Journal of Industrial Engineering 5(2010) 35-41

Outsourcing or Insourcing of Transportation System Evaluation Using Intelligent Agents Approach

Isa Nakhaei Kamalabadi, Parham Azimi, Mohammad Varmaghani*

Islamic Azad Univercity, Qazvin Branch, Department of Industrial Engineering,Qazvin, Iran

Received 15 Nov., 2009; Revised 23 Nov., 2009; Accepted 5 Dec., 2009

Abstract

Nowadays, outsourcing is viewed as a trade strategy and organizations tend to adopt new strategies to achieve competitive advantages in the current world of business. focusing on main copmpetencies, and transferring most of activities to outside resources of organization(outsourcing) is one such strategy is. In this paper, we aim to decide on decision maker agent of transportation system, by applying intelligent agent technology and using learning model which is modeled as a reinforcement learning problem. A Q-learning algorithm is proposed to solve the RL model. Results show that the proposed model given its ability to communicate with environment, adaptability with environment and correcting itself based on learnt data ,the prposed model can be applied as a better and quicker learning model in comparison with other ways of solving of decision making problems.

Key words: Transportation system, Outsourcing, Agent, Reinforcement learning, Pattern x +y.

1. Introduction

One of the main problems faced by product-service units is the high cost of transportation which both results in the loss of enrgy and time and augment the the investment costs such as transprtation equipmen, labor force, storage and safety, thus leading to higher price of final product and consequently compromised qulaity of services and ill-time delivery of services.

Hence, due to competetive pressures, business difficulies, resource limitations, jobs specialization, increased costs and so on, organizations need to reconsider their management policies so as to be able achieve competitive advantages by outresourcing a major part of their activities.

In this paper, the decision making aimed at selecting suppliers is based on two criteri of delivery cost and time which are the most important economic parameters in the context of transportation engineering. An attempt is made to minimize these two crirtria,namely, cost and udartion.

To this end, we consider an agent whose characteristics are defined. Then, using the reinforcement learning, we verify whether outsourcing is economical or not In this model, how agent interacts with each supplier separately is evaluated for purpose of appropraitely making decision regarding the outseorcing or insourcing transpoaraiton system.

* Corresponding author E-mail: mohammad.varmaghani@yahoo.com

2. Literature review

In 1937, an economist named Ranahay Kawz wrote an article named “the nature of company” [7]. He raised the question of why companies regulate their selected structure. He concluded that such issue has has to do with balancing between marketing access costs and problem of non economic scales. This is the time when organization outgrow. During 1975-1985 he developed a conception called “transparency of properties” based on which , exchange cost is determined by characteristics of exchanged goods as well as investment of mother company in supplier company [7].

In 1982, the newly developed organizational strategies emphasized that organizatons need to focus on their main business. This resulted in the development of a new concept of what Peters called "competency-centered "

In 1980, western car manufactures got engaged in new discussions called “trade -centered” talks [7]. At first many companies utilized this approach just for their support activities.

In the following years, various articles dealing with outsourcing were written. They came under the title" making decision on purchasing or manufacturing".

For the first time, Padilo and Daibi viewed outsourcing from a multi- criteria perspective. [9]. They presented a seven -step, multi criteria model of decision analysis for evaluation of purchasing and manufacturing strategies . Some articles studied outsourcing risks. These studies attributed the change in companies's attitude to this approach to the following factors : "hidden costs of outsourcing", lack of transparency in suppliers costs, impossibility of using internal resource for other purposes, supplier's incapability, "complexitites of such approach and the necessity of strict management" .
Lance Del presented a useful conceptual framework for efficient management of outsourcing risks which focuses on the incorporation of competitive advantages of organization. [3]. In some cases, market demand may exceed the production capacity of companies and manager must decide on how much to produce and how much to buy from outside contravctors. Coman and et all, presented a model which considered such a position [2]. They studied outsourcing problem, based on finacial and capacity parameters and converted the problem to a linear programming. To sum up, in realted literaure, decision making regarding outsourcing and insourcing is a multi criteria decision problem (MCDM) which are solved , using linear programming methods, Analytical Hierarchy Process (AHP), Analytical Network Process (ANP), DEMATEL technique or a combination of mentioned methods in fuzzy environments. This innovative article evaluates this problem, using intelligent agent with respect to transportation system.

