55. Veloso MM, Carbonell JG (1993) Derivational Analogy in PRODIGY: Automating Case Acquisition, Storage and Utilization. In: Kolodner JL (ed) Case-Based Learning. Kluwer Academic Publishers, Norwell, MA, pp 55–84
56. Utgoff PE (1989) Incremental induction of decision trees. Machine Learning, 4, 2:161–186
57. West GM, McDonald JR (2003) An SQL-Based Approach to Similarity Assessment within a Relational Database. In: Ashley K, Bridge DG (eds) Proceedings of ICCBR 03. Springer-Verlag, Lecture Notes in Artificial Intelligence 2689, Berlin, Heidelberg, New York, pp 610–621
58. Wilson DC, Leake DB (2001) Mainting Case-based Reasoners: Dimensions and Directions. Computational Intelligence Journal, Vo. 17, No. 2:196–213
59. Wiratunga N, Koychev I, Massie S (2004) Feature Selection and Generalisation for Retrieval of Textual Cases. In: Funk P, Gonzàlez Calero P (eds) Proceedings of ECCBR 04. Springer-Verlag, Lecture Notes in Artificial Intelligence 3155, Berlin, Heidelberg, New York, pp 806–820
60. Yang Q, Cheng H (2003) Case Mining from Large Databases. In: Ashley K, Bridge DG (eds) Proceedings of ICCBR 03. Springer-Verlag, Lecture Notes in Artificial Intelligence 2689, Berlin, Heidelberg, New York, pp 691–702

7

Learning a Statistical Model for Performance Prediction in Case-Based Reasoning

B. Bhanu and R. Wang

Center for Research in Intelligent Systems, University of California at Riverside, California, 92521, USA

Summary. This chapter is concentrated with the performance characterization of a case-based reasoning (CBR) system. Based on the match score and nonmatch score computed from the cases in the case library, we develop a statistical model for prediction. We estimate the size of a subset of cases, called gallery size, that can generate the optimal error estimate and its confidence on a large population (relative to the size of the gallery). The statistical model is based on a generalized two-dimensional prediction model that combines a hypergeometric probability distribution model with a binomial model explicitly and considers the data distortion problem in large populations. Learning is incorporated in the prediction process in order to find the optimal small gallery size and to improve the prediction performance. During the prediction, the expectation–maximization (EM) algorithm is used to learn the match score and the nonmatch score distributions that are represented as mixture of Gaussians. By learning, the optimal size of small gallery is determined and at the same time the upper bound and the lower bound for the prediction on large populations are obtained. Results are shown using a real-world database with the increasing size of the case library.

7.1 Introduction

Case-based approaches are characterized by how the learner represents what it has learned so far, as well as the analogical methods which are used to transfer the learned experience [1]. In CBR, "past" experiences are stored in memory as cases and are used to solve a new problem case. Given a problem to be solved, the case-based method retrieves from the memory the solution to a similar problem encountered in the past, adapts the previous solution to the current problem, and stores the new problem-solution packet as another case in the memory. The major concerns with CBR are the selection of the indexing scheme to organize cases in the memory, the method for choosing the most relevant cases at reasoning time, the adaptation heuristics to modify pervious cases to fit the current problem, and maintenance of the case library.

B. Bhanu and R. Wang: *Learning a Statistical Model for Performance Prediction in Case-Based Reasoning*, Studies in Computational Intelligence (SCI) **73**, 195–212 (2008)
www.springerlink.com

An important problem for a CBR system is the prediction of its performance as the case library grows. In this chapter, we present a statistical model for performance characterization.

Recognition/classification systems can classify images, signals, or other types of measurements into a number of classes. In a simplistic way, we can view a CBR system as a model-based recognition/classification system [2, 3], which stores a set of models in its case library and classifies the incoming cases by performing match with the database models. In this process, the case library will be continuously growing as more cases/models become part of the case base. It is like an incremental learning system for object recognition, which not only involves traditional recognition/classification of complete/occluded objects but also new model acquisition and refinement of existing model [3, 4]. This chapter provides a method for the life-time problem [5] of a CBR system. We define the following terms in the concept of CBR: gallery/training set, probe/testing set, populations, algorithm/system, data, recognition/classification.

We define the following terms in the context of CBR: The gallery set and the probe set are the training set and the testing set, respectively. The algorithm/system is the set of recognition/classification programs. The population is the data that the system under consideration may encounter in its lifetime.

Since the recognition performance of an algorithm/system is usually based on limited data, it is difficult to estimate this performance for additional data: the limited test data may, after all, not accurately represent a larger population. Before we can evaluate and predict the performance of a recognition algorithm/system on large populations, we need to answer some fundamental questions. When we use a small gallery to estimate the algorithm/system performance on large populations how can we find the optimal size of the small gallery and how accurate is the estimation? Since the prediction is based on the selected recognition algorithm/system, we can give the confidence interval for the performance estimation on a large population [6]. The confidence interval can describe the uncertainty associated with the estimation. This gives an interval within which the true algorithm/system performance for the large population is expected to fall, along with the probability that it is expected to fall there [7].

Given limited data we can use the Bayesian parameter estimation or nonparametric estimation methods to estimate the data distribution. The expectation–maximization (EM) algorithm, one of the parameter estimation methods, assumes that the underlying distribution is known. It is an iterative method to estimate the mixture parameters by maximum likelihood techniques. The parzen window and the K-nearest-neighbor are nonparametric estimation methods that are used to estimate the data distribution in case the underlying distribution is unknown. These two methods converge the distribution to the unknown distribution. The parzen window method estimates the density, while the K-nearest-neighbor method determines the K closest neighbors [8].

