Chapter 3

Tennis Style Detection

3.1 Experimental Design

In this experiment, we designed a simple simulator of tennis, to study different people's playing styles. The ball is served automatically from a random position in the upper half field with a random speed within a certain range and a random direction towards the bottom line. A human player can control the racket by moving the mouse. The speed of the racket is proportional to the speed of the mouse, and its orientation is perpendicular to the recent trajectory of the mouse.
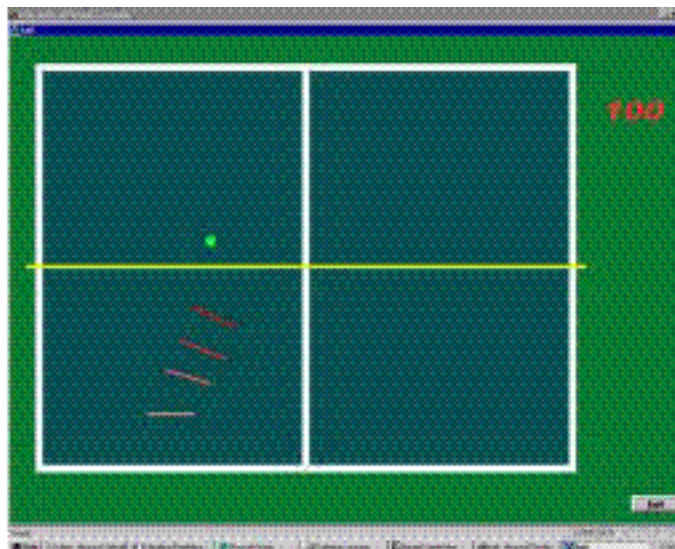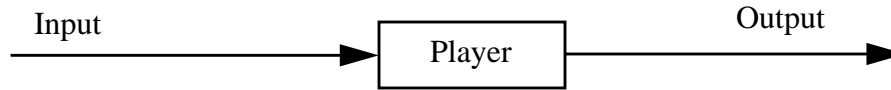

Figure 3-1: Tennis simulator interface.

Input   Output

Player

Input: serve position (x_s, y_s), serve speed (v_s) and orientation (θ_s).
Output: contact position (x_r, y_r), the ball's speed and orientation
    after the contact, (v_r, θ_r).

**Figure 3-2: The tennis simulation system is not dynamic because
there is no feedback from the outputs and no delays for the inputs.**

The short line segments in Figure 3-1 illustrate the recent movement of the racket. When the racket hits the ball, the ball is bounced back as a light beam is reflected by a mirror. Thus, the direction of the ball after contact is decided by both the orientation of the racket and the incident direction of the ball. Concerning the ball's emitted speed, it is decided by three factors: speed of the racket, the incident speed of the ball and the ball's incident angle with respect to the orientation of the racket.

This simulation system is not dynamic. Referring to Figure 3-2, if we regard the human player as a system, the input consists of four variables: the position where the ball is served by the computer, (x_s, y_s), the ball's speed (v_s) and orientation (θ_s) after the serve. The output includes the position where the contact between the racket and the ball happens, (x_r, y_r), the speed and orientation of the ball after the contact, $(v_r$ and $\theta_r)$. We only took records of those shots when the racket hits the ball. If the player was so careless that the racket missed the ball, we did not record that shot. We did not consider the ball's movement after the contact, because we were only interested in distinguishing the different playing styles, instead of evaluating the goodness and drawback of each style. Illustrated in Figure 3-2, there is no time delay in the input, and there is no feedback from the outputs, hence the system is not dynamic. In other words, the time order of the sequence of the data points, $(x_s, y_s, v_s, \theta_s, x_r, y_r, v_r, \theta_r)_t, t = 1, ..., T$, is not important; we can shuffle the order of the data points randomly.

Six people were invited to do the experiment. Each of them played twenty runs; and during each run, they gave one hundred shots. We did not use the data sets of the first three runs because the human players needed some time to learn how to play this game. We did not use the data set of the twentieth run, because when the players realized that they were close to the end, they did not pay enough attention to their performance, instead, they only wanted to finish the experiments as soon as possible. Thus, for each player, we have sixteen valid data sets.

We did not evaluate the merit of the performance, we only want to distinguish the different styles. However, it is an interesting but open question that if we evaluate the performance, whether or not people will adjust their styles so as to pursue higher scores; also, after a long time, whether or not different people will converge to the same style which is preferred by the evaluator.

