Abstract

The performance of biometric recognition systems can become limited in operating environments due to the presence of extraneous noise factors. Biometric fusion alleviates this problem by consolidating information from various biometric sources, thereby achieving a higher recognition rate. One challenge faced in fusion is how to optimally combine information originating from the different sources. We present a framework for fusing matching scores from multiple biometric sources using multivariate $t$-copulas. Our approach involves eliciting suitable models for the joint distribution of matching scores via the $t$-copulas to accommodate generalized marginals (densities having a mixture of both discrete and continuous components). For estimating the unknown parameters in our models, we show that the discrete components can be viewed as missing components in an Expectation-Maximization (EM) framework. The newly developed estimation technique is applied to model the distribution of genuine and impostor matching scores from multiple biometric sources. Biometric fusion is subsequently performed based on the likelihood ratio statistic.

KEY WORDS: Copula functions, $t$-copulas, generalized densities, biometric authentication, biometric fusion.

1. Introduction

Biometric recognition, or biometrics, refers to the automatic recognition of a person based on his/her physiological or behavioral characteristics; see Jain, Bolle & Pankanti (1999) and Maltoni, Maio, Jain & Prabhakar (2003). Biometric recognition offers many advantages over traditional PIN number- and token-based approaches; for example, a biometric trait cannot be easily transferred, forgotten or lost, the rightful owner of the biometric template can be easily identified, and it is difficult to duplicate a biometric trait. Some well-known examples of traits used in biometric recognition are fingerprint, iris, face, signature, voice and hand geometry (see Figure 1). A number of commercial recognition systems based on these traits have been deployed and are currently in use. Biometric technology has now become a viable alternative to traditional recognition systems in many government applications (e.g., US-VISIT program and the proposed biometric passport which is capable of storing biometric information of the owner in a chip inside the passport). With increasing applications involving human-computer interactions, there is a growing need for fast recognition techniques that are reliable and secure.

A biometric system is essentially a pattern recognition system that verifies an individual based on an input biometric trait and a claimed identity. Based on the claimed identity, a template of the biometric trait is retrieved from the stored database. Subsequently, we are interested in accepting inputs which are “close” or “similar” to the retrieved template, and rejecting those that are “far” or “dissimilar”. Biometric recognition can also be cast as a problem of testing statistical hypotheses. Based on the input $Q$ and a claimed identity $I_c$, we are interested in testing the hypothesis

$$H_0 : I_t = I_c \text{ vs. } H_1 : I_t \neq I_c,$$

where $I_t$ is the true identity of the user. In (1), $H_0$ is the null hypothesis that the user is genuine whereas $H_1$ is the alternative hypothesis that the user is an imposter. The testing in (1) is performed by a matcher which computes a similarity measure, $S(Q, T)$, based on $Q$ and $T$; large (respectively, small) values of $S$ indicate that $T$ and $Q$ are close to (far from) each other. A threshold, $\lambda$, is specified so that all similarity values lower (respectively, greater) than $\lambda$ lead to the rejection (acceptance) of $H_0$. Thus, decisions of whether to accept or reject $H_0$ in the verification mode are prone to two types of errors: the false reject rate (FRR), which is the probability of rejecting $H_0$ when in fact the user is genuine, and the false accept rate (FAR), which is the probability of accepting $H_0$ when in fact the user is an imposter. The distribution of $S(Q, T)$ is called the genuine matching distribution when $Q$ and $T$ are a pair of genuine traits (i.e., corresponding to the same user). In the case when $Q$ and $T$ are an impostor pair (corresponding to two different users), the resulting distribution of $S(Q, T)$ is called the impostor matching distribution.

The performance of a biometric authentication system depends on how separated the genuine and impostor distributions are for a particular biometric trait. Often, biometric systems based on a single trait (i.e., unimodal systems) suffer from limited performance due to lack of pattern uniqueness, non-universality and noisy data Jain & Ross (2004). One way to alleviate these problems is to use multi-biometric systems. Multi-biometric systems combine information from its component modalities to arrive at a decision, subsequently achieving better performance.

---

1 A matcher can also compute a distance measure between $Q$ and $T$ in which case similar $Q$ and $T$ will produce distance values that are close to zero and vice versa.
and better separation between the genuine and impostor matching distributions compared to the unimodal systems. For multi-biometric systems, consolidation of information is usually conducted at the matching score level since this is most useful and practical. However, there are several challenges involved in fusing matching scores: First, matching scores from different sources may not be compatible; for example, in the NIST-BSSR1 database (2004), matching scores arising from the two face matchers have ranges in \([-1,1]\) and \([0,100]\). In this case, finding a common domain for combining the vastly different scores can be quite challenging. Further, the scores of different matchers can be either distance or similarity measures, and they may follow different probability distributions. It is also well-known that both the genuine and impostor matching score distributions consist of components that have non-Gaussian distributions and can be highly correlated.

