Comp. by: Amohamed Stage: Galleys ChapterID: 0000872383 Date:22/1/09 Time:18:11:10

M

Au2

Multiple Experts

JOSEF BIGUN
Halmstad University, IDE, Halmstad, Sweden

Synonyms

Decision fusion; Feature fusion; Fusion; Multimodal fusion; Score fusion

Definition

A *Biometric expert*, or an expert in biometric recognition context, refers to a method that expresses an opinion on the likelihood of an identity by analyzing a signal that it is specialized on, e.g. a fingerprint expert using minutiae, a lip-motion expert using statistics of optical-flow. Accordingly, there can be several experts associated with the same sensor data, each analyzing the data in a different way. Alternatively, they can be specialized on different sensor data. Multiple experts can address the issue of how expert opinions should be represented and reconciled to a single opinion on the authenticity of a ► client identity.

Introduction

In biometric signal analysis, the fusion of multiple experts can in practice be achieved as ► feature fusion or *score fusion*. In addition to these, one can also discern *data fusion*, e.g. stereo images of a face, and *decision fusion*, e.g. the decisions of several experts wherein each expresses either of the crisp opinions "client" or "► impostor," in the taxonomy of fusion [1]. However, one can see data fusion and decision fusion as adding novel experts and as a special case of score fusion, respectively. On the other hand, feature fusion is often achieved as concatenation of feature vectors, which is in turn modeled by an expert suitable for the processing demands of the set of the novel vectors. For this reason we only discuss score fusion in this article. The initial frameworks for fusion have been simplistic in that no knowledge on the skills of the experts is used by the ► supervisor. Later efforts to reconcile different expert opinions in a multiple experts biometric system have been described from a probabilistic opinion modeling [2] and a pattern discrimination [3], view points, respectively. From both perspectives, it can be concluded that the weighted average is a good way of reconciling different authenticity scores of individual experts to a single opinion, under reasonable conditions. As the weights reflect the skills of the experts, some sort of training is needed to estimate them. Belonging to probabilistic modeling school, respective discriminant analysis school, Bayesian modeling [4, 5], and support vector machines [6–8] have been utilized to fuse expert opinions. An important issue for a fusion method is, however, whether or not it has mechanisms to discern the general skill of an expert from the quality of the current data. We summarize the basic principles to exemplify typical fusion approaches as follows.

Simple Fusion

This type of fusion applies a rule to input opinions delivered by the experts. The rule is not obtained by training on expert opinions, though they might very well be decided by the human designer of the supervisor with her knowledge of expert skills. Assuming that the supervisor receives all expert inputs in parallel, common simple fusions include, Au1

max	Maximum of the scores,	$M_j = \max(x_{1,j}, x_{2,j}, \ldots, x_{m,j})$
min	Minimum of the scores,	$M_j = \min(x_{1,j}, x_{2,j}, \ldots, x_{m,j})$
sum	Arithmetic mean of the scores,	$M_j = \frac{1}{m}\sum_{i=1}^{m} x_{i,j}$
median	Median of the scores,	$M_j = x_{(\frac{m+1}{2})j}$
Product	Geometric mean of the scores,	$M_j = (\prod_{i=1}^{m} x_{i,j})^{\frac{1}{m}}$

Comp. by: Amohamed Stage: Galleys ChapterID: 0000872383 Date:22/1/09 Time:18:11:11

where M_j is the score output by the supervisor at the instant of operation j, when m expert opinions, $x_{i,j}, i: 1 \ldots m$, are available to it. In addition to a parallel application of a single simple fusion to all expert opinions, one can apply several simple fusion rules serially (one after the other) if some expert opinions are delayed before they are processed by the supervisor(s).

Probabilistic Fusion

Experts can express opinions in various ways. The simplest is to give a strict decision on a claim of an identity, "1" (client) or "0" (impostor). A more common way is to give a graded opinion, usually a real number in [0, 1]. However, it turns out that machine experts can benefit from a more complex representation of an opinion, an array of real variables. This is not surprising to human experience because, a human opinion is seldom so simple or lacks variability that it can be described by what a single variable can afford. A richer representation of an opinion is therefore the use of the distribution of a score rather than a score. Bayes theory is the natural choice in this case because it is about how to update knowledge represented as distribution (prior) when new knowledge (likelihood) becomes available.

