

Strategic Planning Formulation by using Reinforcement Learning

¹Naghi, R. Seyedaghaee, ²S.K. Amirgholipour, ³Hamid Alinejad-Rokny
and ¹Fatemeh Rouhinezhad

¹Department of Computer Engineering, Aliabad Katoul Branch, Islamic Azad University,
Aliabad Katoul, Iran

²Department of Computer, Ramsar Branch, Islamic Azad University, Ramsar, Iran

³Department of Computer Engineering, Science and Research Branch, Islamic Azad University,
Tehran, Iran

Abstract: This study considers the similarities between strategic planning and reinforcement learning, then formulates the strategic planning to a reinforcement learning problem and solves it by using reinforcement learning methods. Increasing the environmental cataclysm and variations in today's world enforce all organizations such as universities, to have a strategic thought to increasing the effect answering time to their environmental needs, resource wasting reduction, leading the organization to a clear and executable direction and problem determination in On the other hand internal and external views., computers have a key role in today's life. We try to relate between these two subjects. We use recommended method to computerize the strategic planning design for Noshahr-Chaloo Islamic Azad Universi of the university placed in compositional ty of Iran. The results show that the general strategic positionstrategies. Also the results present using intelligent methods to computerize the strategic planning have a very good future, although it is on beginning of its long path.

Key words: Formulation, reinforcement learning, strategic planning

INTRODUCTION

Nowadays, because of environmental cataclysm, all organizations especially universities require a strategic thought for an effective answer to its environmental needs. Developed and successful organizations, implement their strategic thought in a strategic planning structure. In strategic thought and planning, to prevent resource wasting and guide the organization to a clear and executable direction and determine the problems in internal and external views, that lead us to discovery and formulate the internal weakness and strength points and external opportunities and threats be considered. It can be used for placing an organization in a better and higher position.

In another word, organizations can impress the environment by using strategic planning and don't let environmental shocks impress their performance, instead of they only react to the environmental events and motives and wait for them. Strategic planning is a powerful management tool which designed to help the organizations to adjust themselves by environment predicted variations in a competitive manner. Strategic

planning process, propose a scheme, analysis of the organization and its related environment and describe the present condition of the organization and company and determine the effective key agents on the organization or company success. Like other humanism discussions, different research and study resources propose different methods, views and opinions on a social and humanity phenomenon. We can't propose a unique opinion on a special phenomenon because of variant opinions in humanism discussions. By considering this fact, we illustrate some of different definitions extracted from strategic planning.

Strategic planning is an organized and regular attempt for an organic decision making and doing vital actions to form organization actions in a legal manner (Bryson, 2004). Strategic planning is a process to attain long-term and vital purpose of organization in a competitive environment. So, difference between long-term planning and Strategic planning is only in vital and competitive environment words (Fred, 2004).

In continue, we review some done works in Strategic planning field. Hokkaido University tried to change its weaknesses points to strength points in new curriculum.

Because of this, the university tried to attend to intermediate-course structure curriculum of low levels, and attend to flexibility in work in a manner that is appropriate to competitive, variable and challenging environments in higher levels. All these mentioned cases and other things that are not mentioned cause to design strategic planning of Hokkaido University in 21st century (Ogasawara, 2002). Dyson emphasizes that SWOT tool is a method to design the required strategy of organizations especially for universities. Finally, researcher measured opportunities, threats, weaknesses and strengths points of university by considering defined variables on interval [1-5] and extracted appropriate strategies for university by comparing these cases, and proposed them to university managements in an applicable manner (Dyson, 2004).

Florida university libraries did some activities related to the strategic planning. This act flowed the strategic thought among managers, and cause of create a lot of interest among staff to achieve data and related information to quality improvement and proposed services to customers and considering their desires. Results of these actions were, improving the relations with customers, and moving toward strategic thought and methods that cause a good decision making to assign available resources (Shorb and Driscoll, 2004). Bowling Green State University met to new challenges that they were cause of library customer's dissatisfaction. So they try to formulate and implement the strategic planning of university to appoint the role of university libraries in high education. A lot of attentions focus on opinions of customers and users of university services by interview with special groups to formulate the university strategic planning (Haricombe and Boettcher, 2004).

Shoemaker and Fischar (2011) described an evidence-informed strategic planning process and framework used by a Magnet-recognized public health system in California. They studied the structure created within nursing for collaborative strategic planning and decision making (Lorie and Fischer, 2011). James *et al.* (2010) reported the findings of an industry-driven research project which aims to identify the available opportunities and the way forward for developing the construction industry of Hong Kong in a sustainable manner. Based on their collective views of experts from the industry, a set of strategic directions was proposed under a construction industry development framework (James *et al.*, 2010). Zhaoguang *et al.* (2010) proposed a new approach called the Integrated Resource Strategic Planning (IRSP). They introduced the necessity and possibility of using their new approach, presented a framework on how to use the approach, and justified the effectiveness of their approach against the traditional power planning approach, with a case study in China (Zhaoguang *et al.*, 2010).