For such multivariate observations, copula functions allow flexible modeling of the joint distribution via its marginal distributions as well as the correlation between pairwise components of the vector of observations. A copula function essentially “couples” the marginal distributions to form a valid joint distribution for a multivariate random vector. More specifically, let \(F_i, i = 1,2,\ldots,D\) denote \(D\) continuous distribution functions on the real line, and let \(F\) denote a joint distribution function on \(R^D\) whose \(k\)-th marginal distribution functions correspond to \(F_k, k = 1,2,\ldots,D\). According to Sklar’s Theorem, see Nelsen (1999), there exists a unique function \(C(u_1,u_2,\ldots,u_D)\) from \([0,1]^D\) to \([0,1]\) satisfying

\[
F(x_1,x_2,\ldots,x_D) = C(F_1(x_1),F_2(x_2),\ldots,F_D(x_D)),
\]

where \(x_1,x_2,\ldots,x_D\) are \(D\) real numbers. The function \(C\) is known as a \(D\)-copula function that couples the one-dimensional distribution functions \(F_k, k = 1,2,\ldots,D\) to obtain \(F\). It can be shown (see, for example, Nelsen (1999)) that copula functions are \(D\)-dimensional distribution functions on \([0,1]^D\) with uniformly distributed marginals. Equation (2) can also be used to construct \(D\)-dimensional distribution functions \(F\) whose marginals are the pre-specified distributions \(F_k, k = 1,2,\ldots,D\): Choose a copula function \(C\) and define the distribution function \(F\) as in (2). It follows that \(F\) is a \(D\)-dimensional distribution function with marginals \(F_k, k = 1,2,\ldots,D\).

Copula-based models are natural in situations where learning about the association between the variables is important, since the effect of the dependence structure, is easily separated from that of the marginals, see Genest, Ghoudi & Rivest (1995). Typically, there is enough data to obtain nonparametric estimates of the marginal distributions, but insufficient information to afford non-parametric estimation of the structure of the association. In such cases, it is convenient to adopt a parametric form for the copula function \(C\) while keeping the marginals unspecified. The flexibility of modeling both the marginals as well as the correlation structure via copula functions have been the reason for their popularity in many fields. Copula-based models, related estimation techniques as well as properties of the estimators been well studied with applications ranging from statistics to mathematical finance and finance risk management; see, for example, Genest, Ghoudi & Rivest (1995), Derart & McNeil (2005), Embrechts, Lindskog & McNeil (2003), Cherubini, Luciano & Vecchiato (2004), Chen & Fan (2005), Chen & Fan (2006) and Frey & McNeil (2001). However, an important distinction is that previous work developed estimation and inference techniques assuming that the marginals are continuous. In our present application, both genuine and impostor matching score distributions contain many discrete components. The reason is as follows: Many matchers assign a default score (a pre-specified value, say 0) if certain “alignment conditions” are not met for the query and template. For example, a fingerprint matcher, see Ratha, Chen, Karu & Jain (1996), may decide that a pair of fingerprints is an impostor pair if their orientation fields (see Dass & Jain (2004)), do not initially align. In such a case, the score 0 is assigned, which accounts for discrete values in the distribution of matching scores. For the same reason, the first face matcher in the NIST-BSSR1 database can output a discrete match score with value −1. Although copula functions are quite successful in modeling multivariate joint distributions, a central assumption for the joint distribution is that the marginals should be continuous. In order to adequately represent the distribution of genuine and impostor matching score distributions, we require to extend the framework to modeling of the joint distribution via copula functions to incorporate marginals with mixed components (both discrete and continuous). In this paper, we work with \(t\)-copulas for eliciting the joint distributions. One reason for our choice is that \(t\)-copulas are able to represent non-zero tail dependence as is usually observed for biometric data.

The paper is organized as follows: Section 2 develops multivariate joint distributions based on \(t\)-copulas where the corresponding marginals are a mixture of discrete and continuous distributions.
continuous components. The identifiability of the parameters associated with the t-copulas is established in this Section. In Section 3, we develop an methodology to estimate the unknown parameters of the t-copulas based on the EM algorithm. A result on the consistency of the parameter estimates as the sample size tends to infinity is presented in Section 4. Finally, results based on simulated examples and a real application related to biometric recognition are presented in Section 5. We summarize our work in Section 6.

