ELG5131 project 2

Name: Siqi Ma

Student number:7070144

2013/4/7

1.Review of the command of project 2

There's 8 field and 300 players. The size of all the fields are 512× 512. Each field contains four different kinds of difficulty area, from 0 to 3. One's moving speed is decided by the difficulty of the area. One go from the left and up corner to the right down corner. At each step, the player can choose to either go forward or go back. Assume at first, the coordinate is (0,0), at last, the coordinate is (512,512). Everyone is allowed to make 1300 moves, otherwise he will fail in the walking. The expectation of this project is to make a count of how many strategies the players used to cross the field.

2.The properties used in the project

2.1 The way to do the project

Form coordinate (0,0) to coordinate (512,512), it goes down and right sided in the whole. Divide the movements in two types. Assume the initial coordinate of the players are (0,0), right movement = 0 and down movement= 0.

If one go from one point to the right or to the right and down side, as shown in figure 2.1.1, then assume the player move one step to

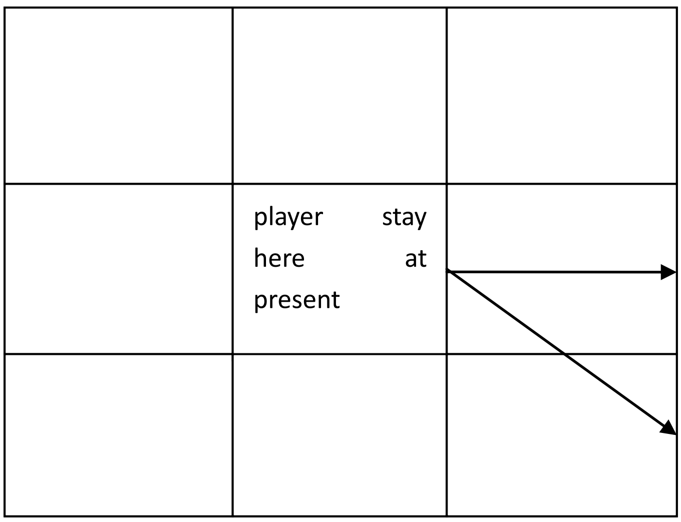

figure 2.1.1

the right direction one time, add right movement by one.

If the player goes down next step, assume the player move down by one time ,as shown in figure 2.1.2, add the down movement by one.

otherwise, ignore the movement because at least in one direction, the player move backward to the destination.

In every field, repeat it for 300 times because there're 300 players. When the next coordinate is (0,0), it means that the former player has finished the walking or the player has failed in the working.

From the two kinds of movements, it is obviously that the whole movement for the coordinate (0,0) to the coordinate (512,512) is :

whole movement = right movement + down movement

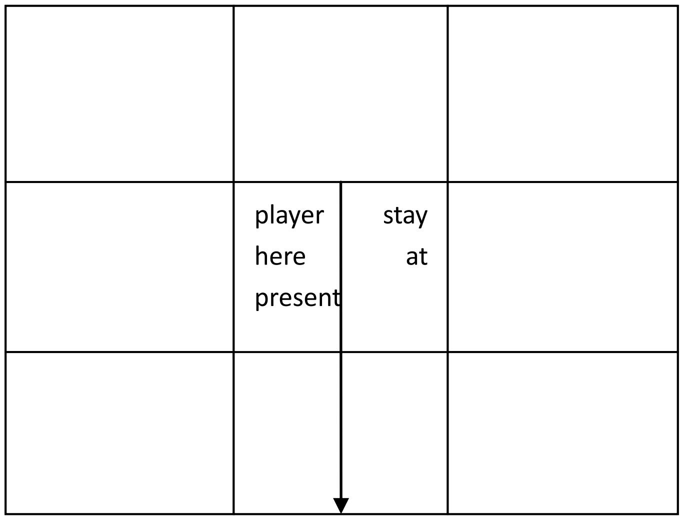

figure 2.1.2

Here plot the movements towards right and down for every player, get a scatter picture. As the question given, the speed one player move depends on the difficulty of the area. The players who move to the destination with similar policy. To confirm this and get a more accurate result, using the EM algorithm to separate the 300 players' strategy of how to move to the place whose coordinate is (512,512) in every field.