Before describing a particular way of constructing a Bayesian supervisor let us illustrate the basic mechanism of Bayesian updating. Let two stochastic variables X_1, X_2 represent the errors of two different measurement systems measuring the same physical quantity. We assume that the errors are independent and are distributed normally as $N(0, \sigma_1^2)$, $N(0, \sigma_2^2)$, respectively. Then their weighted average

$$M = q_1 X_1 + q_2 X_2, \qquad \text{where} \quad q_1 + q_2 = 1 \qquad (1)$$

is also normally distributed with $N(0, {q_1}^2\sigma_1^2 + {q_2}^2\sigma_2^2)$. Given the variances σ_1^2, σ_2^2, if the weights q_1, q_2 are chosen inversely proportional to the respective variances, the variance of the new variable M (the weighted mean) will be smallest provided that

$$q_1 = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \qquad q_2 = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \qquad (2)$$

Notice that the composite variable M is normally distributed always if the X_1, X_2 are independent but the variance is smallest only for a particular choice, (seen earlier) yielding

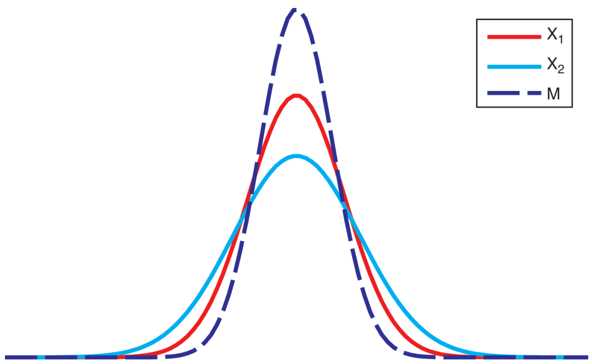
$$\mathrm{var}(M) = q_1^2\sigma_1^2 + q_2^2\sigma_2^2 = \frac{\frac{1}{\sigma_1^4}}{\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^2}\sigma_1^2 + \frac{\frac{1}{\sigma_2^4}}{\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^2}\sigma_2^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \leq \min(\sigma_1^2, \sigma_2^2) \quad (3)$$

The fact that the composite variance never exceeds the smallest of the component variances, and that it converges to the smallest of the two when either becomes large, i.e. one distribution approaches the noninformative distribution $N(0, \infty)$, can be exploited to improve the precision of the aggregated measurements, Fig. 1.

Appropriate weighting is the main mechanism on how knowledge as represented by distributions can be utilized to improve biometric decision making. Bayes theory comes handy at this point because it offers the powerful Bayes theorem to estimate the weights for the aggregation of the distributions, incrementally, or at one-go as new knowledge becomes available. We follow [4] to exemplify this approach. Let the following list describe the variables representing the signals made available by a multiexpert biometric system specialized in making decisions. Next we will discuss the errors of the experts on client and impostor data separately.

- i: Index of the experts. $i \in 1 \ldots m$,
- j: Index of shots (one or more per candidate), $j \in 1 \ldots n$, n_T. It is equivalent to time since an expert has one shot per evaluation time (period). The time n is the last instant in the training whereas n_T is the test time when the system is in operation.
- X_{ij}: The authenticity score, i.e. the score delivered by expert i on shot j's claim of being a certain client
- Y_j The true authenticity score of shot j's claim being a certain client. This variable can take only two numerical values corresponding to "True" and "False"
- Z_{ij} The miss-identification score, that is $Z_{ij} = Y_j - X_{ij}$
- S_{ij} The variance of Z_{ij} as estimated by expert i

One can model the errors (not the scores) that a specific expert makes when it encounters clients. To this end, assume that $Y_i = 1$ and that the conditional stochastic variable Z_{ij} given its expectation value



Multiple Experts. Figure 1 The component distributions with $X_1 \sim N(0,1)$, $X_2 \sim N(0,1.3^2)$ and the composite distribution $M \sim N(0, 1/(1.3^{-2}+1))$, (1).