In this chapter, we use a generalized prediction model that combines a hypergeometric probability distribution model with a binomial model. This prediction model takes into account distortion that may occur in large populations. It also provides performance measurements as a function of rank, large population size, number of distorted images, and similarity score (match and nonmatch score) distributions. While we use the EM algorithm to estimate the match score and the nonmatch score distributions, we introduce learning to feed back similarity scores (match scores and nonmatch scores) to increase the small gallery size. In this way, we can find the optimal size of the small gallery to predict the large population performance. Meanwhile, we provide the upper and the lower bounds for the prediction performance of a large population. In this chapter, we use two different statistical methods – Chernoff's inequality and Chebychev's inequality – to obtain the relationship between the small gallery size and the confidence interval given a margin of error. The small gallery size for prediction that we get from the learning process is smaller than the size determined by statistical methods.

The paper is organized as follows. Related work and contributions are presented in Sect. 7.2. The details of the technical approach are given in Sect. 7.3. It includes the integrated model, the procedure of learning for similarity score distributions in the prediction, and the statistical methods to find the optimal sample size. Experimental results are shown in Sect. 7.4. The integrated model with learning is tested on the NIST-4 fingerprint database. Conclusions are given in Sect. 7.5.

7.2 Related Work and Contributions

7.2.1 Related Work

Until now the prediction models are mostly based on the feature space or similarity scores [9]. The statistical approaches are used by many researchers to estimate the recognition system performance. Wayman [10] and Daugman [11] develop a binomial model that uses the nonmatch score distribution. This model underestimates recognition performance for large populations. Phillips et al. [12] develop a moment model, which uses both the match score and the nonmatch score distributions. Since all the similarity scores are sampled independently, the probability of error is increased and the prediction results underestimate the identification performance. Wang and Bhanu [9] present a binomial model to predict the large fingerprint database recognition performance based on a small gallery. They present their early work on a generalized two-dimensional model, which integrates a hypergeometric probability distribution explicitly with a binomial distribution [13]. This work considers the distortion caused by sensor noise, feature uncertainty, feature occlusion, and feature clutter. However, there is no learning to determine the optimal small gallery size and the bounds on performance. Johnson et al. [14] improve the

moment model by using a multiple nonmatch score set. They average match scores on the whole gallery. For each match score they count the number of times that a nonmatch score is larger than the match score, leading to an error. In this chapter, they assume that the distribution of the match score is uniform. Grother and Phillips [15] introduce the joint density function of the match score and the nonmatch score to estimate both the open-set and the closed-set identification performance. The closed-set identification is the identification for which all potential users are enrolled in the system. The open-set identification is the identification for which some potential users are not enrolled in the system. Since the joint density is generally impractical to estimate, they assume that the match score and nonmatch score are independent and their distributions are the same for large populations. They use the Monte Carlo sampling method to linearly interpolate the match score and the nonmatch score look-up tables.

Providing the upper and lower bounds for the prediction performance is another important topic in the recognition performance prediction. Lindenbaum [16] proposes a probabilistic method to derive bounds on the number of features required to achieve recognition with a certain degree of confidence. This method considered object similarity, bounded uncertainty, and occlusion. A similar approach presented in [17] can be used to analyze object recognition with uncertainty, similarity, and clutter. Boshra and Bhanu [18, 19] present a method to predict upper and lower bounds on the performance prediction. They predict performance by considering feature uncertainty, occlusion, clutter, and similarity simultaneously. In their method performance is predicted in two steps: compute the similarity between each pair of models; use the similarity information along with the statistical model to determine upper and lower bounds for the object recognition performance. Guyon et al. [1] propose guaranteed estimators to determine the test size for the independent identical distribution recognition error and the correlated recognition error, along with the assumption of the underlying probability distribution.

7.2.2 Contributions

In this chapter we address the problems associated with the prediction of performance on large populations and optimal small gallery size. The specific contributions are the following:

1. We use a generalized prediction model that combines a hypergeometric probability distribution model explicitly with a binomial model which takes into account distortions that may occur in large populations. Our distortion model includes feature uncertainty, feature occlusion, and feature clutter. In the prediction model, we use the EM algorithm to estimate similarity score (match score and nonmatch score) distributions and find the number of components of the distributions automatically.

2. We find the optimal size of a small gallery by an iterative learning process [20]. We use the Chernoff inequality and the Chebychev inequality to determine the small gallery size in theory which is related to the margin of error and the confidence interval. We find the upper bound and a good lower bound for predicting recognition performance on a large population.
3. The results are shown on a large data set (NIST-4) of fingerprint images [21]. We show the prediction results with the increasing size of the database.

7.3 Technical Approach

We are given two sets of data: a gallery set and a probe set. The gallery is a set of models saved in the database. The probe is a set of queries for the database. A large population is the unknown data set whose recognition performance needs to be estimated. Based on the given gallery and probe set we would like to estimate the recognition performance on large populations.

7.3.1 Methodology for Determining the Small Gallery Size

Figure 7.1 provides the detailed diagram for the implementation of our proposed approach to get the optimal small gallery size. For a given recognition system whose database size or the number of classes is N, we randomly pick n images from the database N to be our small gallery. By an authentication process, we can get a set of match scores and nonmatch scores for this small gallery. Then, we use the EM algorithm [8] to estimate distributions of the match score and the nonmatch score. Assume that the match score and nonmatch score distributions are Gaussian mixtures. Let $ms(x)$ represents the match score distribution and $ns(x)$ represents the nonmatch score distribution. We have

$$ms(x) = \sum_{i=1}^{m} \alpha_i ms_i(x) \tag{7.1}$$

and

$$ns(x) = \sum_{j=1}^{n} \beta_j ns_j(x) \tag{7.2}$$

where m and n are the number of components, α_i and β_j are the component proportions, $\sum_{i=1}^{m} \alpha_i = 1$, and $\sum_{j=1}^{n} \beta_j = 1$. We have $ms_i \sim N(\mu_{si}, \sigma_{si}^2)$, $ns_j \sim N(\mu_{nj}, \sigma_{nj}^2)$, where μ_s, μ_n, σ_s^2, and σ_n^2 are the mean and variance for the match score distribution and the nonmatch score distribution, respectively.