2. Joint Distributions with Generalized Marginals

For a joint distribution with continuous marginal distributions, it can be shown that there is a unique copula $C$ for which representation (2) holds. If $F_1, \ldots, F_D$ are not all continuous, it can still be shown (for example, see Schweizer & Sklar (1974)) that the joint distribution function can always be expressed as in (2), although in this case $C$ is no longer unique. Uniqueness is an important consideration when estimating the unknown parameters corresponding to the copula function. Since our marginal distributions have discrete components, one way to achieve uniqueness is to restrict $C$ to belong to a particular parametric family. As mentioned before, we restrict our attention to the family of t-copulas, see Nelsen (1999).

Subsequently, we develop some notations for the presentation. A boldfaced letter such as $x$ denotes a $D$-tuple vector of real numbers, that is, $x \equiv (x_1, \ldots, x_D)^T$ where $x_k \in \mathbb{R}$ for $k = 1, \ldots, D$. Also, $\Sigma$ will denote a positive definite matrix of dimension $D \times D$. For such a $\Sigma$, the $D$-dimensional t density with $\nu$ degrees of freedom will be denoted by

$$f_{\nu,\Sigma}(x) = \frac{\Gamma(\frac{\nu+D}{2})}{(\pi \nu)^{\frac{D}{2}} \Gamma(\frac{\nu}{2}) |\Sigma|^\frac{1}{2}} \left(1 + \frac{x^T \Sigma^{-1} x}{\nu}\right)^{-\frac{\nu+D}{2}},$$

(3)

with corresponding distribution function

$$t_{\nu,\Sigma}(x) \equiv \int_{w \leq x} f_{\nu,\Sigma}(w) \, dw.$$  

(4)

The matrix $\Sigma$ with unit diagonal entries corresponds to a correlation matrix and will be denoted by $R$.

The $D$-dimensional t-copula function is given by

$$C_{\nu,R}(u) \equiv \int_{w \leq u_1} f_{\nu,\Sigma}(w) \, dw,$$

(5)

where $t_{\nu}^{-1}(u) \equiv (t^{-1}_u(u_1), \ldots, t^{-1}_u(u_D))^T$, $t^{-1}_u$ is the inverse of the cumulative distribution function of univariate $t$ with $\nu$ degrees of freedom, and $R$ is a $D \times D$ correlation matrix. Note that $C_{\nu,R}(u) = P(X \leq t^{-1}_u(u))$, for $X = (X_1, \ldots, X_D)^T$ distributed as $t^{D}_{\nu,R}$, demonstrating that $C_{\nu,R}(u)$ is a distribution function on $[0, 1]^D$ with uniform marginals. The density corresponding to $C_{\nu,R}(u)$ is given by

$$c_{\nu,R}(u) = \frac{\partial^D C_{\nu,R}(u)}{\partial u_1 \partial u_2 \ldots \partial u_D} = \frac{f_{\nu,R}(t^{-1}_u(u))}{\prod_{k=1}^{D} f_{\nu,R}(t^{-1}_u(u_k))},$$

(6)

where $f_{\nu}$ in (6) is the density of the univariate $t$ distribution with $\nu$ degrees of freedom.

We now turn our attention to eliciting joint distribution functions on $R^D$ with marginal distributions that are a mixture of both discrete and continuous components. Let $F_1, \ldots, F_D$ denote $D$ univariate marginal distributions on the real line with a mixture of discrete and continuous components given by

$$F_k(x) = \sum_{h=1}^{d_k} p_{kh} I_{\{D_{kh} \leq x\}} + (1 - \sum_{h=1}^{d_k} p_{kh}) \int_{w \leq x} f_k(w) \, dw,$$

(7)

where $I_{\{A\}}$ is the indicator function with $I_{\{A\}} = 1$ if $A$ is true, and 0 otherwise; $D_{kh}$ is the $h$-th discrete component of $F_k$ with $P(X = D_{kh}) = p_{kh}$, for $h = 1, \ldots, d_k$, and $f_k(x)$ is a continuous density function with support on the real line. We consider joint distributions of the form

$$F^{D}_{\nu,R}(x) = C_{\nu,R}(F_1(x_1), \ldots, F_D(x_D))$$

(8)

with $C_{\nu,R}$ and $F_k$ defined as in (5) and (7), respectively. It follows from the properties of a copula function that $F^{D}_{\nu,R}$ is a valid multivariate distribution on $R^D$. The identifiability of the marginal distributions $F_k$, $k = 1, \ldots, D$, the correlation matrix $R$, and the degrees of freedom parameter, $\nu$, is established in the following theorem:

Theorem 2.1 Let $F^{D}_{\nu_1,R_1}$ and $G^{D}_{\nu_2,R_2}$ denote two distribution functions on $R^D$ obtained from equation (8) with marginal distributions $F_k, k = 1, \ldots, D$ and $G_k, k = 1, \ldots, D$, respectively. Suppose we have $F^{D}_{\nu_1,R_1}(x) = G^{D}_{\nu_2,R_2}(x)$ for all $x$. Then, $F_k(x) = G_k(x)$ for all $k$, $\nu_1 = \nu_2$ and $R_1 = R_2$.

