R. C. Dixon
P. E. Boudreau

Mathematical Model for Pattern Verification

Abstract: Pattern verification is mathematically defined, an appropriate decision function is derived, and a measure for system evaluation is given. Two basic postulates are set forth to fully define a verification system: each known class is expected, with nonzero probability, to be verified under the correct class label; and the pattern vector extracted during verification should be descriptive of the given class, independent of which class label was entered into the system. Through appropriate use of a priori probabilities, three types of information can be incorporated into the theory: the expected number of times a given class will require verification, the expected use of each class label by a given class, and the likelihood that a particular class is susceptible to "impostor" patterns.

Introduction

This paper describes a mathematical model for pattern verification based upon various concepts from probability theory. Verification will be mathematically defined and contrasted to identification. The correct use of the appropriate a priori probabilities and loss functions is stressed as a necessary contribution to the decision process. However, the theory does not address the practical difficulties of satisfactorily estimating the probabilities of the various patterns. Although this theory was developed with verification of speakers as a prime interest, the theory is applicable to any verification system, such as fingerprint verification and pathological verification.

What happens in a verification system? A pattern, not necessarily a member of a known class, is presented to the system. A label is also presented to the system asserting, truthfully or falsely, that the pattern belongs to a certain one of the classes known to the system. Next, a predetermined set of parameters, possibly depending upon which label was entered, is extracted from the pattern. Finally, the pattern is accepted or rejected as a member of the indicated class, depending upon the values of the various parameters. For example, in a speaker verification system, a given person (the class) utters a predetermined phrase (the pattern) which is analyzed by the system. Upon indicating to the system who he claims to be (the label), the system either accepts or rejects him. Three experimental investigations into speaker verification have been reported in Refs. 1–3.

The authors are at the IBM Systems Development Division Laboratory in Raleigh, North Carolina.

If we consider that all unusable patterns (e.g., an inkblot would be considered unusable in a fingerprint verification system) are never presented to the system, the four principal differences that distinguish a verification procedure from an identification procedure are:

(1) An alien class, for which a priori information is not available, is considered by the system.
(2) Additional information, the class label, is available for the decision.
(3) The class label that is entered can determine which parameters are to be extracted from the pattern.
(4) The decision involves only two states, acceptance or rejection of the pattern.

In identification, all possible classes are presumed known, and the decision amounts to the best match of the pattern to a particular class.

The next section develops a decision theory based upon classical Bayesian statistics (see Refs. 4–8 for fundamentals of the theory), including a description of the a priori probabilities required. The last section defines various measures for evaluating various parts of the verification process as well as an over-all measure of the system's performance.

Decision function

We will be concerned with three random variables: C, the possible classes; L, the possible class labels; and V, the possible sample pattern vectors. We assume the system involves M known classes, $C_1, C_2, \cdots, C_M$ and a single alien class C_0 consisting of all unknown classes. Corresponding to each known class C_j is a label $L = L_j, (j = 1, \cdots, M)$. The sample pattern vector can assume N states, $V = V_k, (k = 1, 2, \cdots, N)$. Each pattern vector is described by an n-component vector extractable from the input pattern to be verified. The classes, labels, and pattern vector states are assumed to be mutually exclusive and exhaustive. In our notation, we denote the probability that $C = C_i$ by $p(C_i)$ and the conditional probability that $C = C_i$ given $L = L_j$ and $V = V_k$ by $p(C_i|L_jV_k)$, etc. Further, we define $\bar{C}_j = \{C_0, C_1, \cdots, C_{j-1}, C_{j+1}, \cdots, C_M\}$ and call it the impostor class.

On the basis of classical decision theory, we will define a decision function for a verification system. We are dealing with a two-state decision process involving two sets of loss functions:

$\lambda(\bar{C}_j|C_j)$ = the loss from classifying C_j as a member of $\bar{C}_j$ (reject a valid pattern),

$\lambda(C_j|\bar{C}_j)$ = the loss from classifying a member of $\bar{C}_j$ as C_j (accept an invalid pattern),

for $j = 1, 2, \cdots, M$.

In order for verification to be meaningful, the above loss functions are restricted to positive values.