Based on these distributions, we use our prediction model, which combines a hypergeometric probability distribution model with a binomial model

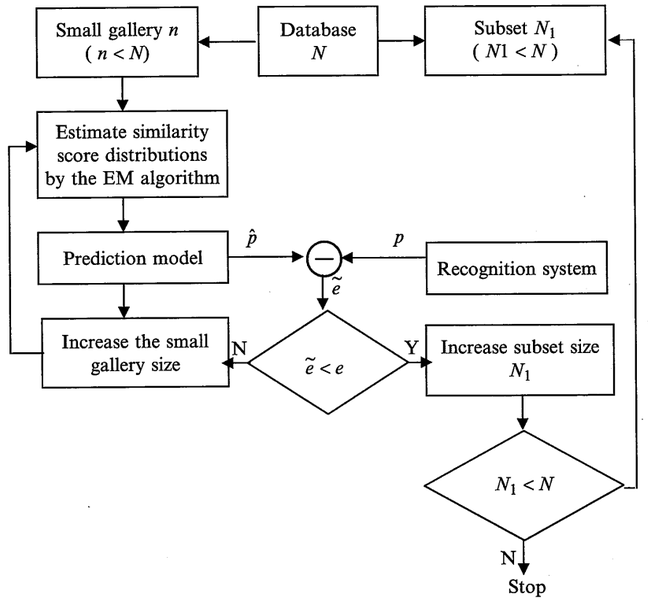


Fig. 7.1. Conceptual prediction model

to estimate the recognition system performance for a large population N_1, a subset of database N ($N_1 < N$). We assume the prediction performance on N_1 is $\hat{p}$. From the recognition system we can obtain the match score and the nonmatch score for N_1. Then, compute the actual recognition performance p for N_1. $\tilde{e}$ is the error between the predicted performance and the actual performance, $\tilde{e} = |\hat{p} - p|$. The margin of error e is the maximum specified error acceptable by the recognition system. If $\tilde{e}$ is larger than the margin of error e then we increase the small gallery size n and feed back match scores and nonmatch scores to the EM algorithm to estimate the similarity score distributions again. Otherwise, we increase N_1, the size of the large population, and repeat this process until the N_1 has increased to N. We use the Chernoff and Chebychev inequalities to find the relationship between the small gallery size and the prediction confidence interval given a margin of error. The small gallery size which we get from the inequalities is used to validate the learned optimal small gallery size. We will explain each part of the diagram in detail in this section.

7.3.2 Prediction Model

Usually a recognition system consists of three stages: image acquisition, feature extraction, and matching. Distortion often occurs in these stages and is caused by sensor noise, feature uncertainty, feature occlusion, and feature clutter. The effects of sensor and image noise are reflected in the feature uncertainty. Our two-dimensional prediction model considers the distortion problem which conforms to reality. Assume we have two kinds of different quality biometric images, group #1 and group #2. Group #1 is a set of biometric images without distortion. Group #2 is a set of biometric images with distortion. Let the size of these two groups be n_1 pairs and n_2 pairs, respectively. We randomly pick n pairs of images from group #1 and group #2 to be our small gallery. Then, the number of pairs of distorted images y which are chosen from group #2 follow the hypergeometric distribution

$$f(y) = \frac{C_{n-y}^{n_1} C_y^{n_2}}{C_n^{n_1+n_2}} \tag{7.3}$$

where $n_1 + n_2$ is the total number of images in these two groups and $n - y$ is the number of images chosen from group #1.

These n pairs of images are our small gallery. We split them into the gallery and the probe set. For each image in the probe set, we compute the match score and the nonmatch score with images in the gallery. Then, we have one match score and $n - 1$ nonmatch scores for this image. We assume that the match score and the nonmatch score are independent. With all these similarity scores we can use the EM algorithm to estimate the match score and the nonmatch score distributions.

From the above discussion, we know that the match score and nonmatch score distributions depend not only on the similarity scores but also on the number of images with distortion. Let $ms(x|y)$ and $ns(x|y)$ represent the distributions of match scores and nonmatch scores given the number of distorted images. If the similarity score is higher then the object are more similar. The error occurs when a given match score is smaller than the nonmatch score corresponding to the same image. For a given number of distorted images, the probability that the nonmatch score is greater than or equal to the match score x is $NS(x)$, where

$$NS(x) = \int_x^\infty \sum_{y=0}^{n} ns(t|y) f(y) dt \tag{7.4}$$

Thus, the probability that the nonmatch score is smaller than the match score is $1 - NS(x)$.

