CSE 40537 / 60537: Biometrics

Multi-Biometrics 3
Dempster-Shafer Theory of Evidence

Assign degrees of belief for uncertain events

*Degree of belief is different from the probability of the event

**Example:**

Suppose matcher is reliable 95% of the time, and unreliable 5% of the time.

We can assign a 0.95 degree of belief to a “match” decision and a 0.0 degree of belief to a “non-match” decision.

Zero belief indicates that there is no reason to believe that the input does not match successfully against the template.
Decision profile and degree of support

Degree of support provided by the $j^{th}$ matcher to the $k^{th}$ class

\[
DP = \begin{bmatrix}
    s_{1,1} & \cdots & s_{1,k} & \cdots & s_{1,M} \\
    \vdots & \ddots & \vdots & \ddots & \vdots \\
    s_{j,1} & \cdots & s_{j,k} & \cdots & s_{j,M} \\
    \vdots & \ddots & \vdots & \ddots & \vdots \\
    s_{R,1} & \cdots & s_{R,k} & \cdots & s_{R,M}
\end{bmatrix}
\]

Degree of support is expressed as:

\[
s_{j,k} = \begin{cases} 
1, & \text{if output of the } j^{th} \text{ matcher is class } y_k \\
0, & \text{otherwise}
\end{cases}
\]
A decision template is:

\[
DT_{j,r}^k = \frac{CM_{j,k,r}}{N_k}
\]

confusion matrix from matcher

number of elements in class

Compute decision profile \( DP^t \) after the decisions of \( R \) matchers are obtained
Decision Template Similarity

\[ \Phi_{j,k} = \frac{(1 + (||DT_j^k - DP_j^t||)^2)^{-1}}{\sum_{r=1}^{M}((1 + (||DT_j^r - DP_j^t||)^2)^{-1})} \]

\( j^{th} \) row of \( DT^k \)
belonging to class \( y^k \)

\( j^{th} \) row of \( DP^t \)
belonging to probe \( x^t \)

Matrix norm.
Degree of Belief

For each matcher, we can compute:

\[ b_{j,k} = \frac{\Phi_{j,k}[\prod_{r=1, r\neq k}^{M} (1 - \Phi_{j,r})]}{1 - \Phi_{j,k}[\prod_{r=1, r\neq k}^{M} (1 - \Phi_{j,r})]} \]

Accumulated degree of belief comes from Dempster’s rule:

\[ g_{k} = \prod_{j=1}^{R} b_{j,k} \]
Let’s walk through another example (starting with some familiar information)

Three users:

Alice  Bob  Charlie

\[
CM^1 = \begin{bmatrix}
65 & 9 & 26 \\
50 & 37 & 13 \\
32 & 32 & 36 \\
\end{bmatrix}
\]

\[
CM^2 = \begin{bmatrix}
49 & 21 & 30 \\
10 & 66 & 24 \\
10 & 19 & 71 \\
\end{bmatrix}
\]
Three decision templates

\[ DT^1 = \begin{bmatrix} 0.65 & 0.09 & 0.26 \\ 0.49 & 0.21 & 0.30 \end{bmatrix} \]

\[ DT^2 = \begin{bmatrix} 0.50 & 0.37 & 0.13 \\ 0.10 & 0.66 & 0.24 \end{bmatrix} \]

\[ DT^3 = \begin{bmatrix} 0.32 & 0.32 & 0.36 \\ 0.10 & 0.19 & 0.71 \end{bmatrix} \]
What’s the decision profile?

Assume for a test image, matcher outputs are \( c_1 = \text{Alice} \) and \( c_2 = \text{Bob} \)

\[
DP^t = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}
\]
Similarity between decision profile and decision templates

\[ \Phi = \begin{bmatrix} 0.39 & 0.33 & 0.28 \\ 0.28 & 0.47 & 0.25 \end{bmatrix} \]

Then compute degree of belief:

\[ b = \begin{bmatrix} 0.23 & 0.17 & 0.13 \\ 0.13 & 0.34 & 0.11 \end{bmatrix} \]
Discriminant functions for the three classes

\[ g_{Alice} = 0.23 \times 0.13 = 0.03 \]
\[ g_{Bob} = 0.17 \times 0.34 = 0.06 \]
\[ g_{Charlie} = 0.13 \times 0.11 = 0.01 \]
Behavior Knowledge Space