The density of $F^D_{\nu,R}$, $dF^D_{\nu,R}$, is a function on $R^D$ that satisfies

$$F^D_{\nu,R}(x) = \int_{w \leq x} dF^D_{\nu,R}(w).$$

(9)

We give a characterization for the density of $F^D_{\nu,R}$ for every point $x \in R^D$. For a fixed $x \in R^D$, every component of $x$ corresponds to either a point of continuity or discontinuity of the corresponding marginal distribution function. For $1 \leq a_1 < \ldots < a_L \leq D$, let $x_{a_l}$ denote the $a_l$-th component of $x$ that corresponds to a discrete component in the marginal distribution $F_{a_l}$, $l = 1, \ldots, L$ (that is, $x_{a_l}$ equals $D_{a_l,h_l}$ for some $1 \leq h_l \leq d_{a_l}$). The remaining components of $x$ are points of continuity of the corresponding marginal distribution functions. These components are denoted by $1 \leq b_1 < \ldots < b_M \leq D$, where $M = D - L$. It follows that the sets $\{a_1, \ldots, a_L\}$ and $\{b_1, \ldots, b_M\}$ are
disjoint and their union is the set of integers \( \{1, \ldots, D\} \). In this case, the density \( dF^D_{\nu,R}(x) \) is given by

\[
dF^D_{\nu,R}(x) = \left( \sum_{m=1}^{M} f_{bm}(x_{bm}) \right) c^*(\nu, R, F_1, F_2, \ldots, F_D)(x),
\]

where \( c^*: R^D \to [0, \infty) \) is a function of \( x \) that depends on the parameters \( \nu \) and \( R \) as well as the marginal distributions \( F_k, k = 1, \ldots, D \). The explicit form of \( c^* \) is given by

\[
c^*(\nu, R, F_1, \ldots, F_D)(x) = \int_{S_{a_1}} \ldots \int_{S_{a_L}} c_{\nu,R}(u) du_{a_1} \ldots du_{a_L}.
\]

3. Estimation of \( R \) and \( \nu \)

3.1 Estimation of \( R \) for fixed \( \nu \)

This section discusses the estimation of \( R \) based on \( n \) iid observations from the joint distribution \( (8) \) for fixed \( \nu \). Let \( X_1, \ldots, X_n \) be \( n \) independent and identically distributed \( D \)-dimensional random vectors arising from the joint distribution \( F^D_{\nu,R} \) in (8). We denote the components of \( X_j \) by \( X_{j,l} = (X_{1,l}, \ldots, X_{D,l})^T \) for \( j = 1, \ldots, n \). The discrete components \( D_{kh}, h = 1, \ldots, d_k \) corresponding to marginal \( F_k \) (see equation (7)) for \( k = 1, \ldots, D \) will be assumed to be known in advance. The remaining components of \( F_k \) consisting of the probabilities \( p_{kh} \) and the density \( f_k(x) \) will be assumed to be unknown. The empirical distribution function corresponding to \( F_k \) evaluated at every \( x_k \in R \) is the sample proportion of observations that are less than or equal to \( x_k \), that is,

\[
\hat{F}_{nk}(x_k) = \frac{1}{n} \sum_{j=1}^{n} I(x_{kj} \leq x_k).
\]

The empirical distribution function \( \hat{F}_{nk} \) is a non-parametric estimate of \( F_k \) and is consistent for \( F_k \) for large \( n \).

