**Zadanie I**

1. 4x3, gamma=1, r=-0.04

**Iteracji**: 25

**Uzytecznosci**:

0.8116, 0.8678, 0.9178, 1

0.7616, 0, 0.6603, -1

0.7053, 0.6553, 0.6114, 0.3879

**Polityka**:

> > > -

^ - ^ -

^ < < <

1. 4x4, gamma=0.99, T=+100, B=-20, r=-1

**Iteracji**: 35

**Uzytecznosci**:

81.9384, 84.2610, 86.5861, 88.8827

81.7354, 84.2724, 87.0596, 91.5547

79.5936, 80.5997, 70.4670, 94.5352

77.4526, 78.2495, 0.0000, 100.0000

**Polityka**:

> > > v

> > > v

^ ^ > v

^ ^ - -

**Wnioski**: polityka omija stan specjalny bo ma niska wartość.

1. 4x4, gamma=0.99, T=+100, B=-1, r=-2

**Iteracji**: 34

**Uzytecznosci**:

79.1789, 82.3571, 85.5201, 87.9433

81.9534, 85.5739, 89.2176, 91.8819

84.2444, 88.4864, 92.9918, 95.9003

81.3109, 84.4961, 0.0000, 100.0000

**Polityka**:

> > v v

> > v v

> > > v

> ^ - -

**Wnioski:** polityka uległa zmianie nie omijamy stanu specjalnego tylko leziemy szybciutko do celu.

1. 4x4, gamma=0.99, T=+100, B=-20, r=-1, p1 = 0.4, p2 = 0.3, p3 = 0.3

**Iteracji**: 196

**Uzytecznosci**:

51.1199, 55.6857, 60.9997, 65.6003

50.1204, 53.8423, 60.2863, 71.1761

46.7156, 47.1158, 43.5009, 82.5154

43.2178, 43.3763, 0.0000, 100.0000

**Polityka**:

> > > >

^ ^ ^ >

^ ^ > >

^ ^ - -

**Wnioski:** polityka uległa zmianie omijamy stan specjalny większym łukiem bo są duże szanse że w niego wpadniemy nawet idąc obok niego.

1. 4x4, gamma=0.9, T=+100, B=-20, r=-1

**Iteracji**: 28

**Uzytecznosci**:

38.4392, 44.8395, 51.9198, 59.6666

41.0585, 49.1295, 58.7455, 70.3109

35.8411, 42.1918, 49.4317, 82.9108

30.7525, 35.3251, 0.0000, 100.0000

**Polityka**:

> > > v

> > > v

> > > v

^ ^ - -

**Wnioski:** polityka uległa zmianie idziemy bezpośrednio do stanu terminalnego.

**Zadanie II**

**e = 0.05**:

**Przebiegow**: 946690

**Uzytecznosci**:

75.84, 84.26, 86.81, 89.17,

79.92, 84.33, 87.77, 92.19,

75.76, 79.59, 90.65, 95.52,

73.04, 74.88, 0.00, 0.00,

**Polityka**:

> > > v

> > > v

> ^ > v

^ ^ - -

**e=0.2:**

**Przebiegow**: 810878

**Uzytecznosci**:

81.71, 84.26, 86.77, 89.15,

81.59, 84.59, 87.78, 92.18,

78.96, 80.13, 90.61, 95.50,

75.00, 74.78, 0.00, 0.00,

**Polityka**:

> > > v

> > > v

^ ^ > v

^ ^ - -

**Wnioski:** polityka jest taka sama jak w przypadku iteracji wartości z pełna wiedzą. Jednakże wartości użyteczności są nieznacznie inne. Większe e(częstsza eksploracja) zwiększa szybkość zbiegania się użyteczności poza optymalną ścieżką moim zdaniem powinna być nawet na poziomie 0.5. Potrzeba naprawdę wielu przebiegów aby użyteczności się zbiegły.

Obsługa i uruchomienie programu:

Aby program działał poprawnie potrzebny jest plik planszy w którym

najpierw mamy wymiary a potem wygląd planszy np.:

*4 4*

*+-------+*

*| | | | |*

*+-+-+-+-+*

*| | | | |*

*+-+-+-+-+*

*| | |B| |*

*+-+-+-+-+*

*|S| |F|T|*

*+-------+*

Następnie jest wartość domyślnej wartości nagrody

Potem wszystkie pozycje specjalne i terminalne w formie

x, y, r (0,0-gorny lewy róg)

na koniec p1, p2, p3

Oprócz pliku planszy jest potrzebny plik konfiguracyjny o nazwie config.txt w formie

*board\_file board.txt* (nawa pliku planszy)

*gamma 0.99* (gamma)

*e 0.05* (exploracyjność)

*iteracja\_wartosci 0* (czy używamy iteracji wartośći)

*q\_learning 1* (czy q-learning-u)

program uruchamia się z jednym parametrem a mianowicie ścieżką do folderu z plikiem konfiguracyjnym i plikiem planszy jeśli nie podamy parametru brany jest aktualny katalog. Po zakończeniu otrzymujemy wyniki w pliku result.txt oraz dane w pliku data.dat. Jeśli w folderze obok wyników umieścimy skrypty gnuplota *col\_counter.gp*, *plot.gp* możemy odpalić ten drugi i otrzymamy ladny wykres użyteczności.