Example: 100 face images from each user are used to construct a look-up table

<table>
<thead>
<tr>
<th>Possible decision vectors ([c_1, c_2])</th>
<th>Number of training instances from each user</th>
<th>Decision label after fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice, Alice</td>
<td>30, 2, 4</td>
<td>Alice</td>
</tr>
<tr>
<td>Alice, Bob</td>
<td>12, 38, 5</td>
<td>Bob</td>
</tr>
<tr>
<td>Alice, Charlie</td>
<td>23, 10, 23</td>
<td>Alice, Charlie</td>
</tr>
<tr>
<td>Bob, Alice</td>
<td>0, 0, 0</td>
<td>Alice, Bob, Charlie</td>
</tr>
<tr>
<td>Bob, Bob</td>
<td>6, 25, 11</td>
<td>Bob</td>
</tr>
<tr>
<td>Bob, Charlie</td>
<td>3, 12, 21</td>
<td>Charlie</td>
</tr>
<tr>
<td>Charlie, Alice</td>
<td>19, 8, 6</td>
<td>Alice</td>
</tr>
<tr>
<td>Charlie, Bob</td>
<td>3, 3, 3</td>
<td>Alice, Bob, Charlie</td>
</tr>
<tr>
<td>Charlie, Charlie</td>
<td>4, 2, 27</td>
<td>Charlie</td>
</tr>
</tbody>
</table>

Image Credit: *Handbook of Multibiometrics* by A. Ross, K. Nandakumar, and A.K. Jain
Score Level Fusion
Bayesian Decision Theory

Given the measurements $x_j, j = 1, \ldots, R$, the input pattern $X$ should be assigned to the class $y_r$ that maximizes the posterior probability:

Assign $X \rightarrow y_r$ if

$$P(y_r | x_1, \ldots, x_R) \geq P(y_k | x_1, \ldots, x_R)$$

$$k = 1, \ldots, M$$

Minimum error-rate classification rule

Duda et al. *Pattern Classification*, 2001
zero-one loss function
Apply Bayes’ rule

\[ P(y_k | x_1, \ldots, x_R) = \frac{p(x_1, \ldots, x_R | y_k)P(y_k)}{\sum_{l=1}^{M} p(x_1, \ldots, x_R | y_l)P(y_l)} \]


Approximations of the above lead to five classifier combination strategies
Assumption: statistical independence of feature representations

Conditional joint probability density

\[ p(x_1, \ldots, x_R | y_k) = \prod_{j=1}^{R} p(x_j | y_k) \]

Product of marginal conditional densities

How correct is this assumption?
Product Rule

If genuine and impostor classes are assigned equal probabilities:

Assign $X \rightarrow y_r$ if

$$\prod_{j=1}^{R} P(y_r|x_j) \geq \prod_{j=1}^{R} P(y_k|x_j)$$
Limitations of product rule

Sensitivity to errors in the estimations of the posterior probabilities

If one classifier outputs a score close to zero, the product will be small and lead to a bad decision

\[
\begin{align*}
0.01 \times 0.92 \times 0.89 &= 0.008
\end{align*}
\]
Sum Rule

If genuine and impostor classes are assigned equal probabilities:

 Assign $X \rightarrow y_r$ if

$$
\sum_{j=1}^{R} P(y_r | x_j) \geq \sum_{j=1}^{R} P(y_k | x_j)
$$
Strength of sum rule

The sum rule is not highly impacted by errors estimated in the posterior probabilities

Commonly used in multi-biometric systems
Max Rule

If genuine and impostor classes are assigned equal probabilities:

\[
\text{Assign } X \rightarrow y_r \text{ if } \\
\max_{j=1}^{R} P(y_r | x_j) \geq \max_{j=1}^{R} P(y_k | x_j)
\]
Min Rule

If genuine and impostor classes are assigned equal probabilities:

Assign $X \rightarrow y_r$ if

$$
\min_{j=1}^{R} P(y_r|x_j) \geq \min_{j=1}^{R} P(y_k|x_j)
$$
Median Rule

Sum rule is not robust in a statistical sense; it is impacted by outliers

If genuine and impostor classes are assigned equal probabilities:

Assign $X \rightarrow y_r$ if

$$\text{median}_{j=1}^{R} P(y_r | x_j) \geq \text{median}_{j=1}^{R} P(y_k | x_j)$$
Relationship between probabilities and match scores

The classifier rules only work if the output of the biometric systems is in the form \( P(y_k|x_j) \)

Most biometric matchers only output a match score \( s_{j,k} \)