If the size of the large population is N, then for the jth image we can have one match score and $N - 1$ nonmatch scores. We rank the match score and the nonmatch score in the descending order. For a given number of images

with distortion, the probability that the match score x is at rank r is given by the binomial probability distribution

$$C_{r-1}^{N-1}(1-NS(x))^{N-r}(NS(x))^{r-1} \tag{7.5}$$

Integrating over all the match scores, for a given number of images with distortion, the probability that the match score is at rank r can be written as

$$\int_{-\infty}^{\infty} C_{r-1}^{N-1}(1-NS(x))^{N-r}(NS(x))^{r-1} ms(x|y)dx \tag{7.6}$$

By summing over the images chosen from group #2, the probability that the match score is at rank r can be written as

$$\int_{-\infty}^{\infty} C_{r-1}^{N-1}(1-NS(x))^{N-r}(NS(x))^{r-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \tag{7.7}$$

In theory, a match score can be any value within $(-\infty, \infty)$. The probability that the match score is within rank r is

$$P(N,r) = \sum_{i=1}^{r} \int_{-\infty}^{\infty} C_{i-1}^{N-1}(1-NS(x))^{N-i}(NS(x))^{i-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \tag{7.8}$$

Given that the correct match takes place above a threshold t, the probability that the match score is within rank r becomes

$$P(N,r,t) = \sum_{i=1}^{r} \int_{t}^{\infty} C_{i-1}^{N-1}(1-NS(x))^{N-i}(NS(x))^{i-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \tag{7.9}$$

When rank $r = 1$ the prediction model with threshold t becomes

$$P(N,1,t) = \int_{t}^{\infty} (1-NS(x))^{N-1} \sum_{y=0}^{n} ms(x|y)f(y)dx \tag{7.10}$$

In this model, we make two assumptions: match scores and nonmatch scores are independent and large populations have distortion with model of feature uncertainty, occlusion, and clutter. We use a small gallery to estimate distributions of $ms(x|y)$ and $ns(x|y)$.

7.3.3 Estimation of The Small Gallery Size Based on Statistical Inequalities

In the following, we discuss the relationship between the prediction confidence interval and the size of the small gallery which could be used to validate the optimal small gallery size that we obtain by the learning process. We use limited data to estimate a large population recognition performance. Therefore,

the prediction value may or may not be very accurate. This question can be mathematically expressed as

$$P\{|(p-\hat{p})| > e\} \leq (1-\alpha) \tag{7.11}$$

where $\hat{p}$ is the predicted performance for the recognition system which can be obtained from our prediction model, p is the actual performance of the recognition system, e is the margin of error for the system, and α is the confidence interval. Then, inequality (7.9) can be written as

$$P\{p > \hat{p} + e\} \leq (1-\alpha) \tag{7.12}$$

or

$$P\{p < \hat{p} - e\} \leq (1-\alpha) \tag{7.13}$$

Here, we consider inequality (7.12) since inequality (7.13) can be solved by the same procedure as inequality (7.12).

We assume that a recognition system recognizes (authentication) individuals with the probability $P\{X_i = 1\} = p$ and $P\{X_i = 0\} = 1-p$, where $X_i = 1$ means an individual with a given object X_i is recognized correctly, $X_i = 0$ means the opposite, $0 \leq p \leq 1$. According to the Chernoff inequality [22], let $X_1, X_2, \ldots, X_n$ be independent random variables. We define the random variable

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7.14}$$

For any $t \geq 0$ we have

$$P\{X \geq E(X) + \frac{t}{n}\} \leq e^{-\frac{2t^2}{n}} \tag{7.15}$$

where $E(X)$ is the mean of X. Comparing with inequality (7.12), we can get

$$1-\alpha = e^{-\frac{2t^2}{n}} \tag{7.16}$$

So,

$$t = \sqrt{-\frac{n\ln(1-\alpha)}{2}} \tag{7.17}$$

Thus, equation (7.15) becomes

$$P\{X \geq E(X) + \sqrt{-\frac{\ln(1-\alpha)}{2n}}\} \leq 1-\alpha \tag{7.18}$$

From inequality (7.12), we know that

$$e = \sqrt{-\frac{\ln(1-\alpha)}{2n}} \tag{7.19}$$

Thus, we get

$$n = -\frac{\ln(1-\alpha)}{2e^2} \tag{7.20}$$

Equation (7.20) is the relationship between the small gallery size and the confidence interval under the given margin of error for the system with the underlying distribution.

In the above we assume that a recognition system can recognize object with a certain distribution. If we do not know the underlying distribution of the recognition system, then we can use the Chebychev inequality [22] which is distribution independent. Assume $X_1, X_2, \ldots, X_n$ are independent random variables. We define X as

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7.21}$$

For any $\varepsilon \geq 0$, we have

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2} \tag{7.22}$$

where σ^2 is the variance of X. Comparing with (7.12), we have

$$1 - \alpha = \frac{\sigma^2}{n\varepsilon^2} \tag{7.23}$$

From the above equation, we obtain

$$\varepsilon = \frac{\sigma}{\sqrt{n(1-\alpha)}} \tag{7.24}$$

From (7.22), (7.23), and (7.24) we have

$$P\{X \geq E(X) + \frac{\sigma}{\sqrt{2n(1-\alpha)}}\} \leq (1-\alpha) \tag{7.25}$$

Then

$$e = \frac{\sigma}{\sqrt{2n(1-\alpha)}} \tag{7.26}$$

So we have,

$$n = \frac{\sigma^2}{2(1-\alpha)e^2} \tag{7.27}$$

From equation (7.27), we obtain the relationship between the small gallery size and the confidence interval under the given margin of error for the system without the assumption of the underlying distribution. It is known that the Chernoff inequality is much tighter than the Chebychev inequality and the Chebychev inequality is distribution independent.

In the above we provide a statistical estimation of the small gallery size. Meanwhile, in our approach presented in Sect. 7.3.1, we learn the similarity

score distribution to find the optimal size of the small gallery. The small gallery size which we get from the statistics can be used as a guide for learning. Under the assumptions that the randomly chosen small galleries can represent the distributions of similarity scores for other galleries of the same size, we use different small galleries with the learned optimal size to predict large population performance. We randomly choose several small galleries of the optimal size to predict the large population performance. Then, we obtain the maximum and minimum prediction performance on the large population. In this way, we can provide the upper bound and the good lower bound for performance prediction on large populations.