3. Reinforcement Learning Model

Reinforcement learning is defined as teaching actions which maximizes a numerical reward signal [8l.
A reinforcement Learning agent is characterized by a knowledge structure, a learning method or rule to update its knowledge and a specific behaviour (policy) [8]. In general, an RL system is considered to be a Markov or Semi markov decision process where the action is controlled by an agent. The simulated environment is characterized by states, rewards and transitions, as discussed in detail below.

Fig. 1 summarizes the communication between the learning agent and its simulated environment. At each decision step, the agent observes the state S_t of the environment and performs an action a_t, selected according to its current decision policy π. As a result of the action taken, the simulated environment makes a transition to a new state S_{t+1} and a reinforcement or reward signal is generated r_t, the reward signal and the new state are received by the RL-agent and through its learning rule is used to update its knowledge about the environment, and consequently it can update its decision making policy π, reward and state transition functions may be in general stochastic, and the underlying probability distribution nor a model of them are assumed not to be known to the RL – agent.

Reinforcement learning is not defined by characteristics of algorithm learning, it is explained by problem of learning features instead. A learning agent in this algorithm must specifically be able to recognize environment state. It reacts against the identified state and beyond this problem, a reinforcement learning agent must have goal or goals concerned with environment.

In fact, the reinforcement learning formulization algorithm includes three aspect of environment perception, performance of the action and purpose [12].

Striking a balance between exploration and exploitation- a concern not covered by other learning methods- is one of the challenges faced by reinforcement learning. In order to obtain far more rewards, an action must be tested several times to obtain an estimate of expected reward. To do so ,a reinforcement learning agent should not only search for its learning results but it also needs to find better responses. It is self evident that, at the beginning of learning process it is necessary to have a high exploration probability [11,5]. This is beacause agent doesn't have savings for exploitation. The more completed learning process leads to the less possibility of exploration, and hence to more exploitation. In other words, agent eploites more experiences and it relies less on chance. If probability of exploration is shown by $P_s$and exploitation probabilty by $P_e$then in each time step this pattern can be written:

$$Ps(i) + Pe(i) = 1 \qquad (1)$$

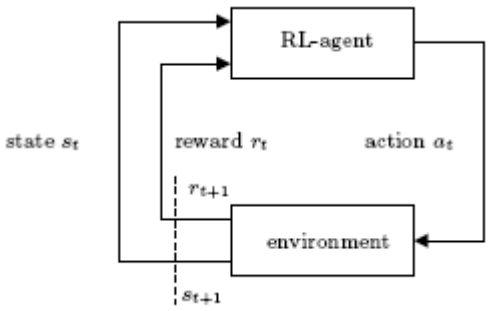

Fig. 1. RL-agent in environment

3.1. Markov decision process

In framework of reinforcement learning, agent makes its decision based on function of signal received from environment which is called environment state. In most cases a series of hidden information exist in environment. In case, agent is aware of them decision making will be approached to optimally, nevertheless agent has no access to those information. A state signal which can hold all related information is called Markov or has Markov properties [12]. This is because the actions have

unfolded in a way which is independent from the route leading to this state. [5]. when there are limited number of states and values in reinforcement learning problem, it can be said that probability of next action depends only on current state and it is independent from the path that has led it to current state [5].

Formula definition of markov property is:

$$\left[\begin{array}{l} p\left\{s_{t+1}=s', r_{t+1}=r \middle| s_t, a_t, r_t, s_{t-1}, a_{t-1}, r_{t-1}, \ldots s_0, a_o, r_0\right\} \\ p\left\{s_{t+1}=s', r_{t+1}=r \middle| s_t, a_t\right\} \end{array}\right] \quad (2)$$

S_t:system status in term t

a_t: the performed action by agent in term t

in other word ,a State signal has markov features only if two above relations are equal.

3.2. Q-Learning algorithm

Q-Learning is one of the Time Difference (TD) methods, so named because the reward and the value of the current state are used to improve the estimate of the previous state [11].

In fact, this kind of learning maps each of state-action pair to an extend called Q-value which is illustrated by $Q(s,a)$. It is also include sum of received rewards when we begin by state and performing action and following such process.