b_i is normally distributed i.e. $(Z_{ij}|\, b_i) \sim N(b_i, \sigma_{ij})$. If Z_{ij} are independent then, according to Bayes theory, the *posterior* distribution $(b_i|\, z_{ij})$, will also be normal

$$(b_i|z_{ij}) \sim N(M_i^C, V_i^C) \tag{4}$$

with mean and variance

$$M_i^C = \frac{\sum_{j=1}^{n^C} \frac{z_{ij}}{\sigma_{ij}^2}}{\sum_{j=1}^{n^C} \frac{1}{\sigma_{ij}^2}} \quad \text{and} \quad V_i^C = \left(\sum_i^{n^C} \frac{1}{\sigma_{ij}^2} \right)^{-1} \tag{5}$$

respectively. In this updating, we see the same pattern as in the example, (1–3). Here C is a label that denotes that the applicable variables relate to clients. This distribution at hand, one can now estimate b_i as the expectation of $(b_i|\, z_{ij})$ which is M_i^C. In this derivation, we updated a noninformative prior distribution, $b_i \in N(0, \infty)$, i.e. "nothing is known about b_i" to obtain the posterior distribution $(b_i|\, Z_{ij}) \in N(M_i^C, V_i^C)$. The resulting distribution is a Gaussian function which attempts to capture the bias of each expert, as well as the precision of each expert, which together represent its skills.

We proceed next to use the observed knowledge about an expert to obtain an unbiased estimate of its score distribution at the time instant $j = n_T$. By re-applying Bayes theorem to update the distribution given in (4) one obtains that,

$$(Y_{n_T}|z_{i,1}, z_{i,2}, \ldots, z_{i,n^C}, x_{i,n_T}) \in N(M_i'^C, V_i'^C) \tag{6}$$

with mean and variance

$$M_i'^C = x_{i,n_T} + M_i^C \quad \text{and} \quad V_i'^C = V_i^C + \sigma_{i,n_T}^2. \tag{7}$$

Consider now the situation that m independent experts have delivered their authenticity scores on supervisor-training shots ($j = 1, 2, \ldots, n^C$) and the test shot n_T. Using the Bayesian updating again, the posterior distribution of b_i, given the scores at the instant $j = n_T$ and the earlier errors, is normal;

$$\begin{aligned}(Y_{n_T}|z_{1,1}, \ldots, z_{1,n^C}, x_{1,n_T}, \ldots, z_{m,1}, \ldots z_{m,n^C}, x_{m,n_T}) \\ \in N(M_i''^C, V_i''^C)\end{aligned} \tag{8}$$

where

$$M''^C = \frac{\sum_{i=1}^{m} \frac{M_i'^C}{V_i'^C}}{\sum_{i=1}^{m} \frac{1}{V_i'^C}} \quad \text{and} \quad V''^C = \left(\sum_{i=1}^{m} \frac{1}{V_i'^C} \right)^{-1} \tag{9}$$

However, to compute these means and variances, the score variances σ_{ij}^2 are needed. We suppose that these estimations are delivered by experts depending on, e.g. the quality of the current biometric sample underlying their scores. This is reasonable because not all samples have the same (good) quality, influencing the precision of the observed score x_{ij}. In case this is not practicable for various reasons, one can assume that x_{ij} has the same variance within an expert i (but allow it to vary between experts). Then, the variances of the distributions of x_{ij} need not be delivered to the supervisor, but can be estimated by the supervisor, as discussed in the following section. Before one can use the distribution $N(M''^C, V''^C)$ as a supervisor, one needs to compare it with the distribution obtained by an alternative aggregation.