2.2 EM algorithm

Maximum expected (EM) algorithm is probabilistic (probabilistic) model to find maximum likelihood parameter estimation or maximum a posteriori estimation algorithm, probabilistic model which can't be observed depends on the hidden variable (Latent Variable)

Let $X = X_1, X_2 \ldots X_n$ be the set of observed random variables. $Z = Z_1, Z_2 \ldots Z_n$ be the set of hidden random variables. $\theta = \theta_1, \theta_2 \ldots \theta_n$ be the set of unknown parameters.

Suppose the model is $P(X, Z|\theta)$, find θ which maximize the following likelihood function using ML:

$$P(\theta_j X) = \mathrm{logP}(X|\theta) = log \sum_z P(X, Z|\theta)$$

Let q(z) be an arbitrary distribution of Z, which may not have any physical meaning. They in fact store as auxiliary variables in the optimization

$$L(q, \theta) = log \sum_z q(z) \log\left(\frac{P(x, z|\theta)}{q(z)}\right)$$

the goal of the EM algorithm is to find the θ maximize the $L(q,\theta)$ by coordinate ascent. Assume the $\theta = (\mu_k, \varphi_k, \alpha_k | k = 1,2,3 \ldots)$, k is the number of clusters to be known. $\mu_k$is the mean, φ_k is the covariance matrix, and α_k is the mixed proportion. Which is shown as figure 2.2.1

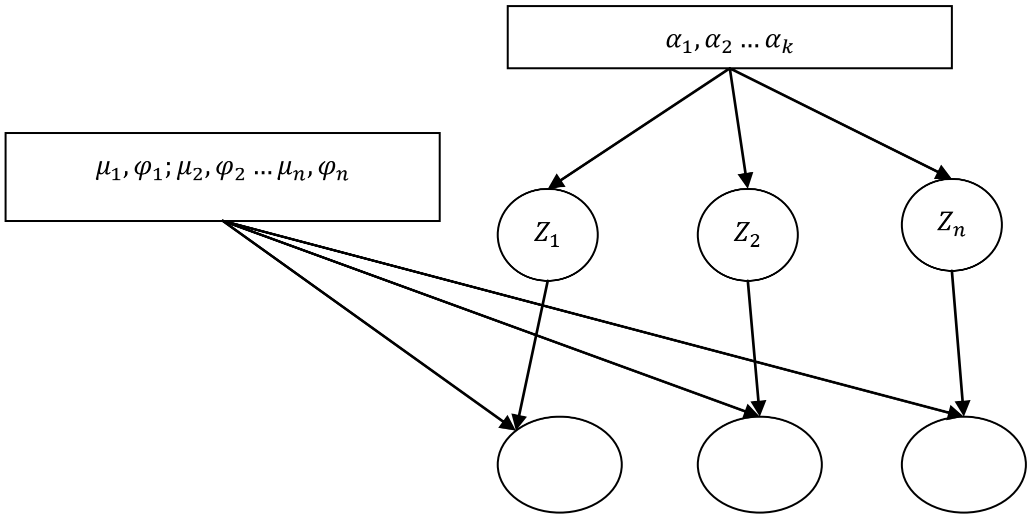

figure 2.2.1

E-step:

$$q^{t+1}(Z_1 \ldots Z_n) = p(Z_1 \ldots Z_n | X_1 \ldots X_n, \theta^t) = \prod_{i=1}^{n} P(Z_i | X_i, \mu^t, \varphi^t, \alpha^t)$$

Denote

$$\mathrm{q}_\mathrm{i}^{t+1}(\mathrm{Z_i}) = \mathrm{p}(\mathrm{Z_i} | X_i, \mu^t, \varphi^t, \alpha^t)$$

then

$$q^{t+1}(Z_1 \ldots Z_n) = \prod_{i=1}^{n} q_i^{t+1}(Z_i).$$

for each fixed i,

$$q_i^{t+1} Z^i = \frac{\alpha_z^i N(X_i, \mu_{Z_1}^t, \varphi_{Z_2}^t)}{\sum_{k=1}^{K} \alpha_k^t \, N(X_i, \mu_k^t, \varphi_k^t)}$$

Find θ to maximize L, the solution is

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{n} q_i^{t+1}(k) X_i}{\sum_{i=1}^{n} q_i^{t+1}(k)}$$

$$\alpha_k^{t+1} = \frac{\sum_{i=1}^{n} q_i^{t+1}(k)}{N}$$

3. Simulation and the result

As the players each use the same strategy in the 8 fields, analysis can be down on one field. Choose field 8 as a sample.

First , plot the scatter of how many kinds of walking the players take. The result is as figure 3.1. From the figure, there're three places the stars gather with each other. And there're still some single stars.

