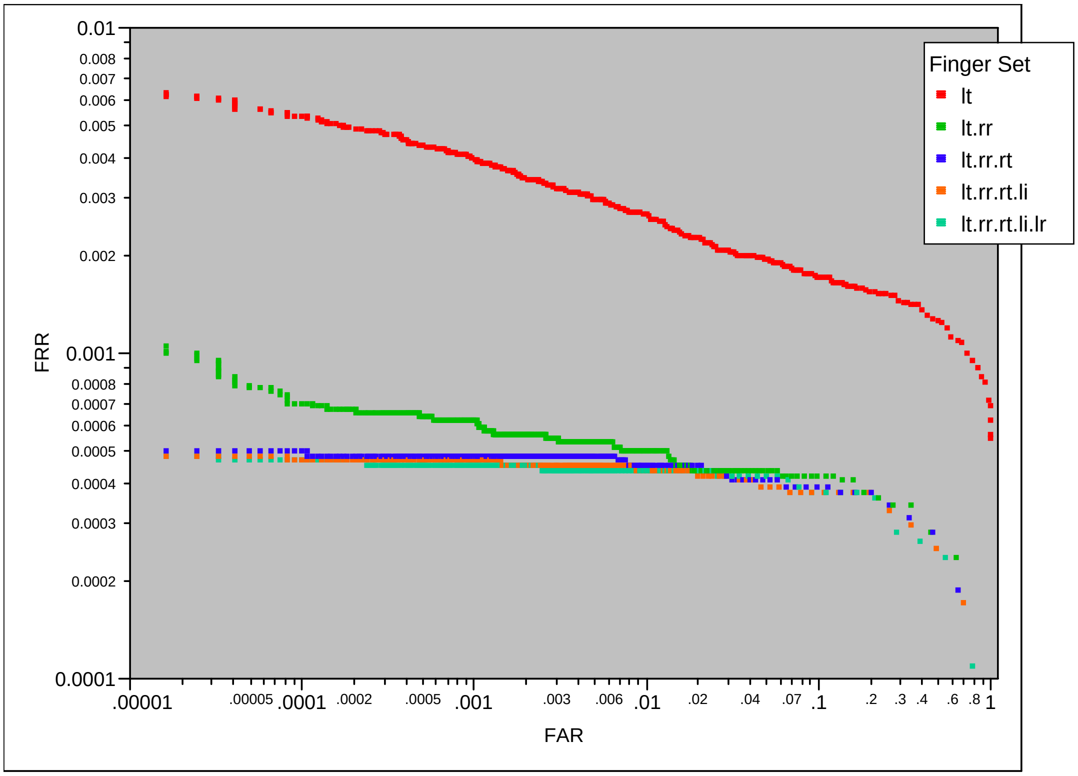

Figure 9: Matcher I: FRR by FAR, by Finger Set

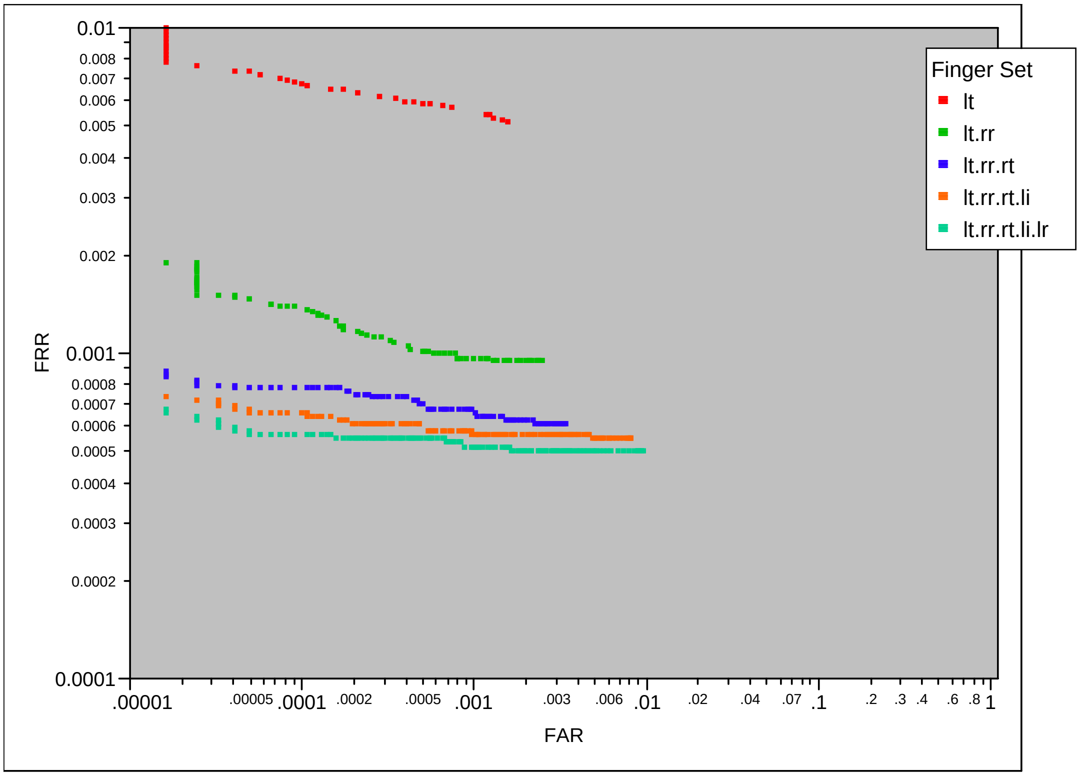

Figure 10: Matcher Q: FRR by FAR, by Finger Set