The log-likelihood function, \( \log(dF^D_{\nu,R}(x)) \), corresponding to the \( n \) iid observations \( X_j, j = 1, \ldots, n \), is given by the expression

\[
\sum_{j=1}^{n} \log(dF^D_{\nu,R}(X_j)) = \sum_{j=1}^{n} \sum_{m=1}^{M} \log(f_{bm}(x_{bm,j})) + \sum_{j=1}^{n} \log c^*(\nu, R, F_1, \ldots, F_D)(X_j),
\]

and will be used to estimate \( R \). Note that the first term after the equality in (13) does not depend on \( R \). Thus, we can focus our attention only on the second term of the log-likelihood, namely,

\[
\tau(\nu, R) = \sum_{j=1}^{n} \log c^*(\nu, R, F_1, \ldots, F_D)(X_j).
\]

Since the expression in (14) also depends on the unknown marginal distribution functions \( F_k, k = 1, \ldots, D \), we plug-in a consistent estimate of \( F_k \), namely, \( \hat{F}_k = \frac{1}{n+1} \hat{F}_{nk} \) (see (12)) in its place. This recasting technique avoids difficulties arising from the unboundedness of \( \log c^*(\nu, R, F_1, \ldots, F_D) \) as some of the \( F_i \)'s tend to 1, and has been employed by Genest, Ghoudi & Rivest (1995). This results in the expression

\[
\hat{\tau}(\nu, R) = \sum_{j=1}^{n} \log c^*(\nu, R, \hat{F}_1, \ldots, \hat{F}_D)(X_j)
\]

which has \( R \) and \( \nu \) as the only unknown parameters. For fixed \( \nu \), our estimate of \( R \) is taken to be the maximizer of \( \hat{\tau}(\nu, R) \), that is,

\[
\hat{R}(\nu) = \arg \max_R \hat{\tau}(\nu, R).
\]

If discrete components are absent (that is, \( F_k \) is continuous for each \( k \)), the resulting likelihood has the expression

\[
\sum_{j=1}^{n} \sum_{k=1}^{D} \log(f_k(x_{kj})) + \sum_{j=1}^{n} \log c_{\nu,R}(F_1(x_{1j}), \ldots, F_D(x_{Dj}))
\]

where \( c_{\nu,R} \) is the density of the \( t \)-copula in (6). In order to estimate the unknown parameters based on the above likelihood, Genest, Ghoudi & Rivest (1995) developed a semiparametric approach that involved plugging \( \hat{F}_k \) in place of \( F_k \) in (17). Subsequently, it was shown that the resulting estimators were consistent as \( n \) tends to infinity; see Genest, Ghoudi & Rivest (1995) for details. Expressions (13) and (16), respectively, are generalizations of (17) and the methodology of Genest, Ghoudi & Rivest (1995), when \( F_k \) contains discrete components. We proceed to describe a computational scheme that can be used to find the estimate of \( R \).

Note that there is a big challenge in maximizing the likelihood \( \hat{\tau}(\nu, R) \): \( \hat{\tau}(\nu, R) \) involves several integrals corresponding to discrete components in \( X_j, j = 1, \ldots, n \); \( c^* \) is not available in a closed form like \( c_{\nu,R} \) in (17). This difficulty can be overcome with the use of the EM algorithm. Details of the algorithm to the present context are presented in the technical report: Huang and Dass (2007).

3.2 Selection of the degrees of freedom, \( \nu \)

The above estimation procedure for \( R \) is carried out for a collection of degrees of freedom \( \nu \in A \), \( A \) finite. For each fixed \( \nu \), we obtain the estimate of the correlation matrix \( \hat{R}(\nu) \) based on the EM algorithm above. We select the degrees of freedom in the following way: Select \( \hat{\nu} \) such that

\[
\hat{\nu} = \arg \max_{\nu \in A} \frac{1}{n} \hat{\tau}(\nu, \hat{R}(\nu)).
\]
4. Consistency

The consistency of the parameter estimates \((\hat{\nu}, \hat{R}(\hat{\nu}))\) is established in this section. Let

\[
KL(F_0, F_{\nu,R}) = \int_{\mathbb{R}^p} \log \frac{dF_0}{dF_{\nu,R}}(x) \, dF_0(x) \tag{19}
\]
denote the Kullback-Liebler distance between the true distribution of \(X, F_0\), and a candidate distribution, \(F_{\nu,R}\), defined as in (8). One assumption needed for the KL distance to be defined (a finite number) is that the support of \(F_0\) and \(F_{\nu,R}\) should be identical, and this is the assumption we make here. The following theorem establishes consistency:

**Theorem 4.1** Let \((\hat{\nu}, \hat{R}(\hat{\nu}))\) be obtained as in (16) and (18). Then, as \(n \to \infty\), \((\hat{\nu}, \hat{R}(\hat{\nu})) \to (\nu^*, R^*)\) where \((\nu^*, R^*) \equiv \arg \min_{(\nu,R), \nu \in \mathcal{A}} KL(F_0, F_{\nu,R})\).

It is nature to have the following corollary:

**Corollary 1** If \(F_0 = F_{\nu_0,R_0}^D\), then as \(n \to \infty\), \((\hat{\nu}, \hat{R}(\hat{\nu})) \to (\nu_0, R_0)\).