Recognizing that the decision depends upon the class label L_j entered and the state of the extracted pattern vector V_k, we define the decision function

$$G_j(V_k) = \lambda(\bar{C}_j|C_j)p(C_j|L_jV_k) - \lambda(C_j|\bar{C}_j)p(\bar{C}_j|L_jV_k).$$

The accept/reject rule is to accept the pattern as valid if $G_j(V_k) > 0$ and reject the pattern if $G_j(V_k) \leq 0$. Using this rule together with the fact that the sum of the two loss functions associated with a particular class is positive, and defining the ratio

$$\theta_j = \frac{\lambda(C_j \mid \bar{C}_j)}{\lambda(C_j \mid \bar{C}_j) + \lambda(\bar{C}_j \mid C_j)},$$

we can rewrite the decision function as

$$G_j(V_k) = p(C_j|L_jV_k) - \theta_j.$$

Using Bayes' theorem and the laws of total probabilities, we write

$$p(C_i \mid L_j V_k) = \frac{p(V_k \mid C_i L_j)p(L_j \mid C_i)p(C_i)}{p(V_k L_j)},$$

to show that the problem of estimating $p(C_i|L_jV_k)$ can be reduced to estimating $p(C_i)$ for $i = 0, 1, \cdots, M$; $p(L_j|C_i)$ for $j = 1, 2, \cdots, M$; and $p(V_k|L_jC_i)$ for $k = 1, 2, \cdots, N$. In addition, a choice concerning the θ_j's must be made.

Before we can show how one might determine the above probabilities, we require the following two basic postulates to fully define a verification system:

Postulate I: $p(V_k|L_jC_i) = p(V_k|C_i), (i = 0, 1, 2, \cdots, M)$.

In the presence of the knowledge of the class C_i, the pattern vector produced, V_k, should be independent of the class label which was entered.

This is not saying that V and L are independent, but that V is independent of L when C is given. If this postulate does not hold for a particular system, the particular parameters used in V_k are less than ideal, and the verification system would not be as successful.* In other words, the pattern vector V should be descriptive of the class C, independent of which class label has been entered into the system.

Postulate II: $p(C_jL_j) > 0, \qquad (j = 1, 2, \cdots, M)$.

The joint event that the class is C_j and the label L_j is entered has nonzero probability.

This says that each of the M known classes are expected, with a nonzero probability, to be verified under their correct label. Otherwise, $p(C_jL_j) = 0$ would imply that the class C_j should be considered as an alien (a member of C_0) and only $(M - 1)$ known classes should be included in the system. Further reasons for these two postulates will become apparent in the following.

We will assume that a set of labeled pattern vectors is available from which one could estimate the conditional probability distributions $p(V_k|C_i), (i = 0, 1, 2, \cdots, M)$, for each class. Applying Postulate I, we can then give an estimate of the probability $p(V_k|L_jC_i)$. In order to estimate the distribution for the alien class, $p(V_k|C_0)$, it seems conceivable that a representative set of pattern vectors, not from the members of the M known classes, could be used. This would require a large number of patterns extracted from a wide spectrum of possible alien patterns. We are left with the need to estimate the a priori probabilities $p(C_i)$ and $p(L_j|C_i)$. These could be readily estimated if it were not for the inclusion of the unknown class, represented by the alien class C_0. For example, if $p(C_0) = 0$, one should be able to give a fair estimate of the $p(C_i)$ for each known class $(1 \leq i \leq M)$. However, it would seem quite difficult to give an estimate for the verification system usage by members of the alien class. Thus, the a priori probabilities, $p(C_i)$, for all classes cannot be estimated, but only the relative frequency, f_i of each known class C_i:

$$f_i = \frac{p(C_i)}{1 - p(C_0)} = \frac{p(C_i)}{p(\bar{C}_0)}, \qquad (i = 1, 2, \cdots, M).$$

* If a pattern from an impostor class can be made to appear like a typical pattern from the true class, the verification system should accept it as valid. Thus, characteristics should be extracted from the pattern so that they are not easily duplicated by an impostor. If this postulate is not assumed, knowledge of the joint probability distribution $p(V, L, C)$ would be required. This appears very impractical, if not impossible, for most verification systems.

Having a priori knowledge of each of the M known classes, one should be able to estimate the conditional probabilities, $p(L_j|C_i)$. That is, how often is a particular known class, C_i, going to require verification under a particular label, L_j? For most systems, we propose that the $p(L_i|C_i)$ will be quite large, with the $p(L_j|C_i)$, $j \neq i$, being small. For lack of specific information, one could choose these latter conditional probabilities to be equal, once the $p(L_i|C_i)$ have been specified.

To complete the specification of our problem, we need to determine the probabilities $p(C_0)$ and $p(L_j|C_0)$.

The likelihood that a particular class is susceptible to impostors is information available to a verification system that an identification system doesn't have. We can use this to show that an indirect method is available for estimating the a priori probabilities.

We need to specify the susceptibilities $p(C_j|L_j)$, ($j = 1, 2, \cdots, M$). There exist upper limits for these probabilities once values have been assigned to the relative frequencies, f_i, and to the a priori probabilities $p(L_j|C_i)$, ($i, j = 1, 2, \cdots, M$). We have

$$0 < p(C_j \mid L_j) = \frac{p(L_jC_j)}{\sum_{i=0}^{M} p(L_jC_i)} \leq \frac{p(L_jC_j)}{\sum_{i=1}^{M} p(L_jC_i)}$$

with equality on the right whenever $p(L_jC_0) = 0$.