Assume now that we perform this training with n^I impostor samples ($Y_j = 0$) i.e. that we compute the bias distribution $N(M_i^{I}, V_i^{I})$ when expert i evaluates impostors, and the final distribution $N(M''^I, V''^I)$, with I being a label denoting "Impostor." We do not write the update formulas explicitly as these are identical to (5,7,9) except that the training set consists of impostors, for simplicity. One of the two distributions $N(M''^C, V''^C)$, and $N(M''^I, V''^I)$, represents the true knowledge better than the other at the test occasion, $j = n_T$. At this point one can choose the distribution that achieves a resemblance that is most bona-fide to its role, e.g.

$$M'' = \begin{cases} M''^C, & \text{if} \quad 1 - M''^C \le M''^I; \\ M''^I, & \text{otherwise.} \end{cases} \tag{10}$$

In other words if the client-supervisor has a mean closer to its goal (one, because $Y_j = 1$ represents client) than the impostor-supervisor's mean is to its goal (zero) then the choice falls on the distribution coming from the client-supervisor and vice-versa. An additional possibility is to reject to output a distribution in case the two competing distributions overlap more than a desired threshold. One could also think of a hypersupervisor to reconcile the two antagonist ► supervisor opinions.

In practice most experts can deliver scores that are between 0 and 1. However, there is a formal incompatibility of this with our assumptions because the distributions of Z_{ij} would be limited to the interval $[-1,1]$ whereas the concept we discussed earlier is based on normal distributions taking values in$]-\infty,\infty[$. This is a classical problem in statistics and is addressed typically by remapping the scores so that one works with "odds" of scores

$$X_{ij} = \log \frac{X'_{ij}}{1 - X'_{ij}} \tag{11}$$

where $X_{ij}' \in]0,1[$. It can be shown that the supervisor formula (10) and its underlying updating formulas hold for X_{ij}' as well. The only difference is in the conditional distributions which will be log normal yielding, in particular, the expected value $\exp(M'' + V''/2)$ and the variance $\exp(2M'' + 2V'') - \exp(2M'' + V'')$ for Y_{n_T}, (8).

Quality estimations for Bayesian supervisors. There are various ways to estimate the variance of a score distribution associated with a particular biometric sample on which an expert expresses an opinion of authenticity. The Bayesian supervisor expects this estimate because it works with distributions to represent the knowledge/opinion concerning the current sample as well as the past experience, not scalars. It makes most sense that this information is delivered by the expert or by considering the quality of the score. Next we discuss how these can be entered into update formulas.

One can assume that the experts give the precisions correctly except for an individual proportionality constant.

$$s_{ij} = a_i \sigma_{ij}^2 \tag{12}$$

Applying the Bayes theory again, i.e. a_i is first modeled to be a distribution rather than a scalar, then the distribution of $(a_i | (z_{i,1}, s_{i,1}), \ldots, (z_{i,n}, s_{i,n}))$ can be computed (it is a beta distribution under reasonable assumptions [4]). In turn this allows one to estimate the conditional expectation of $\frac{1}{a_i}$, yielding a Bayesian estimate of the score-error variances

$$\begin{aligned} \bar{\sigma}_{i,j^2} &= E(\sigma^2_{in_T} | s_{in_T}, (z_{i,1}, s_{i,1}), \cdots (z_{i,n}, s_{i,n})) \\ &= s_{ij} E\left(\frac{1}{a_i}\right) = s_{ij}\alpha_i = s_{ij} \frac{(G_i - D_i)}{n-3} = \end{aligned} \tag{13}$$

with

$$\alpha_i = E\left(\frac{1}{a_i}\right) = \frac{(G_i - D_i)}{n-3}, \qquad G_i = \sum_{j=1}^{n} \left(\frac{z_{ij}^2}{s_{ij}}\right) \quad \text{and}$$