5. Experimental Results

5.1 Simulated Data

In this section, results based on simulated data for three experiments are presented. The true distribution for the observations \(X_j, j = 1, \ldots, n\) in all three experiments is taken to be \(F_0(x) = F_{\nu_0,R_0}^D(x)\) with \(F_{\nu,R}^D\) defined as in (8).

In the first experiment, we took \(D = 2\) with \(\nu_0 = \infty\) corresponding to the Gaussian copula function (see Nelsen (1999)). Since \(D = 2\), the correlation matrix \(R_0\) is determined by a single parameter \(\rho_0\). We take \(\rho_0 = 0.75\) here. The cumulative distribution functions corresponding to the two marginal distributions are given by

\[
F_1(x_1) = 0.25I_{(x_1 \geq 0.2)} + 0.6I_{(x_1 \geq 10)}(x_1) + 0.15I_{(x_1 \geq 0.7)},
\]

and

\[
F_2(x_2) = 0.2I_{(x_2 \geq 0.3)} + 0.5I_{(x_2 \geq 10)}(x_2) + 0.3I_{(x_2 \geq 0.6)}.
\]

Here \(p_{11} = 0.25, p_{12} = 0.15, p_{21} = 0.2, p_{22} = 0.3, D_{11} = 0.2, D_{12} = 0.7, D_{21} = 0.3, \) and \(D_{22} = 0.6\). For now, we select \(\mathcal{A} = \{3, 5, 10, 15, \infty\}\) for illustrative purposes. \(\mathcal{A}\) contains more small values since the \(t\) distributions are most different for the smaller values.

The sample size \(n\) is taken from \(n = 500\) to \(n = 2500\) in increment of 500. For each \(n\), the experiment was repeated 50 times. In each experiment, the estimate of \(\rho_0\) for each \(\nu \in \mathcal{A}\) was found using the EM algorithm described in the previous section. Subsequently, \(\hat{\nu}\) was selected according to the criteria stated in (18). The simulation results, including percentage of times (out of 50) of getting the true value of \(\nu_0\), mean of \(\hat{\rho}(\hat{\nu})\), and the \(MSE\) of \(\hat{\rho}(\hat{\nu})\), are presented in Table 1.

<table>
<thead>
<tr>
<th>Sample size (n)</th>
<th>Percentage of times (\hat{\nu} = \nu_0)</th>
<th>Mean (\hat{\rho}(\hat{\nu}))</th>
<th>(MSE(\hat{\rho}(\hat{\nu})) \times 10^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>86%</td>
<td>0.7361</td>
<td>0.6312</td>
</tr>
<tr>
<td>1000</td>
<td>94%</td>
<td>0.7392</td>
<td>0.3869</td>
</tr>
<tr>
<td>1500</td>
<td>100%</td>
<td>0.7413</td>
<td>0.3808</td>
</tr>
<tr>
<td>2000</td>
<td>100%</td>
<td>0.7429</td>
<td>0.3035</td>
</tr>
<tr>
<td>2500</td>
<td>100%</td>
<td>0.7450</td>
<td>0.2861</td>
</tr>
</tbody>
</table>

Table 1: Simulation results for Experiment 1 with \(\nu_0 = \infty\) and \(\rho_0 = 0.75\).

In the second experiment, we took \(D = 2, \nu_0 = 10\) degrees of freedom corresponding to a \(t\)-copula. The two marginal distributions are taken to be the same marginals as in Experiment 1. The correlation parameter \(\rho_0\) is selected to be 0.20 and 0.75, respectively. The results are presented in Table 2. From the entries of Table 1 and Table 2, we see that the estimation procedure is more effective in selecting the true degrees of freedom when \(\nu_0 = \infty\) compared to \(\nu_0 = 10\), the reason being that the distribution corresponding to \(\nu_0 = \infty\) is further away from all the other candidate distributions in \(\mathcal{A}\). Also, the estimation procedure is less effected by the value of \(\rho_0\) as illustrated by the percentage of times \(\hat{\nu} = \nu_0\) column in Table 2.