Verlinde et al. Int. Symp. on Pattern Recognition 1999

\[
s_{j,k} = g(P(y_k|x_j)) + \beta(x_j)
\]

monotonic function error made by biometric matcher
Score conditions

Good scenario for fusion:

\[ \beta = 0 \]

Reasonable to approximate \( P(y_k|x_j) \) by \( P(y_k|s_{j,k}) \)

Bad scenario for fusion:

\[ \beta \neq 0 \]

Not possible to directly apply the classifier combination rules
Score normalization

Scores from different matchers need to be placed on a common basis before fusion

0.5677  101.14  3.4511

Change the location and scale of the score distributions
Types of score normalization

**Fixed score normalization:** normalization parameters are determined using fixed training set

**Adaptive score normalization:** normalization parameters are estimated based on the match score of the current sample
Min-Max

- Simplest approach
- Can only be used when the bounds of the scores produced by a matcher are known

\[ n s^t_j = \frac{s^t_j - \min_{i=1}^N s^i_j}{\max_{i=1}^N s^i_j - \min_{i=1}^N s^i_j} \]

min-max normalized score
Decimal Scaling

- Can be applied when scores are on a logarithmic scale
- Example: Matcher 1 [0,10], Matcher 2 [0, 1000]

\[ ns_j^t = \frac{s_j^t}{10^{n_j}} \]

\[ n_j = \log_{10} \max_{i=1}^{N} s_j^i \]
Z-score

- Most commonly used technique
- Simple Gaussian normalization with parameters estimated from a training set

\[ n s_{j}^t = \frac{s_t^j - \mu_j}{\sigma_j} \]

\(n\) is the mean, \(s_t^j\) is the standard deviation, and \(\mu_j\) and \(\sigma_j\) are parameters estimated from a training set.
Median and MAD (median absolute deviation)

- Insensitive to outliers and points in the extreme tails of the distribution
- Bad when score distribution is not Gaussian

\[ \text{median} \sum_{i=1}^{N} s_j^i \]

\[ n s_j^t = \frac{s_j^t - \text{med}_j}{\text{MAD}_j} \]

\[ \text{median} \sum_{i=1}^{N} |s_j^i - \text{med}_j| \]
Double Sigmoid

- For multibiometric systems that combine different fingerprint matchers

\[
n s_j^t = \begin{cases} 
  \frac{1}{1+\exp\left(-2\left(\frac{s_j^t - \tau}{\alpha_1}\right)\right)} & \text{if } s_j^t < \tau, \\
  \frac{1}{1+\exp\left(-2\left(\frac{s_j^t - \tau}{\alpha_2}\right)\right)} & \text{otherwise}
\end{cases}
\]

\(\alpha_1, \alpha_2\): left and right edges of the region where the function is linear

reference operating point
Double Sigmoid

\[ \tau = 200, \alpha_1 = 20, \alpha_2 = 30 \]

Image Credit: Handbook of Multibiometrics by A. Ross, K. Nandakumar, and A.K. Jain
tanh-estimators

- Robust and highly efficient
- Requires careful parameter estimation

\[ n s^t_j = \frac{1}{2} \left\{ \tanh \left( 0.01 \left( \frac{s^t_j - \mu_{GH}}{\sigma_{GH}} \right) \right) + 1 \right\} \]

mean of genuine score distribution given by Hampel estimators

standard deviation of genuine score distribution given by Hampel estimators
Hampel Estimators

\[ \psi(u) = \begin{cases} 
  u & 0 \leq |u| < a, \\
  a \times \text{sign}(u) & a \leq |u| < b, \\
  a \times \text{sign}(u) \times \left( \frac{c-|u|}{c-b} \right) & b \leq |u| < c, \\
  0 & |u| \geq c 
\end{cases} \]

where

\[ \text{sign}\{u\} = \begin{cases} 
  +1, & \text{if } u \geq 0, \\
  -1, & \text{otherwise.} 
\end{cases} \]
Hampel influence function

Image Credit: *Handbook of Multibiometrics* by A. Ross, K. Nandakumar, and A.K. Jain
Summary of normalization techniques

<table>
<thead>
<tr>
<th>Normalization Technique</th>
<th>Robustness</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min-max</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>Decimal scaling</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>Z-score</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>Median and MAD</td>
<td>Yes</td>
<td>Moderate</td>
</tr>
<tr>
<td>Double sigmoid</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Tanh-estimators</td>
<td>Yes</td>
<td>High</td>
</tr>
</tbody>
</table>

Image Credit: *Handbook of Multibiometrics* by A. Ross, K. Nandakumar, and A.K. Jain