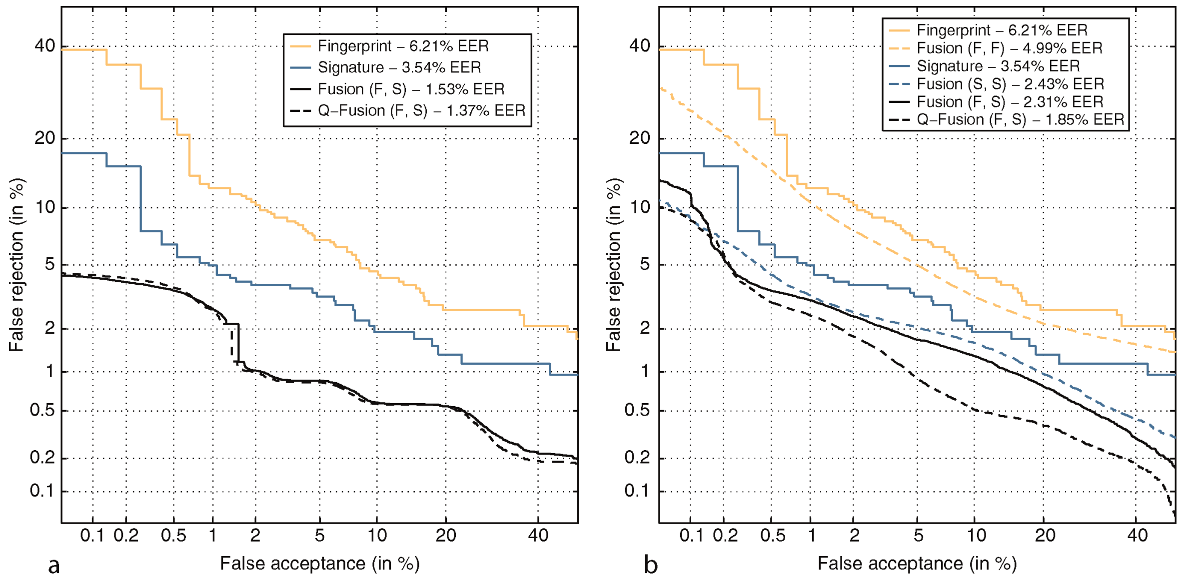
$$D_i = \left(\sum_{j=1}^{n} \left(\frac{z_{ij}}{s_{ij}}\right)\right)^2 \left(\sum_{j=1}^{n} \left(\frac{1}{s_{ij}}\right)\right)^{-1} \tag{14}$$

Note that, n will normally represent the number of biometric samples in the training set and equals to either n^C or n^I. From this result it can also be concluded that if an expert is unable to give a differentiated quality estimation then its variance estimation s_{ij} will be constant across the biometric samples it inspects and the $\bar{\sigma}_{ij}^2$ will approach gracefully to the variance of the error of the scores of the expert (not adjusted to sample quality).

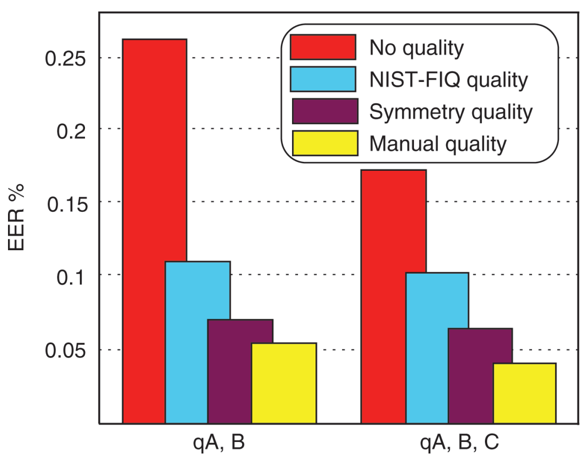
The machine expert will, in practice, be allowed to deliver an empirical quality score p_{ij} because these are easier to obtain than variance estimations, s_{ij}. At this point, one can assert that these qualities are inversely proportional to the underlying standard deviations of the score distributions, yielding

$$s_{ij} = \frac{1}{p_{ij}^2} \tag{15}$$

where p_{ij} is a quality measure of the biometric sample j as estimated by the expert i. If it is a human expert that estimates the quality p_{ij} it can be the length of the interval in which she/he is willing to place the score x_{ij}, so that even human and machine opinions can be reconciled by using the Bayesian supervisor. In Fig. 2 (1), the performance of using this Bayes supervisor in a recognition system relying on a fingerprint and a signature expert is shown. The quality scores are generated by human experts independent of the experiment. To automatically find quality scores p_{ij} for biometric samples is an emerging field of study [8–11]. The results of Bayes supervisor are illustrated by Fig. 3 where three fingerprint (machine) experts' opinions



Multiple Experts. Figure 2 The graphs, from [12], illustrate the recognition performance of two supervisors on the same data-set (1) a probabilistic supervisor (Bayesian) (2) a discrimination based supervisor (SVM). The used experts were common, F: fingerprint, and S: signature.