<table>
<thead>
<tr>
<th>Sample size (n)</th>
<th>Percentage of times (\hat{\nu} = \nu_0)</th>
<th>Mean (\hat{\rho}(\hat{\nu}))</th>
<th>(MSE(\hat{\rho}(\hat{\nu})) \times 10^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>84%</td>
<td>0.1861</td>
<td>0.7532</td>
</tr>
<tr>
<td>1000</td>
<td>88%</td>
<td>0.1877</td>
<td>0.6871</td>
</tr>
<tr>
<td>1500</td>
<td>92%</td>
<td>0.1889</td>
<td>0.6095</td>
</tr>
<tr>
<td>2000</td>
<td>94%</td>
<td>0.1923</td>
<td>0.4173</td>
</tr>
<tr>
<td>2500</td>
<td>100%</td>
<td>0.1944</td>
<td>0.3664</td>
</tr>
</tbody>
</table>

Table 2: Simulation results for Experiment 2 with \(\nu_0 = 10\). The two correlation values considered are \(\rho_0 = 0.2\) and \(\rho_0 = 0.75\).

In the third experiment, we took \(D = 3\). The first

<table>
<thead>
<tr>
<th>Sample size (n)</th>
<th>Percentage of times (\hat{\nu} = \nu_0)</th>
<th>Mean (\hat{\rho}(\hat{\nu}))</th>
<th>Mean (\hat{\rho}_2(\hat{\nu}))</th>
<th>Mean (\hat{\rho}_3(\hat{\nu}))</th>
<th>Total (MSE) (\times 10^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>85%</td>
<td>0.1990</td>
<td>0.2760</td>
<td>0.1885</td>
<td>0.7143</td>
</tr>
<tr>
<td>1000</td>
<td>90%</td>
<td>0.1830</td>
<td>0.2805</td>
<td>0.2005</td>
<td>0.6732</td>
</tr>
<tr>
<td>1500</td>
<td>95%</td>
<td>0.1795</td>
<td>0.2845</td>
<td>0.1970</td>
<td>0.6714</td>
</tr>
<tr>
<td>2000</td>
<td>100%</td>
<td>0.1820</td>
<td>0.2880</td>
<td>0.1880</td>
<td>0.6126</td>
</tr>
<tr>
<td>2500</td>
<td>100%</td>
<td>0.1955</td>
<td>0.2920</td>
<td>0.1930</td>
<td>0.1335</td>
</tr>
</tbody>
</table>

Table 3: Simulation results for Experiment 3.
two marginal distributions are same as before. The third marginal distribution is taken to be a t-distribution with 10 degrees of freedom (thus, having no points of discontinuity). We took the correlation matrix \( R_0 \) as:

\[
R_0 = \begin{pmatrix}
1 & 0.2 & 0.3 \\
0.2 & 1 & 0.2 \\
0.3 & 0.2 & 1
\end{pmatrix}_{3 \times 3} = \begin{pmatrix}
1 & \rho_1 & \rho_2 \\
\rho_1 & 1 & \rho_3 \\
\rho_2 & \rho_3 & 1
\end{pmatrix}_{3 \times 3}.
\]

For different sample sizes, the experimental were repeated 20 times instead of 50 times to reduce computational time. The estimates of \( \rho_i \), \( \hat{\rho}_i \), \( i = 1, 2, 3 \), were obtained based on the iterative procedure outlined in Section 3. Since the maximization step involves another loop within the M-step, the objective function was maximized over the \( \rho \)-intervals in steps of 0.01 to reduce computational time (see Huang and Dass (2007) for details). The iterative procedure within the M-step was not required when \( D = 2 \) which enabled us to maximize the objective function over a finer grid (steps of 0.0001). The results are given in Table 3; note that (i) the estimates converge and (ii) the MSE reduces as \( n \) tends to infinity.

5.2 Real Data

Our first experiment was carried out for the MSU-Multimodal database, consisting of multiple samples of the face, fingerprint and hand-geometry biometric collected in our lab, see PRIP. Here, we focus on the fingerprint biometric of the MSU-Multimodal database. The fingerprint biometric consists of 8 samples per user for 100 users. Features for matching a pair of fingerprints consist of minutiae locations and directions extracted from the fingerprint images. The genuine and impostor matching scores of the fingerprint biometric were generated using two different matchers. Recall that a genuine score \( S_k(Q, T) \), \( k = 1, 2 \), is obtained if \( Q \) and \( T \) are samples from the same user. Thus, the total number of genuine scores is 2,800, and we can set \( X_j = (S_1(Q, T), S_2(Q, T))^T \) for \( j = 1, \ldots, n = 2,800 \).