The style is relevant to the distribution of the eight variables. Some people tended to hit the ball when the ball was close to the bottom line; the others gave a quick response once the ball came across the net. Some people wanted the ball to go in a direction as far as possible from the serving direction; others preferred the ball going back along the way it came, because this action is safer and easier. However, we cannot distinguish the styles only relying on the distribution of any one variable, because it is influenced by the other variables. As a matter of fact, we found that the speed of racket, v_r, was the best single feature to distinguish different players. But comparing with OMEGA, the single-feature-based classifier's accuracy is very low (Section 3.3).

Since there are six players, and each player has sixteen data sets, totally there are ninety-six data sets. Randomly we picked out one from the ninety-six datasets, and asked OMEGA to detect who was the underlying player by using the other ninety-five datasets as the training datasets. By comparing OMEGA's result with the real underlying player, we could tell for this data set, whether or not OMEGA's detection is correct. Similarly, we selected another data set to do this test, thus, we repeated the experiment for ninety-six times. The number of times that OMEGA succeeded to detect the correct underlying players can be used as a measurement of

OMEGA's accuracy. In the same way, we can measure the accuracy of the other methods, like the single-feature-based classifier.

3.2 OMEGA Result Analysis

This subsection discussed the experiment, which was to test if OMEGA could detect the underlying player correctly. We picked out one data set from each player's sixteen data sets as the testing set, and used the other fifteen data sets as the memory data sets. To detect who was the underlying player of the testing data set, OMEGA compared the testing data set with the six players' memory data sets one by one. Hence, we got six average negative log likelihoods, $\overline{-\text{lik}(S_p)}$'s. In Figure 3-3, 3-4, 3-5, the six curves correspond to the six players' $\overline{-\text{lik}(S_p)}$'s with respect to different numbers of data points involved in the calculation. The horizontal axis is the number of data points in the unlabeled data set. Thus, the tails of the $\overline{-\text{lik}(S_p)}$ curves tell who were most likely to be the underlying players.

Shown in Figure 3-3 (a) and (b), OMEGA detected Marianne and Colonel were the underlying players of the concerned data sets. These results are correct. For the ninety-six data sets, OMEGA did correct jobs for eighty-five times. It made mistakes for four times and was confused for seven times[1]. Figure 3-4 (a) shows a confused case, while Figure 3-4 (b) is a wrong one. Even in the wrong cases and the confused ones, OMEGA always found that the tails of the real players' likelihood curves were closer to the horizontal axis than most of the others.

Sometimes the likelihood curves are bumpy. This is because the player performed in an unusual way that hasn't been observed in memory. If a performance was so strange that it rarely happened to all the players, including the underlying player himself, then all the likelihood curves are bumpy, and roughly paralleling each other. In the case illustrated by Figure 3-4 (a), the ninth ball was served from a position very close to the right edge and also close to the net, with a sharp angle towards the left edge of the opposite field. Although the speed was not too fast, it