Multiple Experts. Figure 3 The graphs, from [11], illustrate the recognition performance of two (qA, B) and three (qA, B, C) fingerprint experts combined by the Bayesian supervisor with automatically extracted quality measures (attached to qA).

are weighted to yield the supervisor opinion. The experts are called A, B, and C and the quality measures used were (1) no quality, i.e. $p_{ij} = 1$ (2) An automatic quality measure [10], (the method is publicly made available by NIST), (3) another automatic quality measure based on local symmetries [11], (4) Quality measures provided by human experts. At each experiment, one of the four quality measures is attached to the scores of A (so that this expert is called qA) in a two or three expert configurations to evaluate the effect of using sample adaptive quality measures in machine supervisors. As can be seen, using quality measures does improve the recognition performance. It is not surprising that human experts perform better in quality estimation, as this is a very complex task in which human experts still excel. However, the machine-delivered quality estimates are fairing quite well, not too far away behind human assessments of the quality. It is also worth noting that the final decisions are suggested by the machine supervisor which processed both human and machine delivered opinion parameters transparently.

Discrimination Functions Based Fusion

Discrimination functions are frequently used in pattern classification and can also be used to fuse decisions of biometric experts. Discriminative statistics differs from Bayesian approach in that distributions of random variables are not necessary for decision making. Instead, modeling the decision boundaries is the focus of attention. Here we exemplify the use of this approach in fusion by Support Vector Machines, SVMs [3].

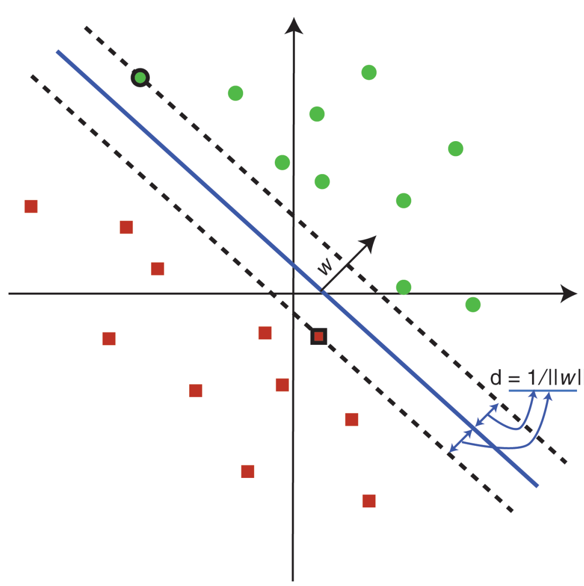
Assume that we are given a set of observations

$$\{\mathbf{x}_1, y_1\}, \{\mathbf{x}_2, y_2\} \cdots \{\mathbf{x}_n, y_n\} \qquad (16)$$

where $\mathbf{x}_j = (x_{1,j}, \ldots, x_{1,j})^T$ is a feature vector of dimension m, and y_j is the class-label of the latter (relative to the classes, "client" and "impostor"), respectively. Assume further that the two classes are separable by a hyperplane. Then there is an optimal hyperplane in a high dimensional space to which $\mathbf{x}$ is mapped. For simplicity we assume that the mapping is the (trivial) identity transformation but other tranformations using e.g. polynomials, or radial basis functions, can be used with little impact on the discussion that follows next [3], provided that appropriate kernel functions are used whenever scalar products are utilized in computations. The separation hyperplane

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = 0 \qquad (17)$$

can be made to have maximal distance d_{max} to samples belonging to the two classes, Fig. 4. The equation



Multiple Experts. Figure 4 Illustration of two classes (circle:client and square:impostor) that are separable by a hyperplane with direction **w**. The support vectors that define the separation hyperplane are represented by the outlined square and the circle on dashed hyperplanes. The width of the separation zone is $2/\|\mathbf{w}\|$ which is maximized by SVM.