In terms of the previously specified frequencies and probabilities, we have the relationship

$$0 < p(C_j \mid L_j) \leq \frac{p(L_jC_j)f_j}{\sum_{i=1}^{M} p(L_j \mid C_i)f_i}.$$

Note that if $p(L_j|C_i) = \delta_{ji}$, the Kronecker delta, the above upper limit becomes one. Otherwise, the limit is less than one, reflecting the fact that known classes, other than C_j, will attempt to be verified under the label L_j. When $p(C_j|L_j)$ is assigned a value less than the upper limit, it indicates that some members of the alien class C_0 are expected to attempt verification under class label L_j.

Once these M probabilities are specified, subject to the above limits, one can easily determine the remaining unknown quantity, $p(C_0)$, the probability of an alien class being presented to the system, from the relationship

$$\frac{1}{p(\bar{C}_0)} = \sum_{j=1}^{M} \frac{p(L_j \mid C_j)f_j}{p(C_j \mid L_j)}.$$

Since $p(C_jL_j) > 0$, by Postulate II, the above function will always be defined. Note that if $p(L_j|C_j) = p(C_j|L_j)$, which implies $p(L_j) = p(C_j)$, the above equation gives $p(\bar{C}_0) = 1$. In other words, no alien patterns are ever presented to the system.

We now have specified the information that is required to carry out verification of an unknown pattern. This consists of the quantities θ_j, $p(V_k|C_i)$, f_i, $p(L_j|C_i)$, and $p(C_j|L_j)$. Given these probabilities, one can easily determine all other probabilities related to the three random variables, C, L, or V. Specifically, we will need

$$p(C_i) = \frac{f_i}{\sum_{j=1}^{M} \frac{p(L_j \mid C_j)f_j}{p(C_j \mid L_j)}}, \qquad (i = 1, 2, \cdots, M),$$

and

$$p(L_j) = \frac{p(L_j \mid C_j)p(C_j)}{p(C_j \mid L_j)}, \qquad (j = 1, 2, \cdots, M).$$

Using these quantities, we can now evaluate the various decision functions for each class through the following equation:

$$G_j(V_k) = \frac{p(V_k \mid C_j)p(L_j \mid C_j)p(C_j)}{p(V_kL_j)} - \theta_j.$$

In the next section, we will use these probabilities to give an expression for evaluating the verification system.

One final but very important remark needs to be made. For verification under a particular class label, say L_j, only one a priori susceptibility is required; namely, $p(C_j|L_j)$. The other $(M - 1)$ susceptibilities do not enter into the calculation of the decision function. This is not obvious from the above equations, since all M susceptibilities are required to calculate the $p(C_i)$'s for the known classes. To prove this, one can expand $p(V_kL_j)$ as

$$p(V_kL_j) = p(V_k \mid C_0)p(L_j) + \sum_{i=1}^{M} [p(V_k \mid C_i) - p(V_k \mid C_0)]\, p(L_j\, C_i).$$

Using the expression for $p(L_j)$, we are able to express the decision function in terms of the relative frequencies, f_i, as

$$G_j(V_k) = \Bigg\{[p(V_k \mid C_j)p(L_j \mid C_j)f_j] \div \Bigg[P(V_k \mid C_0)\frac{p(L_j \mid C_j)}{p(C_j \mid L_j)} f_j + \sum_{i=1}^{M} [p(V_k \mid C_i) - p(V_k \mid C_0)] \times p(L_j \mid C_i)f_i\Bigg]\Bigg\} - \theta_j.$$

Note that only the true class susceptibility, $p(C_j|L_j)$, occurs in the above equation, and the probability of the alien class, $p(C_0)$, does not explicitly appear. This means that the $p(C_j|L_j)$ can be adjusted at the time of the verification on the basis of other available information, such as an observer's estimation of how certain he is about the true identity of the class before the pattern vector is extracted.

This enables an observer, or the system for that matter, to affect the outcome of the decision process.

The strong effect of this a priori susceptibility on the resulting decision can be easily seen if we assume that $p(L_j|C_i) = \delta_{ij}$, which implies that all known classes will only require verification under their own class label. This reduces the decision function to

$$G_j(V_k) = \frac{p(V_k \mid C_j)p(C_j \mid L_j)}{p(V_k \mid C_j)p(C_j \mid L_j) + p(V_k \mid C_0)p(\bar{C}_j \mid L_j)} - \theta_j .$$

For this special case, we have shown that the value of the $p(C_j|L_j)$ can range between zero and one. Thus, an observer can cause the verification to always reject or always accept the pattern regardless of the pattern vector extracted, or can influence the decision one way or the other to any degree he desires.