Q-learning can be expressed as follows:

$$Q\left(s_t, a_t\right) \leftarrow (1-\alpha) Q(s_t, a_t) + \alpha\left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a)\right]$$

In this paper α as learning rate stands for "condition and history of each suppliers, whether in supplying quality or meeting commitment of administrative and legal commitment" and γ as a time value parameters stand for "conditions and situation which a company supplier of transportation service provide from the point of capability of suppling and responsibility toward needs". The amount of both α and γ is considered to be positive number smaller than one.

The estimate for $Q(s_t, a_t)$,the value of the state-action pair at time t is updated using the best estimated value of the next state in the follow we show the proposed algorithm based on Q-learning for solving TSO problem

4. Modeling based on agents

reinforcement learning mechanism discussion always revolves around an interaction between agent and surrounding environment (fig.2). In this model, agent interacts with each of suppliers separately, making decision on outsourcing and insourcing transportation services by evaluating each supplier,drawing on both criteria of transporting cost and delivery time. It also makes decision by reinforcement learning mechanism. In this problem objective function is:

$$TC = \alpha T + \beta C \quad (3)$$

TC is total costs for every supplier, T stands for delivery time index and C transportation cost, α and β have a constant coefficient.. therefore minimizing this function is desired.

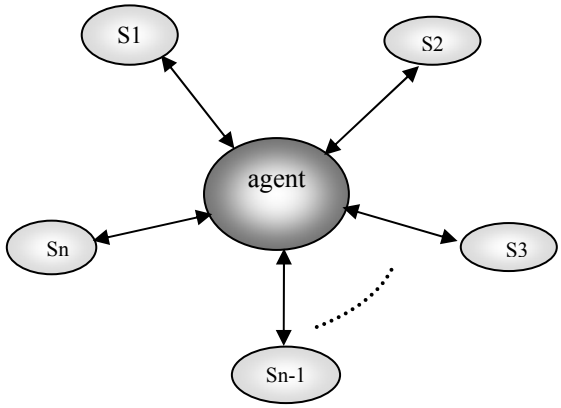

Fig.2. Agent-environment interaction in proposed model.

4.1. RL modelling of outsourcing problem in transportation system

In this section, characteristics of reinforcement learning model in outsourcing problem of transportation (TSO) has been defined. Key parameters of model including variable state, reward function, value function and system policy have been specified. Elements of decision- making mechanism based on RL (RLDM) have been explained. Q-learning algorithm has been applied for solving the problem.

4.1.1. State variable

As mentioned, the agent makes decision based on a function of signals received from environment which is "called environment state" and is markov in nature.

Even when such state is non markov, it is suitable to assume that it has great similarities to Markovies property in reinforcement learning [1]. Since the next state of system uses X+Y pattern directly which is based on its previous state, then markovies property is formed. In this paper environment state, available time for delivering good, in other words the spent time for

transporting or considered amount of cost will be i period. In TSO problem is shown by following vector:

Here s(i) stands for vector of system state related to each of suppliers in i period and $S_c(i)$, $S_t(i)$ are states related to transportation cost transporting and delivery time of each supplier separately.

4.1.2. policy

policy in reinforcement learning algorithm is a mapping of perceived states to actions which must have been done in that state[10]. In other words policy is action that agent does. In TSO problem policy is specifying amount of y for estimation of transporting cost and delivery time of each supplier. Hence policy is a vector as follow:

$$Y_S(i) = \left\{Y_{S,C}(i)\ ,\ Y_{S,T}(i)\right\} \tag{4}$$

Notice that actions done by agent are amount of time delivery and cost of transporting therefore
A vector in s state can be shown as follow:

$$A_S(i) = \left\{A_{S,C}(i),\ A_{S,T}(i)\right\}$$

That:

$$A_{S,C}(i) = X + Y_{S,C}(i) \tag{5}$$

$$A_{S,T}(i) = X + Y_{S,T}(i)$$

It needs to be mentioned that y amount can be positive, negative or zero

4.1.3. Reward function

A reward function in i period which is shown as $r(i)$ is, a mapping of perceived state (or State- action pair) Which record an action as a unified number called reward[12]. Such number is an indication of desirability of system state (or performed action in that state). reinforcement learner agent aims to maximise total reward in long term. Hence reward function must be a definition of goals in reinforcement learning. The main purpose in this paper is to estimate costs of transporting and delivery time connected with intended transportation suppliers in order to decide on both outsourcing and insourcing of transportation services. So reward function must be defined in a way so as to take this purpose into consideration. Reward function represents the next system state, that is by action of agent, with respect to two parameters of transporting cost and delivery time. It can be defined as follow:

$$s(i) + a(i) \tag{6}$$

4.1. 4. value function

The value of a state is total rewards that an agent can expect to accumulate at the beginning of this state. While rewards indicate the suitability of temporary state of environment, values indicate the time length of states which is desirable [10,11].
The action-value function of taking an action a in a state s is defined as:

$$Q(s,a) = E\left\{R_t \middle| s_t = s, a_t = a\right\} = E\left\{\sum_{k=o}^{\infty} \gamma^k r_{i+k+1} \middle| s_t = s, a_t = a\right\} \tag{7}$$

this is the expected return under policy π, starting from taking action ain state S.
As shown, in order to specify the value of an action in special state, a range of rewards must be calculated in the system. Since reward continuity depends on system state which in turn is a function of environment conditions (and is unknown),. The extent of relation related action value function can be calculated. Calculating the value function is not as simple as reward function, because the next system states are uncertain. Therefore, the value of Q(s, a) must be estimated. Estimating Q(s, a) for each state-action pair, the best action in each state will be specified and optimal policy can be derived [1]. In our model, Q-function is estimated by a mechanism based on Q-learning which has already been explained.

4.2. the used rule in proposed algorithm

The pattern applied for determining the amount of delivery time is X+Y pattern (see Fig. 3)which is the rule used in reinforcement learning algorithm [6]. According to this rule, if in an ideal state, delivery time or costs is X units, during different time steps such amount can be increased or decreased. Part Y can be positive, negative or zero. In fact , this learning algorithm aims to obtain policies for determining Y amount during different states of system so as to cause a desirable increase.

Fig. 3. X+Y rule

4.3. proposed algorithm for solving RL decision making model

In previous sections, the problem was modeled based on reinforcement learning . Now, we will apply Q-learning algorithm and use our recommended law to solve the problem . In proposed algorithm , the amount of value function must be learnt iteratively (fig.4). At the end of learning process, the optimized policy is adopted based on previous learning as well as the values of Q function.then, the best action available that is, the action

with the least valuesQ(s,a) is selected. Then based on that amount, delivery time and transporting cost is estimated. In proposed algorithm, learning model has been simulated to specific numbers and during the simulation ,amount of operation value in any state is updated. Following is the relation of updating the values [13]:

$$Q(s,a)=Q(s,a)+\alpha\left[r+\gamma\max_{a'}Q(s',a')-Q(s,a)\right] \quad (8)$$

In this pattern r is reward function which is an estimation of current state as well as the amount of a performance a.

I- set the initial learning conditions include:
Iteration=0, i=0, Q(s,a)=0 for all s,a

II- While Iteration ≤ MAX Epoach

select a random T and C from states space

While i ≤ n

(1) With probability p_e select an action vector with maximum Q(s,a), otherwise take a random action from action spase

(2) Calculate r(i+1)

(3) If the next state is s', update Q(s,a) using:

$$Q(s,a)=Q(s,a)+\alpha\left[r(t+1)+\gamma\max_{a'}Q(s',a')-Q(s,a)\right]$$

(4) Do action a and update the current state vector

(5) Increase the p_e according to the scheme(i.e. linearly)

(6) i=i+1

i=0

Epoach=Epoach+1

By a greedy search on the Q(s,a), best action in each state is distinguished

Fig. 4. Proposed algorithm based on Q-learning for solving TSO problem

5. execution of proposed model

In this section, the proposed algorithm has been implemented by applying Q-learning mechanism and using simulation(see table 1) process as well as by considering assumed carecteristics concerned with time and costs of transporting given by four suppliers (including the main company itself) (fig. 5). The result obtained following the performance of algorithm as well as the values determined for parameters and their respective outputs have been shown(table 2,3).Since decision making is based on two parameters of transportstion time and cost, it is necessary to evaluate suppliers outsourcing and insourcing using a combined linear programming of these two parameters. In order to select each supplier as well as for purpose of outsourcing them ,the collected data having different dimensions(cost and time) are normailized, yielding the total cost fucnction (table 4). For each state, the minimium value based on which the suppliers are chosen is detremnined. Fig. 6.