can be multiplied by a nonzero constant such that $\|\mathbf{w}\| = 1/d_{max}$. We can (using this freedom) represent thereby the two class-labels as $+1$ and -1, to follow the convention of SVM litterature. Then, we have

$$\begin{cases} f(\mathbf{x}_j) \geq 1 & \text{if } y_j = 1 \\ f(\mathbf{x}_j) \leq -1 & \text{if } y_j = -1 \end{cases} \qquad (18)$$

Equivalently, the distance is maximized if $\frac{1}{2}\|\mathbf{w}\|^2$ is minimized under the constraints given by (18). If we know $\mathbf{w}$ and b, the function f will be a discrimination function, i.e. $f(\mathbf{x}) \geq 0$ prompts for a decision $y_i = 1$. The parameters $\mathbf{w}$ and b can be found by solving a quadratic problem with linear constraints.

In case the classes are not separable by a hyperplane, slack variables ξ_j are introduced so as to allow a classification that makes an error, but that this error is the smallest on the training/observation set. The corresponding problem is still a quadratic optimization problem

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + \sum_j C\xi_j \qquad (19)$$

subject to the constraints

$$\begin{cases} f(\mathbf{x}_j) \geq 1 & \text{if } y_j = 1 - \xi_j \\ f(\mathbf{x}_j) \leq -1 & \text{if } y_j = -1 + \xi_j \end{cases} \qquad (20)$$

The constant C assures that there is a limit on the amount of change the training vectors can introduce to the solution.

The SVM formulation allows one to construct a supervisor that is able to assign a class label y_j to the score vector of m experts $\mathbf{x}_j = (x_{1,j}, x_{2,j} \ldots x_{m,j})^T$ (at $j = n_T$). However, such a supervisor would not be quality adaptive yet, because the contribution of each sample to the total cost function would be uniform due to C's being a constant. This can be changed such that the cost depends on the quality of the biometric sample by means of a heuristically chosen function. In the following section we follow the description of [8] to obtain such a sample adaptation and a final supervisor.

Let the original quality measure delivered by the expert i be $p_{ij} \in [0, p_{max}]$ and

$$q_{ij} = \sqrt{p_{ij} \cdot \bar{p}_i} \qquad (21)$$

where $\bar{p}_i$ is the average quality measure of the signals that expert i delivered for the training samples, which is also in the range of $[0, p_{max}]$. Then one can train m SVMs, each having its own discrimination function

f_i, by defining the cost coefficients for the respective function as follows.

$$C_{ij} = C\left(\frac{\prod_{i' \neq i} q_{i',j}}{p_{max}^{m-1}}\right)^{\gamma_1} \quad (22)$$

The coefficient C_{ij} represents now the cost of not influencing the biometric sample j by expert i and it is measured as the product of the quality measures of other experts (excluding expert i) on the current sample j. The training samples of f_i are $\mathbf{x}_j^{\mathrm{i}}$, $j : 1 \ldots n$, which equals to $\mathbf{x}_j$ except that its component corresponding to expert i has been removed.

$$\mathbf{x}_j^i = (x_{1,j}, \cdots, x_{i-1,j}, x_{i+1,j}, \cdots x_{m,j})^T \quad (23)$$

Here the use of as superscript is in the sense of label, not exponent, signifying that the data of expert i is lacking. The exponent γ_1 is an empirically chosen constant the purpose of which is to adjust the overall influence of quality based discrimination on the final decision. In a similar fashion an additional discrimination function f_0 can be computed, except that the cost coefficients are now defined as

$$C_j = C\left(\frac{\prod_{i=1}^{m} q_{i,j}}{p_{max}^{m}}\right)^{\gamma_2} \quad (24)$$