7.4 Experimental Results

In all the experiments, we use fingerprints from the *NIST Special Database 4* (NIST-4). It consists of 2,000 pairs of fingerprints. Each of the fingerprints is labeled with an ID number preceded by an "f" or an "s," which represents different impressions of the same fingerprint. The images are collected by scanning inked fingerprints from paper. The resolution of the fingerprint image is 500 DPI and the size of the image is 480×512 pixels.

7.4.1 Prediction Model

Distorted Data

Since large populations will have distortions which may not be presented in the small gallery, we simulate the distortion in our prediction model to estimate the recognition performance based on small galleries. The minutiae features used for the fingerprint recognition can be expressed as $f = (x, y, c, d)$, where x and y are the locations of a minutiae, c is the class of the minutiae which represents whether the minutiae is endpoint (0) or bifurcation (1), and d is the direction of the minutiae. We define the amount of the minutiae distortion for a fingerprint as $g\%$. In this chapter, we choose $g = 5\%$. Assume the number of minutiae is num_j. Usually one pair of fingerprints has a different number of minutiae so $j = 1, 2, \ldots, 4000$. We apply the distortion model [13] to these 2,000 pairs of fingerprints as follows:

(a) *Uncertainty*: Uniformly choose $U = 5\% \times num_j$ minutiae features out of num_j features and replace each $f_i = (x, y, c, d)$ with f_i' chosen uniformly from the set $\{(x', y', c', d')\}$, where $(x', y') \in 4NEIGHBOR(x, y)$, $c' = 1 - c, d' = d \pm 3°, i = 1, 2, \ldots, U$.

(b) *Occlusion*: Uniformly choose $O = 5\% \times num_j$ minutiae features out of num_j features and remove these minutiae.

(c) *Clutter*: Add $C = 5\% \times num_j$ additional minutiae, where each minutiae is generated by picking a feature uniformly at random from the clutter region. Here we choose the clutter region as $CR = \{(x, y, c, d),\ 50 \leq x \leq 450,\ 60 \leq y \leq 480,\ c = \{0, 1, 2, 3, 4\}, 10° \leq d \leq 350°\}$.

In our experiments we use the uniform distribution as the uncertainty *PDF* and the clutter *PDF*. The number of features with uncertainty, occlusion, and clutter is the same. We use the algorithm provided in [23] to extract minutiae and algorithm [21] for matching.

Verification

We use the algorithm provided in [23] to extract minutiae. Suppose there are M and Q minutiae in the gallery and query fingerprints, respectively. Δ_{m} and Δ_{q} are potential corresponding triangles. We assume $F(s, \theta, t_x, t_y)$ is the transformation between the query and gallery fingerprints, where s is a scale parameter, θ is a rotation parameter, t_x and t_y are translation parameters. If these parameters are within limits [21], then we apply this transformation as the transformation between potential corresponding triangles Δ_{m} and Δ_{q}. The details of how to estimate the transformation parameters are explained in [21]. Based on the transformation $F(s, \theta, t_x, t_y)$, we compute the distance d by using equation (7.28),

$$d = \underset{i}{\arg\min} \left\{ \left| \hat{F}\left(\begin{bmatrix} x_{j,1} \\ x_{j,2} \end{bmatrix} \right) - \begin{bmatrix} y_{i,1} \\ y_{i,2} \end{bmatrix} \right| \right\} \tag{7.28}$$

where $(x_{j,1}, x_{j,2})$ and $(y_{i,1}, y_{i,2})$ are two sets of minutiae in the gallery and query fingerprints, $j = 1, 2, \ldots M$ and $i = 1, 2, \ldots Q$. If d is smaller than a threshold, then we can say that $(x_{j,1}, x_{j,2})$ and $(y_{i,1}, y_{i,2})$ are the corresponding points. If the number of corresponding points is greater than a threshold [21], then we define Δ_{m} and Δ_{q} as the corresponding triangles between the template and the query fingerprints. The final match score is the number of corresponding triangles between the query and template fingerprints.

Prediction Results

We randomly choose 50 pairs of fingerprints from two kinds of fingerprint pairs (with and without distortion) as our small gallery following a hypergeometric distribution. For this small gallery, we get 50 match scores and 2,450 nonmatch scores. After we obtain these similarity scores we use the EM algorithm to estimate the match score distribution and the nonmatch score distribution. The EM algorithm can find the number of components automatically [24] and for each component the EM algorithm finds its mean, variance, and weight. In this chapter, the similarity scores are the number of matched triangles between two fingerprints, the match scores are positive integers and the nonmatch

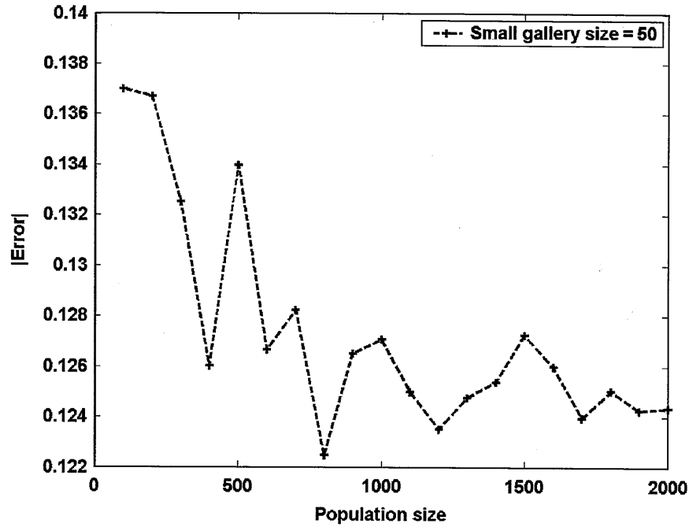


Fig. 7.2. Absolute error between the prediction and the actual performance when the small gallery size is 50

scores are close to 0. By applying the prediction model, we can estimate the fingerprints recognition performance on 2,000 pairs of fingerprints based on these 50 pairs of fingerprints. We repeat the experiment seven times and average the results to obtain the prediction performance which is shown in Fig. 7.2. Here, we choose the subset size $N_1 = 100$ and the margin of error $e = 0.06$. From this curve, we can see that for the large population size 100 the error between the prediction performance and the actual performance is 0.137, which is larger than the margin of error.