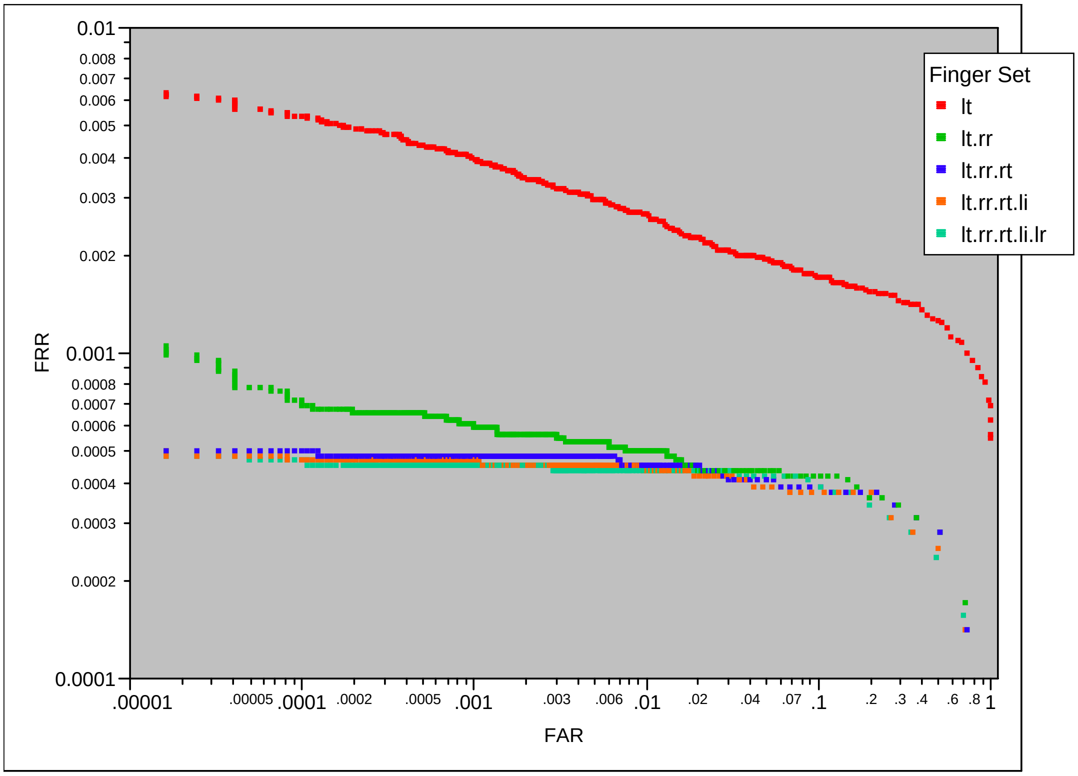

Figure 11: Matcher I: FRR by FAR, by Finger Set – Using 1st –Order Polynomials and summing the single-finger models