---

1. The definition of confusion refers to Chapter 2.2., Hypothesis testing, with significance level $\alpha = 5\%$.
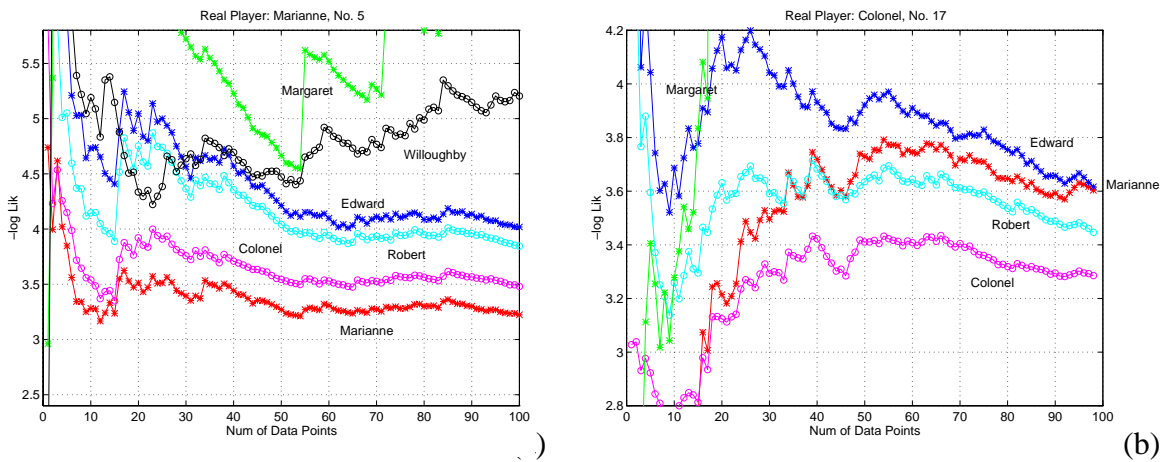
Figure 3-3: Likelihood curves of six human players. Two sample of the correct cases.
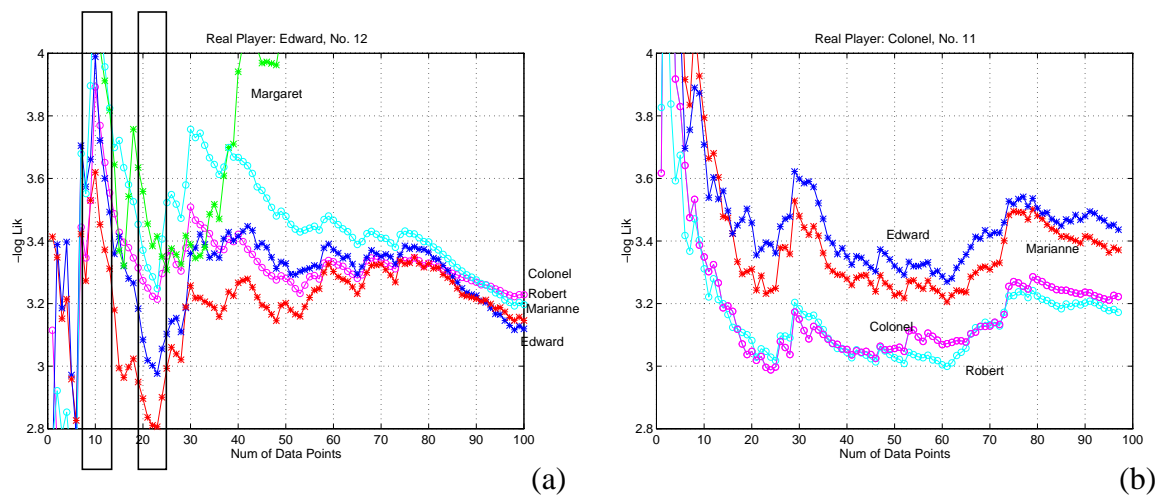

Figure 3-4: A confused case and a wrong case. **(a)** Confused: OMEGA can hardly distinguish Edward from Marianne. **(b)** Wrong: The real play should be Colonel, but OMEGA decided it was Robert. However, OMEGA did figure out Colonel was also very likely to be the player.

left Edward little time to react. Because Edward is right-handed, any ball coming from the right made him uncomfortable. Therefore, Edward's action for the ninth hit was totally a failure: the ball did not go across the net before it went out of the tennis court. Not only that, it seemed Edward did not recover from this shock until the twelfth hit. In the eleventh hit, he hardly touched the ball, because the ball's direction did not change too much after the contact. Hence, the likelihood curves in Figure 3-4 (a) rose to a peak at the eleventh hit. Fortunately, the twelfth
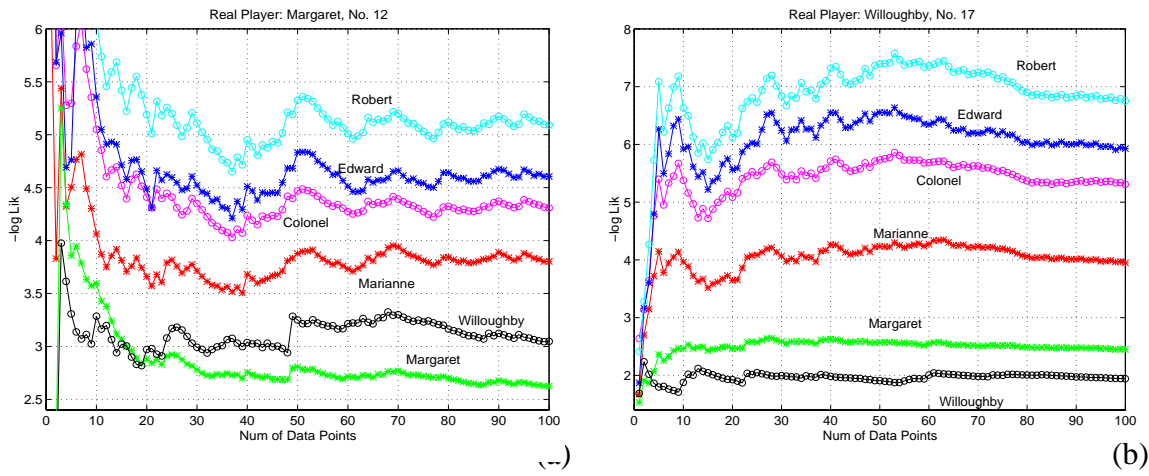
Figure 3-5: Willoughby and Margaret behaved similarly all the time. But they are different from others. Willoughby played more consistently, referring to his likelihood curve in (b) which is smoother than others.