Then calculate the probability separately that each play moves right and down in the whole movements. plot the scatter picture again. The distribution of the probability is shown as figure 3.2

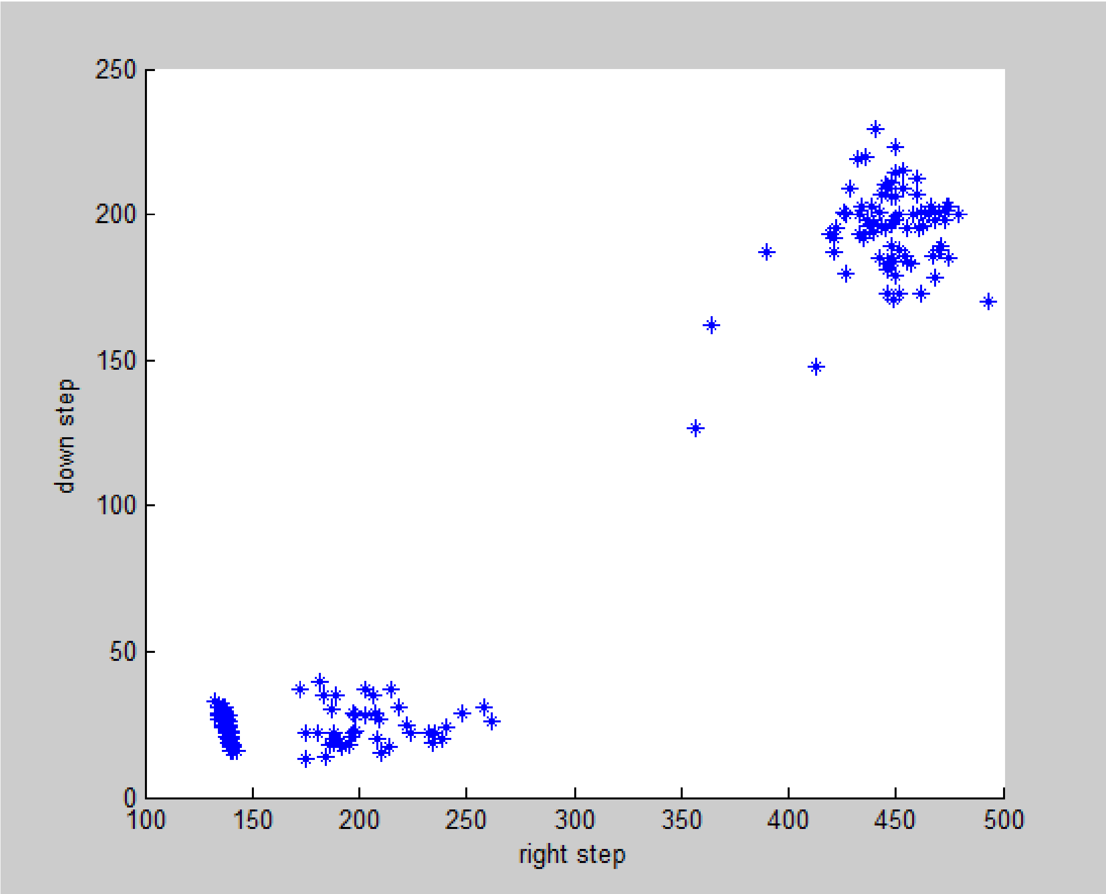

figure 3.1

From figure 3.1, It is shown that it is almost distributed into 3 or 4 parts, so choose 4 as the parameter of the separate. In MATLAB, we can use the function gmdistribution.fit to estimate Then using the EM algorithm to do the separation for the probability, we got