Now, we apply learning to the prediction process. We increase the small gallery size to $n = 100$. We feed back the match score and the nonmatch score from the randomly selected 100 pairs of fingerprint and repeat this process seven times. When the large population size is 100, the absolute error between the prediction performance and the actual performance is 0.135, which is greater than the margin of error 0.06. So, we increase the small gallery size to $n = 200$ and repeat the same process seven times. The absolute error is 0.09 when the large population size is 100. Then, we increase the small gallery size to $n = 300$ and repeat the same process seven times. The absolute error is 0.042 when the large population size is 100. We increase the large population size in steps of 100 until the large population size $N = 2,000$. For these three small galleries, most of the nonmatch scores are 0. Table 7.1 shows the

Table 7.1. Match score distributions estimated by the EM algorithm

Size	Component #	Mean	Variance	Weight
100	2	17.152658	334.452802	0.535764
		299.015489	55459.580193	0.450296
200	5	57.348298	1026.825771	0.160830
		3.615611	20.189071	0.362406
		585.278037	66686.529667	0.151087
		206.327514	7334.980411	0.191394
		27.106400	131.423073	0.133465
300	4	3.581775	21.569950	0.395165
		420.142835	64933.952657	0.236481
		35.420091	423.267100	0.228275
		143.774430	3016.000039	0.139634

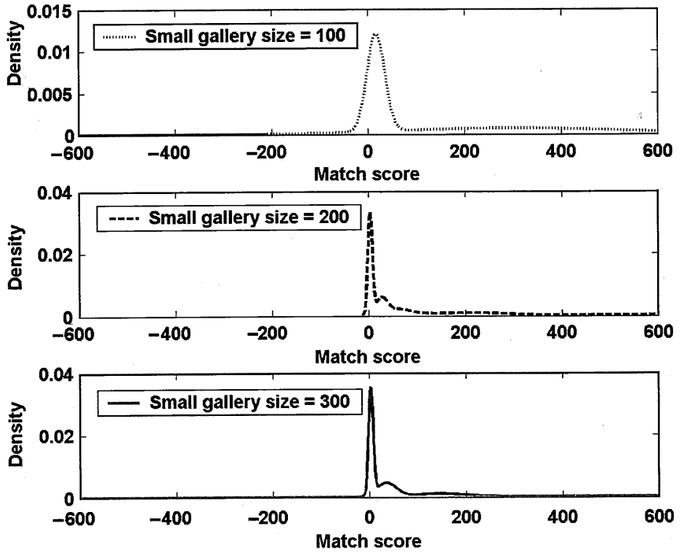


Fig. 7.3. Match score distributions for different small gallery sizes

estimation of the match score distributions with different small gallery sizes. The distributions are represented by the Gaussian mixture model. For each component we have its mean, covariance, and weight. Figure 7.3 shows the match score distribution curves on different small gallery sizes. For each small

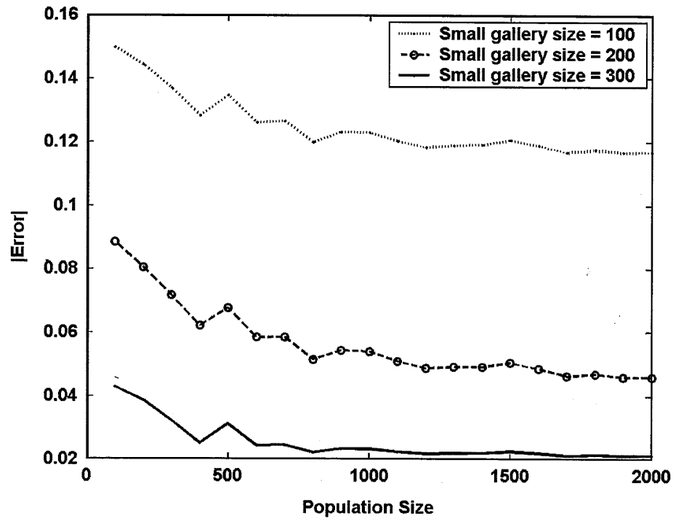


Fig. 7.4. Absolute error between the predicted and actual performance for different small gallery sizes

gallery size we provide two figures with different range of match score (X-axis) so that the distribution can be more closely examined. Figure 7.4 shows the absolute error between the prediction and the actual performance decreases when the gallery size increases. When the small gallery size is $n = 300$, the absolute error for the large population is smaller than the margin of error 0.06. At this point we can stop learning the small gallery size.

We use different small galleries with the learned optimal size to predict large population performance. Then, we select the maximum and the minimum prediction performance as our upper bound and lower bound for the performance prediction on the large population. Figure 7.5 gives the upper bound and lower bound on the prediction of large population performance when the small gallery size $n = 300$. Since we have 2,000 pairs of fingerprints, the actual recognition performance for the distorted images is shown in Fig. 7.5. Beyond this population size we can give the bounds for the prediction. From Fig. 7.5 it can be seen that the actual performance is within the upper bound and lower bound except when the population size is very small. Our experiments show that when the small gallery size is $n = 300$ the prediction error is less than 0.05.