In a similar fashion, an impostor score \( S_k(Q, T) \) is obtained if \( Q \) and \( T \) are samples from different users, making the total of impostor scores for this database to be 4,950. For the impostor matching scores, we can set \( X_j = (S_1(Q, T), S_2(Q, T))^T \) with \( j = 1, \ldots, n = 4,950 \). Figure 2 gives histograms of the genuine and impostor matching score distributions for both matchers based on the observed matching scores. Note that each marginal distribution in Figure 2 is highly non-Gaussian. Further, the two matchers were applied on the same database, and therefore, we expect the matching scores to be highly correlated. Both matchers output the discrete score ‘0’ if certain “initial conditions” are not met, resulting in a spike at 0 in the four histograms (with the most pronounced spike in panel (c)). For the reasons mentioned above, joint distribution functions of the form (8) with \( D = 2 \) are appropriate for strongly correlated biometric data as we have here. For both the genuine and impostor

![Figure 2: Histograms of matching scores.](image-url)
The joint distributions estimated from the genuine and impostor matching scores in the previous section can be used to develop fusion rules for a multimodal biometric system (biometric system with two or more modalities).

Since we assume that the genuine and impostor distributions are of the form (8), the test of hypotheses (1) can be re-stated in terms of the matching score distributions under $H_0$ and $H_1$, namely,

$$H_0 : F_Q(x) = F^{D}_{\nu_0, R_0}(x) \quad \text{vs.} \quad H_1 : F_Q(x) = F^{D}_{\nu_1, R_1}(x)$$

(21)

for some $(\nu_0, R_0)$ and $(\nu_1, R_1)$. For the testing in (21), the optimal decision rule (fusion rule) turns out to be the likelihood ratio

$$LR(x) = \frac{dF^{D}_{\nu_0, R_0}(x)}{dF^{D}_{\nu_1, R_1}(x)}$$

(22)

from the Neyman-Pearson Lemma. However, the fusion rule $LR$ cannot be used in the current form since the parameters $\nu_0, R_0, \nu_1$ and $R_1$ are unknown. The methodology developed here can be used to obtain estimates of all of these parameters, thus, obtaining the estimated likelihood ratio statistic

$$\hat{LR}(x) = \frac{dF^{D}_{\hat{\nu}_0, \hat{R}_0}(x)}{dF^{D}_{\hat{\nu}_1, \hat{R}_1}(x)}.$$

(23)

The effectiveness of the (estimated) LR fusion rule is evaluated based on a $K$-fold cross validation procedure. In the $k$-th iteration, $k = 1, \ldots, K$, a random subset of $n_0 < n$ genuine matching scores, say, $S_0$, is selected for estimating the parameters $\nu_0$ and $R_0$. Similarly, a (random) subset $S_1$ of $n_1 < n$ impostor scores are selected to estimate the parameters $\nu_1$ and $R_1$. The remaining genuine and impostor scores are used to obtain an estimate of the false accept and genuine accept rates (FAR and GAR, respectively) for each threshold $\lambda$. The relevant formulas are

$$FAR(\lambda) = \frac{1}{n-n_1} \sum_{j \in S_0^c} I_{\{LR(x_j) > \lambda\}}$$

and

$$GAR(\lambda) = \frac{1}{n-n_0} \sum_{j \in S_1^c} I_{\{LR(x_j) > \lambda\}},$$

where $S_0^c$ and $S_1^c$ are the complements of $S_0$ and $S_1$, respectively. The ROC (Receiver Operating Characteristics) curve is the plot of $FAR(\lambda)$ versus $GAR(\lambda)$ with higher ROC values indicating better recognition performance.

Our experiments were carried out on the MSU-Multimodal and NIST databases discussed in Section 5.2. We set $K = 10$. For each $k = 1, 2, \ldots, 10$, $80\%$ (that is, $\frac{n_0}{n} = 0.8$) of the genuine and impostor matching scores were randomly selected to estimate the marginal distributions as well as $\nu$ and $R$. The remaining 20% of the genuine and impostor scores were used to generate the ROC curves. The fusion results are presented in Figure 3: panel (a) and (b), respectively, gives the ROC curve for the LR fusion rule for the MSU-Multimodal and NIST databases. Note that there is an dramatic overall improvement of the performance.

6. Summary and Conclusion

Joint distributions with mixed marginals are developed via $t$-copula functions. Estimation of the correlation matrix as well as the degrees of freedom corresponding to the $t$-copula are carried out based on the EM algorithm. We show large sample consistency results of our estimates
Figure 3: Performance of copula fusion on (a) the MSU-Multimodal database and (b) the NIST database.

and demonstrate this based on several simulation examples. Finally, the methodology is applied to real data consisting of matching scores from various biometric sources. Fusion based on the generalized distributions give superior performance compared to the individual unimodal systems. In future work, we will consider extensions to copula functions derived from general elliptical distributions.

REFERENCES


Pattern Recognition and Image Processing Lab, PRIP. Department of Computer Science And Engineering, Michigan State University. Online: http://biometrics.cse.msu.edu/.