Table. 1
Characteristic of learning model

variable	S.1	S.2	S.3	S.4
α	0.05	0.1	0.07	0.17
γ	0.7	0.5	1	0.3
Y_i (cost)	(time) Y_i	S_C	S_T	
[-100 , 100]	[-2 , 2]	[300 , 600]	[8 , 14]	

Table. 2
Output of proposed algorithm (time)

state	S.1	S.2	S.3	S.4	Min	Selection
8	10	8	9	10	8	S.2
9	11	8	8	11	8	S.2 , S.3
10	11	8	9	9	8	S.2
11	9	9	9	9	9	S.1 , S.2 , S.3 , S.4
12	14	14	14	14	14	S.1 , S.2 , S.3 , S.4
13	11	15	11	12	11	S.1 , S.3
14	13	13	13	14	13	S.1 , S.2 , S.3

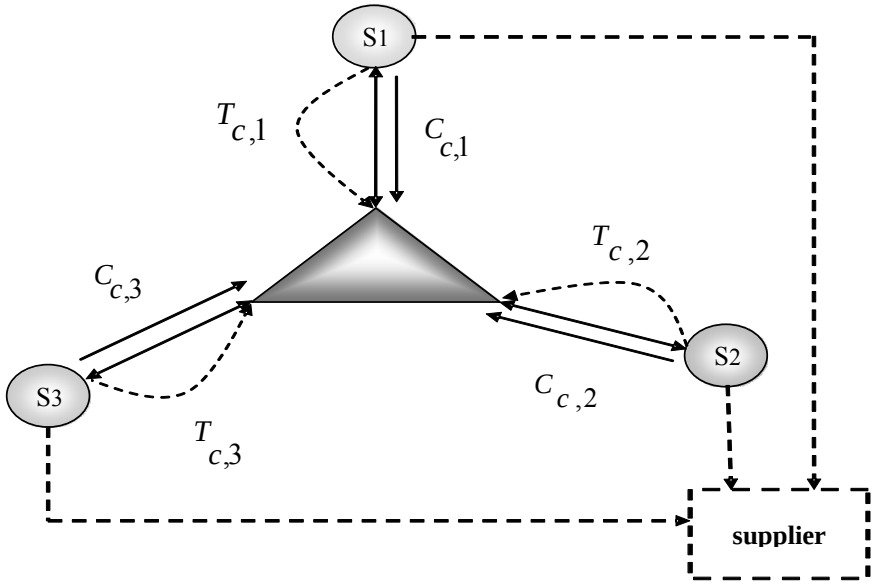

fig. 5. Center factory interaction whit each of supplier

Table. 3
Output of proposed algorithm (cost)

S.1	S.2	S.3	S.4	Min	Selection
350	400	450	350	350	S.1 , S.4
450	450	400	300	300	S.4
400	550	500	500	500	S.1
350	350	350	350	350	S.1 , S.2 , S.3 , S.4
450	450	450	450	450	S.1 , S.2 , S.3 , S.4
500	500	550	550	500	S.1 , S.2
550	550	550	600	550	S.1 , S.2 , S.3

Table 4
Decision making about supplier selection

S_{T-S}	S.1	S.2	S.3	S.4	Min	Selection
۸-۶۰۰	0.57	0.53	0.54	0.60	0.53	S.2
۹-۵۵۰	0.56	0.50	0.55	0.59	0.50	S.2
۱۰-۵۰۰	0.50	0.44	0.46	0.45	0.44	S.2
۱۱-۴۵۰	0.43	0.43	0.43	0.42	0.42	S.4
۱۲-۴۰۰	0.57	0.68	0.67	0.57	0.57	S.4 , S.1
۱۳-۳۵۰	0.53	0.65	0.51	0.46	0.46	S.4
۱۴-۳۰۰	0.52	0.55	0.59	0.54	0.52	S.1
۸-۵۵۰	0.54	0.50	0.54	0.57	0.50	S.2
۹-۵۵۰	0.50	0.44	0.44	0.50	0.44	S.2 ,S.3
۱۰-۴۵۰	0.47	0.41	0.43	0.42	0.41	S.2
۱۱-۴۰۰	0.46	0.55	0.51	0.51	0.46	S.1
۱۲-۳۵۰	0.60	0.62	0.59	0.54	0.54	S.4
۹-۴۵۰	0.47	0.41	0.41	0.47	0.41	S.2 ,S.3
۱۰-۴۰۰	0.50	0.53	0.51	0.51	0.52	S.1
۸-۵۰۰	0.51	0.44	0.46	0.48	0.44	S.2