ball was served in a manner Edward preferred: from the top left corner of the court towards the lower-right one, with a slow speed. This gave Edward a break to rebuild his confidence. He played in normal way again. Therefore, the likelihood curves start to go downhill. The twenty-second ball was another triumph. It started not far from center of the upper field, slowly and straightly downward. This was a great chance for Edward to exaggerate all his unique characteristics: he moved his racket rapidly to hit the ball when it arrived the center of the lower half field; after the contact, the ball rushed towards to the top right corner. Therefore, in Figure 3-4 (a), we see a great peak around the eleventh data point and a deep valley at the twenty-second.

The bumpiness implies the consistency of the players. Willoughby was the most consistent players among the six, because comparing Figure 3-5 (b) with other figures, Willoughby's curves are smoother than the others'.

The distances among the likelihood curves imply whose performances are similar. In this experiment, Margaret and Willoughby behaved similarly, referring to Figure 3-5 (a) and (b). But they are quite different from the others. As in Figure 3-3, 3-4, their curves were so much higher than the others that they are off the graphs.
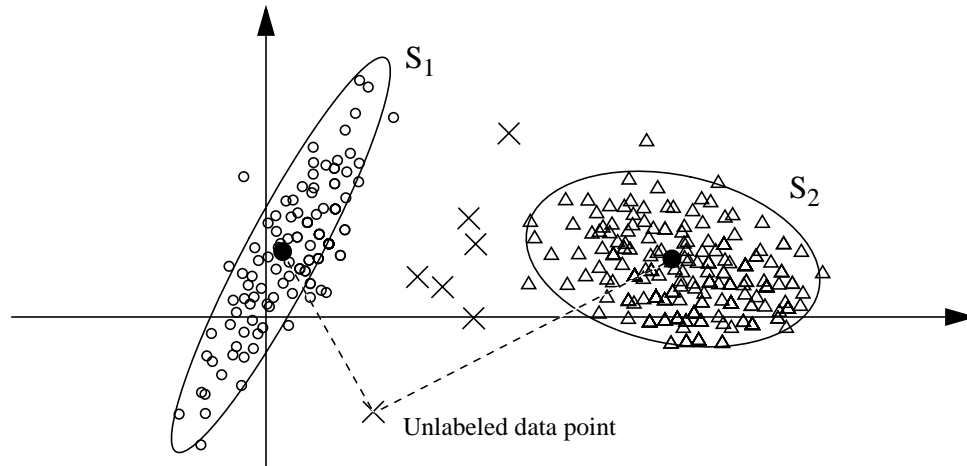
Figure 3-6: Bayes classifier assumes the distributions of the candidate system's data points are all of Gaussian distributions.

The likelihood curves tend to be more bumpy or chaotic at the beginning phase than afterwards. Recall that with limited testing data points, OMEGA is still able to start the classification job; but with more data points, OMEGA may improve its precision. Therefore, OMEGA is an ideal on-line classification technique.

3.3 Comparison with Other Methods

In this section we compare OMEGA's performance with those of other methods, like Bayes classifier and linear regression, because Bayes classifier is a popular statistical classifier while linear regression represents the linear control system approach. We also used the best single feature to do the classification. The purpose was to show that it is not easy to distinguish different tennis playing styles.

Bayes classifier

Bayes classifier assumes the memory data points of each candidate system are of Gaussian distribution, in plain words, each candidate system's memory data points cluster in a shape more or less like an ellipse. In Figure 3-6, there are two candidate systems, S_1 and S_2, whose data

points are represented by the circles and the triangles respectively. The horizontal axis may be the input, the vertical one may be the output; but this is not a requirement. As a matter of fact, Bayes classifier does not distinguish the input and output, instead, it treats all the input and output as features. By adjusting the scales of the axes, Bayes classifier can discriminate the importance of different features. In Figure 3-6, if the scales of the axes are changed, the elliptical shape of the clusters will be different. To classify an unlabeled data point, like the cross in Figure 3-6, we can measure the distances from the unlabeled data point to the centroids of the elliptical clusters. The shortest distance indicates to which candidate system (represented by the ellipse) the unlabeled data point belong to. Given a set of unlabeled data points, we can do the classification one by one, then make an overall judgement.