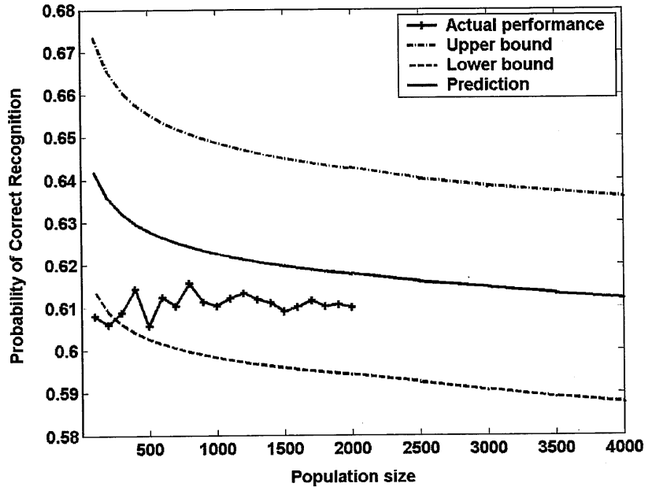


Fig. 7.5. The upper bound and lower bound on the large population when the small gallery size is 300. Note that the upper bound and lower bound are within 5%.

Table 7.2. Values of the confidence interval, the margin of error, and the small gallery size for Chernoff inequality and Chebychev inequality ($\sigma^2 = 1$)

$1-\alpha$	0.05	0.05	0.1	0.1	0.15	0.15
e	0.06	0.04	0.06	0.04	0.06	0.04
n(Chernoff)	417	937	320	720	264	593
n(Chebychev)	2,778	6,250	1,389	3,125	926	2,083

7.4.2 Estimation of The Small Gallery Size Based on Statistical Inequalities

Table 7.2 shows different small gallery sizes given different confidence intervals and margins of error for Chernoff inequality and Chebychev inequality ($\sigma^2 = 1$). From the table we ascertain that the Chernoff inequality is much tighter than the Chebychev inequality. We compare our learning small gallery size with the Chernoff inequality. When the confidence interval $\alpha = 95\%$ and margin of error $e = 0.06$ then the small gallery size $n = 417$. From our experiment for the same margin of error the small gallery size is 300 and the confidence interval is $\alpha = 95\%$. Note that statistical methods give us a loose estimate of the small gallery size. Based on our recognition system we find a more accurate small gallery size by learning.

7.5 Conclusions

We focused on the problem of performance characterization of a simplified CBR system. In particular, we addressed the following questions: what is the optimal size of the small gallery that can give good error estimation and what is the confidence in the estimation? We use a generalized prediction model that combines a hypergeometric probability distribution model with a binomial model, taking into account distortion in large populations. We incorporate learning in the prediction process to find the optimal small gallery size and provide the upper and lower bounds for the performance prediction on large populations. The Chernoff inequality and the Chebychev inequality are used as a guide to obtain the small gallery size and the confidence interval given a margin of error. Experimental results show that the small gallery size obtained from the statistical methods are loose compared to the size provided by the proposed learning method. We believe that the methodology and results of this research will be useful for a wide range of applications of CBR in signal processing, image processing, computer vision, and pattern recognition.

References

1. Kolodner J (1993) Case-based Reasoning. Morgan Kaufmann Publisher.
2. Perner P, Perner H, Jänichen S (2006) Recognition of airborne fungi spores in digital microscopic images. J Arificial Intelligence in Medicine AIM, Special Issue on CBR, vol. 36, no. 2, pp. 137–157.
3. Perner P, Jänichen S (2004) Case acquisition and case mining for case-based object recognition. In: Peter Funk, Pedro A. González Calero (Eds.), Advances in Case-Based Reasoning, ECCBR2004, Springer Verlag 2004, vol. 3155, pp. 616–629.
4. Ming J, Bhanu B (1997) ORACLE: An integrated learning approach for object recognition. J Pattern Recognition and Artificial Intelligence, vol. 11, no. 6, pp. 961–990.
5. Minor M, Hanft A (1999) Cases with a life-cycle. In: Proc. Intl. Conf. on Case-Based Reasoning Workshops, 1999, pp. 3–8.
6. Guyon I, Makhoul J (1998) What size test set gives good error rate estimates? J IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 52–64.
7. Mitchell TM (1997) Machine Learning. McGraw Hill.
8. Duda RO, Hart PE, Stork DG (2000) Pattern Classification. Wiley-Interscience Publication.
9. Wang R, Bhanu B (2007) Predicting fingerprint biometric performance from a small gallery. J Pattern Recognition Letters, vol. 28, pp. 40–48.
10. Wayman JL (1999) Error-rate equations for the general biometric system. J IEEE Robotics & Automation Magazine, vol. 6, issue 1, pp. 35–48.
11. Daugman J (2003) The importance of being random: statistical principles of iris recognition. J Pattern Recognition, vol. 36, no. 2, pp. 279–291.
12. Phillips PJ, Grother P, Micheals RJ, Blackburn DM, Tabassi E, and Bone M (2003) Face recognition vendor test 2002, Evaluation Report.