This represents the alternative cost of using all individual quality measures including those delivered by expert i. The discrimination function f_0 is obtained by an SVM training on full length expert score vectors $\mathbf{x}_j$, $j : 1 \ldots n$, as opposed to f_i, $i : 1, \ldots, m$ which trains on $\mathbf{x}_j^{\mathrm{i}}$, lacking the opinion of expert i. When the system is operational at time $j = n_T$, the m quality scores q_{i,n_T} as well as m expert scores x_{i,n_T} are available. The quality measures q_{i,n_T}, as well as the corresponding scores q_{i,n_T} and the discrimination functions f_i, $i : 1,\ldots,m$, are re-indexed such that $q_{1,n_T} \leq \cdots \leq q_{m,n_T}$. A final supervisor can then be obtained by aggregating f_0 with $f_1, \ldots f_m$ as follows:

$$f_Q = \beta_1 \sum_{i=1}^{m-1} \frac{\beta_i}{\sum_{i'=1}^{m-1} \beta_{i'}} f_i(\mathbf{x}_{n_T}^i) + (1 - \beta_1) f_0(x_{n_T}) \quad (25)$$

where

$$\beta_i = \left(\frac{q_{m,n_T} - q_{i,n_T}}{p_{\max}}\right)^{\alpha_2} \quad (26)$$

The results of this supervisor is shown in Fig. 2 (2) where fingerprint and signature traits are fused using human expert opinions. Again, one can conclude that skill and sample adaptation do help to improve the recognition performance.

Related Entries

▶ Quality Measures

References

1. Ross, A., Jain, A.K.: Information fusion in biometrics. Pattern Recogn. Lett. **24**(13), 2115–2125, (2003)
2. Lindley, D.V.: Making decisions. Wiley, London (1990)
3. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
4. Bigun, E., Bigun, J., Duc, B., Fischer, S.: Expert conciliation for multi modal person authentication systems by bayesian statistics. In: Bigun, J. Chollet, G. Borgefors G. (eds.) Audio and Video based Person Authentication – AVBPA97, pp. 291–300. Springer, Heidelberg (1997)
5. Kittler, J., Li, Y., Matas, J., Sanchez, M.R.: Combining evidence in multimodal personal identity recognition systems. In: Bigun, J. Chollet, G. Borgefors G. (eds.) Audio and Video based Person Authentication - AVBPA97, pp. 327–334. Springer, Heidelberg (1997)
6. Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E. Fusion of face and speech data for person identity verification. IEEE Trans. Neural Networks **10**(5), 1065 (1999)
7. Gutschoven, B., Verlinde, P.: Multi-modal identity verification using support vector machines (SVM). In: Proceedings of the Third International Conference on Information Fusion. IEEE Press (2000). http://citeseer.ist.psu.edu/303146.html;ftp://ftp.elec.rma.ac.be/user/verlinde/pub/fusion_2000.ps.gz
8. Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J., Bigun, J.: Discriminative multimodal biometric authentication based on quality measures. Pattern Recogn. **38**, 777–779 (2005)
9. Lim, E., Toh, K.A., Saganthan, P.N., Jiang, X.D., Yan, W.Y.: Fingerprint image quality analysis. In: International Conference on Image Processing, pp. II: 1241–1244 (2004). URL http: //dx.doi.org/10.1109/ICIP.2004.1419530
10. Tabassi, E., Wilson, C.L.: A novel approach to fingerprint image quality. In: ICIP (2), pp. 37–40 (2005). URL http://dx.doi.org/10.1109/ICIP.2005.1529985
11. Fronthaler, H., Kollreider, K., Bigun, J., Fierrez, J., Alonso-Fernandez, F., Ortega-Garcia, J.: Automatic image quality estimation and its application to multi-algorithm verification. IEEE Transactions on Information Forensics and Security p. accepted (2008)
12. Fierrez-Aguillar, J.: Adapted fusion schemes for multimodal biometric authentication. Phd thesis, Universidad Politecnica de Madrid, ETSIT (2006)

Author Query Form

Encyclopedia of Biometrics
Alpha: M

Query Refs.	Details Required	Author's response
AU1	Kindly confirm whether the sentence,"The rule is not obtained by training on expert opinions, though they might very well be decided by the human designer of the supervisor with her knowledge of expert skills." is correct.Is it "the designer of the supervisor"?	
AU2	Please confirm if you are the copyright holder for the figure. If it is from another source, please confirm whether the necessary permissions have been obtained to reproduce the figure. Also, please provide us the permission details to be carried while citing the figure source.	