System evaluation

As a measure for evaluating the resulting verification system, assuming we know all the probabilities given in the last section, we will use the average expected loss per verification. To derive this measure, we need to know the probability that a loss will occur. As explained earlier, there are only two types of losses in a verification system: accepting a member of the imposter class, or rejecting a member of the true class.

Denoting the event "acceptance" by the letter A, we define the probability of acceptance as

$$p(A) = \sum_{\substack{j=1 \\ G_j(V_k)>0}}^{M} \sum_{k=1}^{N} p(L_j V_k),$$

where the summation includes only those terms which satisfy the indicated inequality. That is, the random variables L and V have to give a positive value to the decision function to cause acceptance of the pattern vector.

Since we are concerned with only two types of errors in a verification system, the following two probabilities are of prime importance:

(1) The probability of acceptance given that a known class, C_j, is to be verified under its own label,

$$p(A \mid C_j L_j) = \sum_{\substack{k=1 \\ G_j(V_k)>0}}^{N} p(V_k \mid C_j), \qquad (j = 1, 2, \cdots, M).$$

(2) The probability that it really was the true class, C_j, when a pattern was accepted under the label L_j,

$$p(C_j \mid AL_j) = \frac{p(A \mid C_j L_j)p(L_j \mid C_j)p(C_j)}{\sum_{i=0}^{M} p(A \mid C_i L_j)p(L_j \mid C_i)p(C_i)} .$$

We can now define the average loss per verification, Λ, as

$$\Lambda = \sum_{j=1}^{M} [\lambda(\bar{C}_j \mid C_j)p(C_j L_j R) + \lambda(C_j \mid \bar{C}_j)p(\bar{C}_j L_j A)],$$

where R represents the event of being rejected. The two joint probabilities occurring in the above equation can be readily calculated from previous expressions.

It is instructive to examine the range of possible values for Λ. For a perfect verification system,

$$p(\bar{C}_j L_j A) = p(C_j L_j R) = 0,$$

and the average loss per verification is zero. The maximum average loss results when the wrong decision is made for each verification. Mathematically, this requires

$$p(\bar{C}_j L_j A) = p(\bar{C}_j L_j)$$

and

$$p(C_j L_j R) = p(C_j L_j),$$

which gives

$$\Lambda_{\text{Max}} = \sum_{j=1}^{M} [\lambda(C_j \mid \bar{C}_j)p(\bar{C}_j L_j) + \lambda(\bar{C}_j \mid C_j)p(C_j L_j)] .$$

This enables us to calculate the maximum average loss based upon a priori information only.

Summary

This mathematical model should prove useful in the final design and evaluation of verification systems, once a reliable estimate can be obtained of the probability distributions of the various class pattern vectors. We have shown that:

(1) Alien class members can be accounted for by proper use of the appropriate a priori probabilities.
(2) An observer can affect the outcome of the decision by presetting the relevant a priori susceptibility.
(3) The probability that a particular class is accepted under a particular class label can be used to obtain a measure for evaluating the verification system.

In order that all probabilities used in the decision function and other expressions are defined, we have required two basic postulates:

(1) The pattern vector probability distribution from a particular class should be independent of the class label entered into the system.
(2) The joint event that a known class is to be verified under its correct class label should have nonzero probability.

Most important, we have attempted to define explicitly the difference between pattern verification and pattern identification. This difference is accounted for primarily by the inclusion of an additional variable, the class label under which the verification is to be performed.

References

1. K. P. Li, J. E. Dammann, and W. D. Chapman, "Experimental Studies in Speaker Verification Using an Adaptive System," *JASA* **40,** 966-978 (November, 1966).
2. J. E. Luck, "Description of a Completely Automatic Speaker-Verification System," presented at 77th meeting of Acoustical Society of America, Philadelphia, April, 1969.
3. G. Beakley and F. B. Tuteur, "Application of Distribution-Free Tolerance Regions to Automatic Speaker Verification," presented at 77th meeting of Acoustical Society of America, Philadelphia, April, 1969.
4. A. I. Khinchin, *Mathematical Foundations of Information Theory,* Dover Publications, New York, 1967.
5. R. M. Fano, *Transmission of Information,* J. Wiley and Sons, New York, 1961.
6. N. J. Nilsson, *Learning Machines,* McGraw-Hill, New York, 1965.
7. G. S. Sebestyen, *Decision-Making Processes in Pattern Recognition,* Macmillan, New York, 1962.
8. E. Parzen, *Modern Probability Theory and Its Applications,* John Wiley and Sons, New York, 1960.

Received January 14, 1969
Revised May 1, 1969