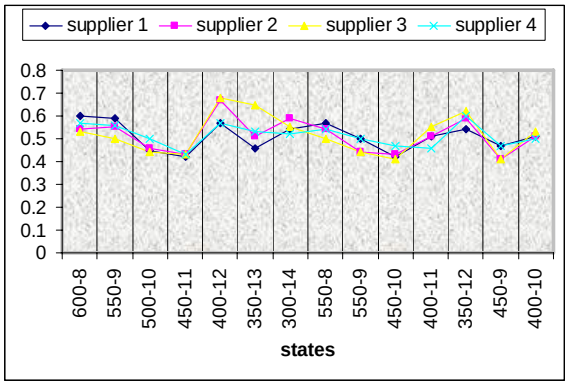

Fig. 6. utility of each supplier in diferent states.

Conclusion

This paper presents the way decision making concerned with both outsourcing and insourcing of transport system by applying intelligent agent approach. To do so, an intelligent model based on interaction of agent with environment was presented. At first, the problem was modeled in an intensive form. The details and characteristics of such model was defined and then Q-learning was utilized to solve.

This method (which is a subdivision of artificial intelligent), has been applied due to being simply understood , higher speed, adaptability of the problem based on intelligent agent learning model, the ability of agent to understand environment,and its interaction especially in an unknown environment. To sum up it can be total that presented model is mechanism of decision making based on agent. Due to its capability of environmental adaptation. Ability to relate with environment, and self reformation based on the learnt information it can be apply has a learn model concerned with decision making problems. Due to innovational aspect of this paper it is natural that it can be faced with imperfection and deficiency at early steps. Hence many investigational fields for expansion and practicality of its ability can be designed an performed. At the end only sum points will be mentioned as follow:

1. Weighted quality criteria and their normalizion can be included in values table and hence in this model. Although some of important qualitative criteria have been applied directly as parameters of Q-learning algorithm.
2. This paper considers one agent while we can have a multi- agent system.
3. This study assumes the values of α and γ (time and cost criteria)for each supplier to be exclusively independent,yet they may be dependent. which are independent. In fact these can be related with each other. Although independence problem of α, γ coefficients have been partly solved by normalization and sensitivity analysis of results that in future researches a linear combination can be considered.

References

[1] S.K. Chaharsooghi, S.H. Zegordi, J. Heydari, reinforcement learning model for supply chain ordering management: An application to the Beer Game, 949-959, 2008.

[2] Coman, B. Ronen, Production outsourcing: a linear programming model for the Theory-Of-Constraints. International Journal of Production Research, 38, 1631-1639, 2000.

[3] J.M. Downey, Risks of outsourcing applying risk management techniques to staffing methods. Facilities, 13, 38-44, 1995.

[4] S. Ishii, W. Yoshida, J. Yoshimoto, Control of exploitation-exploration meta-parameter in reinforcement learning, Neural Networks, 665-687, 2002.

[5] S. S. Keerthi, B. Ravindran, Rinforcement Learning, In Fiesler, E., and Beal, R., editors, handbook of neural computation. Oxford Univercity press, USA, 1997.

[6] S.O. kimbrough, D.J. wu, F. zhong, Computers play the beer game: can artifical agents manage supply chains? Decision support systems 33, 323-333, 2002.

[7] C. Lonsdale, Effectively managing vertical supply relationships: a risk management model for outsourcing. International Journal of Supply Chain Management, 4, 176-183, 1999.

[8] E. Martinez, Solving batch process scheduling/planning tasks using reinforcement learning, computers and chemical Engineering supplement, 23, S527-S530, 1999.

[9] J. M. Padillo, M. Diaby, A multiple-criteria decision methodology for the make-or-buy problem. International Journal of Production Research, 37, 3203-3229, 1999.

[10] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, A Bradford Book, The MIT Press Cambridge, Massachusetts London, England, 2005.

[11] R. S. Sutton, A. G. Barto, learning to predict by the method of temporal differences.machine learning, 487-206, 1998.

[12] C. J. C. H. Watkins, Learning from deleyed rewards, phD thesis, Combridge Univercity, Combridge England, 1989.

[13] C. J. C. H. Watkins, P. Dayan, Q-learning. Machine Learning, 279-292, 1992.