Gaussian mixture distribution with 4 components in 2 dimensions

Component 1:
Mixing proportion: 0.117971
Mean: 0.0261 0.0073

Component 2:
Mixing proportion: 0.035528
Mean: 0.2342 0.0733

Component 3:
Mixing proportion: 0.381856

Mean: 0.0081 0.0024

Component 4:
Mixing proportion: 0.464646
Mean: 0.0026 0.0004

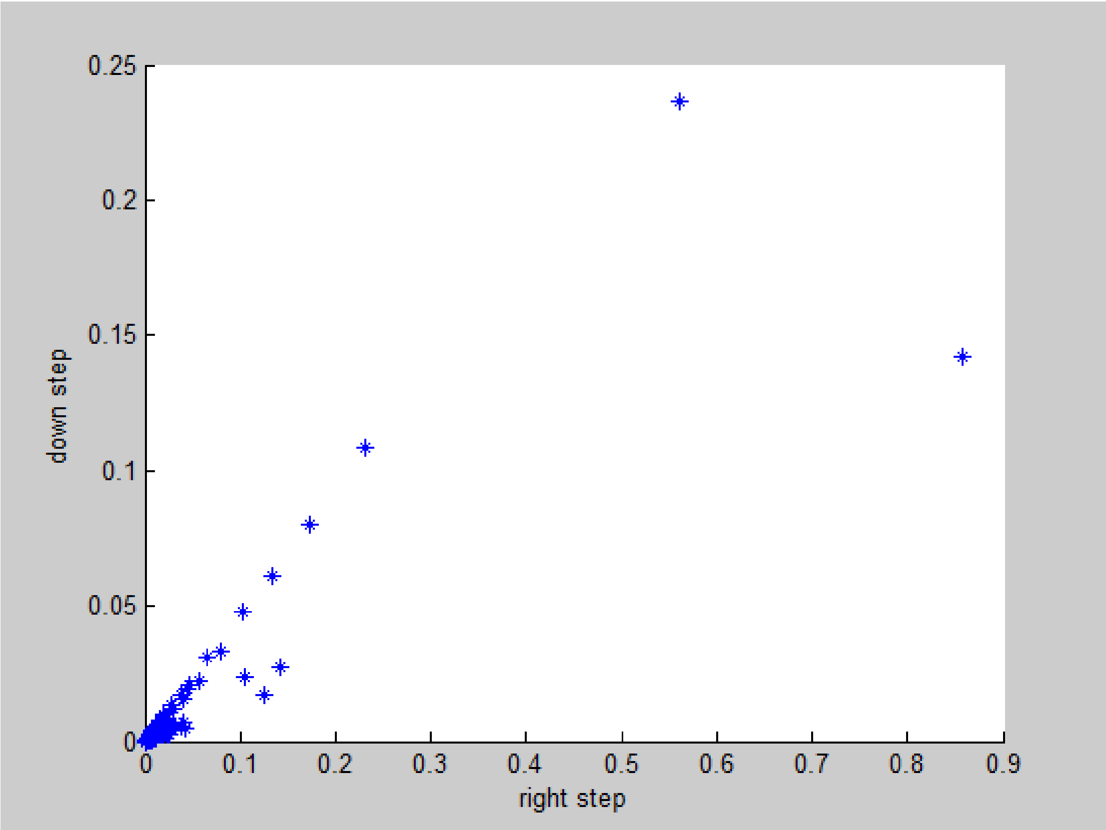

figure 3.2

Appendix

code:

```
load('playOnfield.mat')
[a,b]=size(playOnfield8);
as=playOnfield8;
right=0;
down=0;
distance=0;
j=1;
pr = zeros(300,1); %the probability of going right
pd = zeros(300,1); %the probability of going down
for i=2:158711
    if ((as(i+1,1) == 0 && as(i+1,2) == 0) || (as(i,1) == 255 && as(i,2)
== 255))
        %one person reach the destination or failed in the walking
        %distance = right + down;
        pr(j,1)=right/distance;
        pd(j,1)=down/distance;
        a1(j,1) = right;
        a2(j,1) = down;
        j = j+1;
        right = 0;
        down = 0;
    else if as(i,1) > as(i-1,1) && as(i,2) == as(i-1,2)
            %go to the right, plus right and distance
        right=right+1;
        distance=distance+1;
        else if as(i,1) > as(i-1,1) &&as(i,2) > as(i-1,2)
                %go to the right down side, plus right, and distance
            right=right+1;
            %down=down+1;
            distance=distance+1;
            else if as(i,1) == as(i-1,1) && as(i,2)>as(i-1,2)
                    %go to the dowm, plus down and distance
                down=down+1;
                distance=distance+1;
                else
                    right = right;
                    down = down;
                    distance = distance;
            end
        end
```

```
        end
    end
end
prob_matrix=[pr,pd]

%way_of_all_people=[pr,pd];% the percentage of everyperson go down and
go right
c = 0
scatter(a1,a2,'*') %plot the distribution of the steps towards right and
%down.
hold on
xlabel('right step');
ylabel('down step');
figure
scatter(pr*100,pd*100,'*');
xlabel('probability of going right(%)');
ylabel('probability of going down (%)');

%EM algorithm
obj = gmdistribution.fit(prob_matrix,4)
```