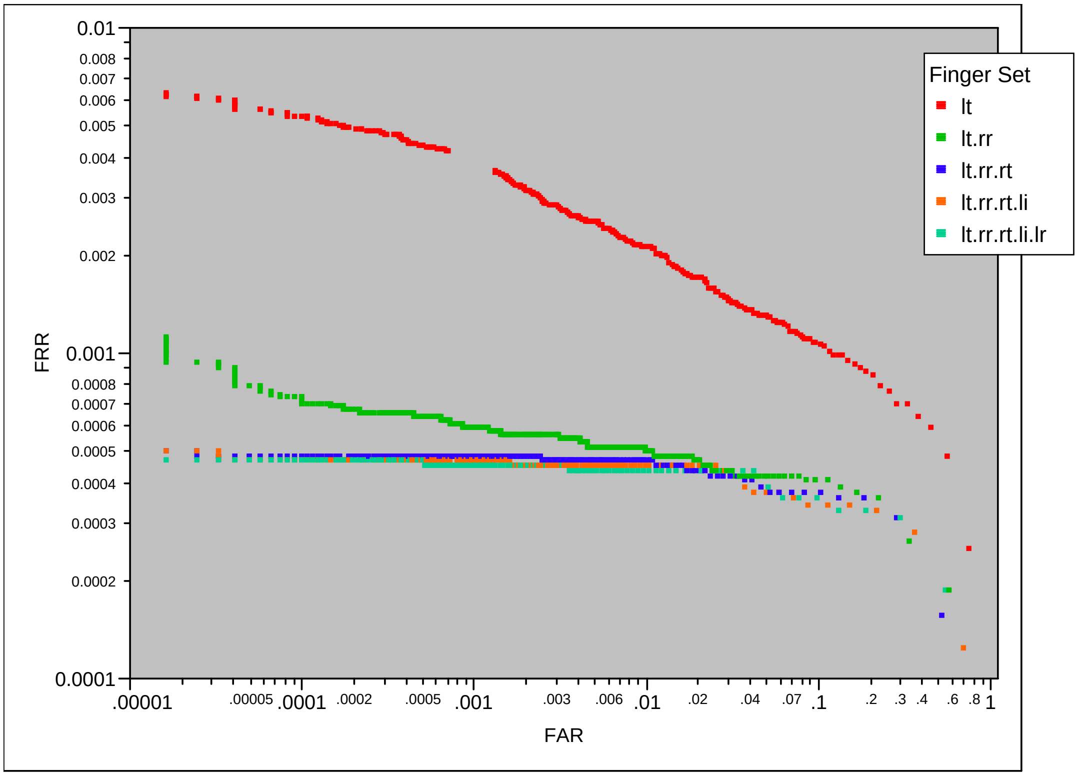

Figure 12: Matcher I: FRR by FAR, by Finger Set – Using 4th –Order Polynomials and summing the single-finger models

---

5 Conclusion

Classification by Logistic Regression is a useful alternative to classification by density ratio. It is strongly related to the density ratio method, but produces a classification that can be summarized in a relatively small number of numerical parameters. The classifier is reasonably flexible, in that it is not necessarily a linear combination of the raw scores. While not needed for the current data set, the method can be applied even where one wishes to include modeling the dependence among the raw scores. Finally, the method is widely implemented among commonly used statistical software packages.

---

6 References

[Bianco-96] A M Bianco and V J Yohai; "Robust Estimation in the Logistic Regression Model", in *Robust Statistics, Data Analysis, and Computer Intensive Methods*, 17-34; Lecture Notes in Statistics 109, Springer Verlag, Ed. H Rieder. New York, 1996.

[Croux-03] C Croux and G Haesbroeck (2003) "Implementing the Bianco and Yohai Estimator for Logistic Regression", *Computational Statistics & Data Analysis*, 44, 273-295.

[Ghosh-04b] A Ghosh and P Chaudhuri (2004) "Optimal Smoothing in Kernel Discriminant Analysis", *Statistica Sinica* 14, 457-83.

[Griffin-05] P. Griffin; "Optimal Fusion for Multi-Biometric Identity Verification"; Identix Research Preprint RDNJ-05-0001, Jan. 2005.

[Griffin-05] P. Griffin; "Optimal Fusion for Multi-Biometric Identity Verification"; Identix Research Preprint RDNJ-05-0001, Jan. 2005.

[Hettmansperger-77] T P Hettmansperger and J W McKean (1977) "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models", *Technometrics* 19, 275-284.

[Jain-99c] A. K. Jain, S. Prabhakar and S. Chen; "Combining Multiple Matchers for a High Security Fingerprint Verification System"; *Pattern Recognition Letters*, Vol 20, No. 11-13, pp. 1371-1379, 1999

[Pepe-05] M S Pepe, T Cai, Z Zhang, and G M Longton; "Combining Predictors for Classification using the Area Under the ROC Curve", *University of Washington Biostatistics Working Paper Series* 238; 2005.

[Verlinde-00] P Verlinde, G Chollet, and M Acheroy; "Multi-modal Identity Verification Using Expert Fusion", *Information Fusion* 1, 17-33; 2000.