13. Wang R, Bhanu B, Chen H (2005) An integrated prediction model for biometrics. In: Proc. Audio- and Video-based Biometric Person Authentication, New York, pp. 355–364.
14. Johnson AY, Sun J, Boick AF (2003) Using similarity scores from a small gallery to estimate recognition performance for large galleries. In: Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures, pp. 100–103.
15. Grother P, Phillips PJ (2004) Models of large population recognition performance. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, pp. 68–75.
16. Lindenbaum M (1995) Bounds on shape recognition performance. J Pattern Analysis and Machine Intelligence, vol. 17, no. 7, pp. 666–680.
17. Lindenbaum M (1997) An integrated model for evaluating the amount of data required for reliable recognition. J Pattern Analysis and Machine Intelligence, vol. 19, no. 11, pp. 1251–1264.
18. Boshra M, Bhanu B (2000) Predicting performance of object recognition. J Pattern Analysis and Machine Intelligence, vol. 22, no. 9, pp. 956–969.
19. Boshra M, Bhanu B (2001) Predicting an upper bound on SAR ATR performance. J IEEE Trans. on Aerospace and Electronic Systems, vol. 37, no. 3, pp. 876–888.
20. Wang R, Bhanu B (2005) Learning models for predicting recognition performance. In: Proc. IEEE International Conf. on Computer Vision, vol. 2, pp. 1613–1618.
21. Tan X, Bhanu B (2002) Robust fingerprint identification. In: Proc. IEEE Int. Conf. on Image Processing, vol. 1, pp. 277–280.
22. Mood AM, Graybill FA, Boes DC (1974) Introduction to the Theory of Statistics. McGraw,Hill.
23. Bhanu B, Boshra M, Tan X (2000) Logical templates for feature extraction in fingerprint images. In: Proc. Int. Conf. on Pattern Recognition, vol. 3, pp. 850–854.
24. Figueiredo MAT, Jain AK (2002) Supervised learning of finite mixture models. J IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 3, pp. 381–396.

8

A CBR Agent for Monitoring the Carbon Dioxide Exchange Rate from Satellite Images

J.M. Corchado, J. Aiken, and J. Bajo

Departamento Informática y Automática
Universidad de Salamanca
Plaza de la Merced s/n, 37008, Salamanca, Spain
corchado@usal.es
Centre for Air–Sea Interactions and fluxes
Plymouth Marine Laboratory, Prospect Place, Plymouth, PL1 3 DH, UK
jai@mail.pml.ac.uk
Universidad Pontificia de Salamanca
Compañía 5, 37002, Salamanca, Spain
jbajope@upsa.es

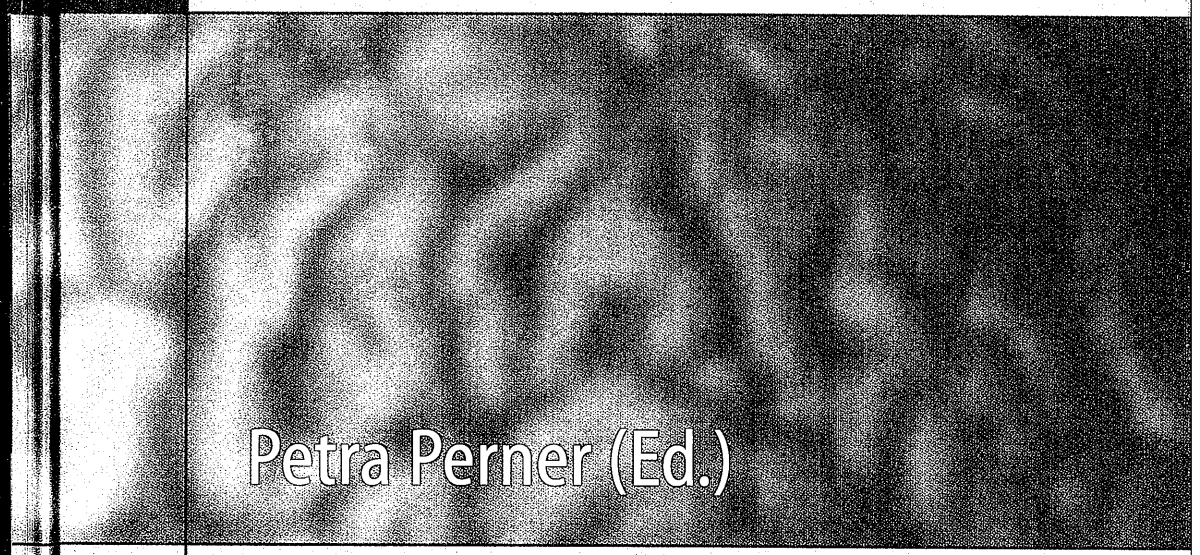
Summary. This work presents a multiagent system for evaluating automatically the interaction that exists between the atmosphere and the ocean surface. monitoring and evaluating within the ocean carbon dioxide exchange process is a function requiring working with a great amount of data: satellite images and in situ Vessel's data. The system presented in this work focuses on Ambient Intelligence (AmI) technologies since the vision of AmI assumes seamless, unobtrusive, and often invisible but also controllable interactions between humans and technology. The work presents the construction of an open multiagent architecture which, based on the use of deliberative agents incorporating Case-Based Reasoning (CBR) systems, offers a distributed model for such an interaction. This work also presents an analysis and design methodology that facilitates the implementation of CBR agent-based distributed artificial intelligent systems. Moreover, the architecture takes into account the fact that the working environment is dynamic and therefore it requires autonomous models that evolve overtime. In order to resolve this problem an intelligent environment has been developed, based on the use of CBR agents, which are capable of handling several goals, constructing plans from the data obtained through satellite images and research Vessels, acquiring knowledge, and of adapting to environmental changes, are incorporated. The artificial intelligence system has been successfully tested in the North Atlantic ocean, and the results obtained will be presented within this work.

8.1 Introduction

Ambient intelligent environments are characterized by their ubiquity, transparency, and intelligence [2]. The agents and multiagent systems (MASs) have become increasingly relevant for developing distributed and dynamic

J.M. Corchado et al.: *A CBR Agent for Monitoring the Carbon Dioxide Exchange Rate from Satellite Images*, Studies in Computational Intelligence (SCI) **73**, 213–246 (2008)
www.springerlink.com

Studies in Computational Intelligence 73



Petra Perner (Ed.)

Case-Based Reasoning on Images and Signals

Springer