The Gaussian assumption of Bayes classifier is too restrictive for the tennis style domain. Therefore, Bayes classifier's performance as shown in Table 3-1 is very poor compared with OMEGA.

Linear regression approach

Linear regression assumes the function relationship between the inputs and the outputs are linear. Furthermore, *global* linear regression assumes the function relationship (the parameters of the function) is fixed anywhere around the input space. If the function parameters of a certain system is distinguishable from the others, the classification job is feasible. In this experiment, we did the global linear regression of each candidate system based on its memory data points. In other words, we determined the parameters, β's, in the following linear equations for every candidate system,

$$\begin{cases} x_r = \beta_{10} + \beta_{11}x_s + \beta_{12}y_s + \beta_{13}v_s + \beta_{14}\theta_s + \xi_1 \\ y_r = \beta_{20} + \beta_{21}x_s + \beta_{22}y_s + \beta_{23}v_s + \beta_{24}\theta_s + \xi_2 \\ v_r = \beta_{30} + \beta_{31}x_s + \beta_{32}y_s + \beta_{33}v_s + \beta_{34}\theta_s + \xi_3 \\ \theta_r = \beta_{40} + \beta_{41}x_s + \beta_{42}y_s + \beta_{43}v_s + \beta_{44}\theta_s + \xi_4 \end{cases} \tag{3-1}$$

in which the definitions of the input variables, x_s, y_s, v_s, θ_s, and the output variables x_r, y_r, v_r, θ_r refer to Section 3.1. When a unlabeled data set came, to detect its underlying player, we temporarily assume the unlabeled set was generated by the first player. Since we have already estimated the first player's function parameters (the β's), we picked out a data point $(x_s, y_s, v_s, \theta_s, x_r, y_r, v_r, \theta_r)_t$ from the unlabeled data set, we could predict the outputs $(x_r, y_r, v_r, \theta_r)_t$ corresponding to the input $(x_s, y_s, v_s, \theta_s)_t$. If the residual between the predicted outputs and "real" observed output is small, the first player is likely to be the underlying player. We repeated this test with respect to all the six players, the smallest residual responds to the most likely player.

We used the estimated β's to predict the outputs, then compare the predicted outputs with the real outputs. Usually there is a residual between the predictions and the real outputs. The system with the least residuals is most likely to be the underlying system which generates the testing dataset.

Referring to Table 3-1, global linear regression can hardly distinguish the variant human players, because in most cases, global linear regression is "confused". To improve it, we can do two things: (1) We can extend the linear equations in Equation 3-1 to polynomials with higher degrees. In this way, the function is capable of describing more complicated relationship between the input and output. (2) Instead of assuming there is one fixed global linear function, we can assume in any local region, the input and output relationship is linear, but the linear parameters may vary with different inputs.

In Table 3-1, we notice that quadratic model does not do any better than the linear models, but local paradigm does help. However, the local approach, even the local models with quadratic items, is still worse than OMEGA by a large margin. The reason is that in this tennis playing style domain, even for the identical serves, the same player may react in different ways. That means, the conditional distribution of the output with respect to a certain input may be of multimodal, instead of uni-modal as the linear model assumes. Therefore, the linear models are not proper for the tennis playing style domain, either.

Table 3-1: Comparison experiment for tennis domain

	Correct	*Wrong*	*Confused*
One Feature	21	57	18
Bayes	34	40	22
k-Nearest Neighbors[a]	17	14	67
Global Linear	9	12	75
Global Quadratic	9	12	75
Local Linear	17	8	71
Local Quadratic	20	5	71
OMEGA	85	4	7

a. We used 9 nearest neighbors here. Also, we tried 3 nearest neighbors as well as 6, the results do not deviate from those values in the table significantly.

3.4 Summary

In this chapter, we used OMEGA to classify different human operators' behavior in a game mimicking tennis. Although the simulation system is not dynamic, the classification job is not easy, especially because the distribution of the input and output is complicated. OMEGA performs very well in this domain, which demonstrates that OMEGA is a good classification technique. Although originally it was explored to classify time series, OMEGA is also a general purpose classification tool, which is capable of handling both time series and non-time series.

Experiments have been done to compare OMEGA with other methods. Although we have tuned up those methods to perform as well as possible, they still are not competitive with OMEGA.