Nowadays, using computers for doing actions as secondary instruments is inevitable. Usually learning in machine is a computer program that provides the learning

ability for computers, and in intelligent system means that an agent can use its experiments. In other word, this system can learn and improve its performance. In learning problems follow things should be considered: What action should be done? How gain the experiment? How an assessment done? Learning is synonym of intelligence, because we can say that an existent is intelligent if it can learn about its experiments and improve its performance. Reinforcement learning is the problem faced by an agent that learns behavior through trial-and-error interactions with a dynamic environment. An agent doesn't have answer of problem, but by doing each action, get a reward if it done correctly, otherwise get a punishment (Littman *et al.*, 1996).

As it mentioned, we are encountered with an environment that it changes perpetual, in strategic planning of organizations. In strategic planning running stage, an organization assesses its present position and does some actions to improve its position. Then organization observes the results of these actions, and modifies the direction of the organization to achieve the goals by using previous experiments and basis of new position, if modifying be necessary. We are encounter with the environment which it is unknown for an agent in reinforcement learning, too. The agent recognizes its position in each time and does some actions. Doing this action cause change the position of agent. The environment gives a reward against the action that done by agent and by considering its position. This reward can be positive or negative. The agent remembers the traversed paths and uses them for choosing actions in the future (Mitchell, 1997).

In this study, we consider the similarities between strategic planning and reinforcement learning to computerize the strategic planning. We convert the strategic planning to a reinforcement learning problem, and solve it by using reinforcement learning methods. The implement results are presented on Noshahr-Chalooos Islamic Azad University data. The results show the computerized method estimated the optimum cost and path correctly-very close to the expert's perspective.

Reinforcement learning: Reinforcement learning is a word that extracted from words collection related to animals learning by Minsky and used in machine learning field. Reinforcement learning is assigned to a class of algorithms and learning and training functions. These learner systems learn a relational mapping by maximizing a grade of its action environmental reward. In this relation X is the set of states and A is the set of possible actions (Littman *et al.*, 1996).

If $x \in X$ is given to us in t moment, learner system tries to find $a \in A$ - the most appropriate action in learner system view, and environment returns a numeric grade from (x, a) function immediately (This determines how much action a is near to $P(x)$). In the standard reinforcement

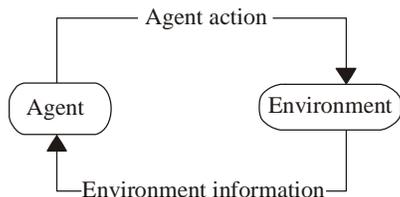


Fig. 1: A standard reinforcement learning model

learning model, an agent is connected to its environment via perceptions and actions. On each step, the agent receives *ith* input-some indication of the environment current state *s*; then the agent chooses an action *a*, to generate output. The action changes the state of the environment, and the value of this state transition is communicated to the agent through a scalar reinforcement signal, *r*. The agent chooses the actions by doing this policy. Figure 1 illustrates standard reinforcement learning.

A simple model of reinforcement learning is consisting of following parameters:

- The environment is a set of *S* possible states.
- On each time *t*, the agent can do one of *A* possible actions.
- The agent receives reward *r* for a set of actions that it does. This reward can be positive or negative (punishment).
- The agent moves in the environment and remembers the related states and rewards.
- The agent's behavior should choose actions that tend to maximize the reward function.
- $R(t)$ is the sum of rewards that an agent collected in *t* lapsed time (Littman *et al.*, 1996).

Strategic planning and reinforcement learning:

Formulate the strategic planning to a reinforcement learning problem: To convert the strategic planning to reinforcement learning problem and then solve it, we should specify the mentioned parameters in previous section for strategic planning problem. The first parameter is the number of environment states. By considering Table 1, each place of these nine places divided to 81 slots and this table modify to 27*27 slots. The place (27*27) is the best position for an organization and all organizations assign their goals to reach this place finally. The total number of the environment states is 729. On each time the organization (agent) can be placed in one of these places and chooses appropriate actions by

considering its position and the actions that it can do to reach its final goal finally, with lowest cost and time. This final goal is the same organization vision.

Second parameter is the number and kind of actions that an organization (agent) can do in the different states. Number of actions that we can specify for the organization is 40, and is numbered from 1 to 40. These numbers can be increased or decreased. Each action has the follow features:

- The length of each action can be one of numbers 1, 2 or 3
- The cost of each action can be one of numbers 1, 2, 3, 4 or 5
- The effect of each action on each internal and external environment sections can be one of values between -5 (or high negative effect) to +5 (or high positive effect)

The environment against each done action of organization (agent), return a reinforcement value. This value can be positive (reward) or negative (punishment) based on improving the organization position. Assigning an appropriate reinforcement function for an agent that gives it the reinforcement value after doing each action, is the key point in the every reinforcement learning problem. The reinforcement function for strategic planning is defined follow:

- If after selected done action, the agent reaches to the final state (27, 27), $R=100$, *R* is the environment feedback (Reinforcement signal).
- If after selected done action, both of internal and external sections have been improved $R = 2*(9/(L*C))$, *L* and *C* are the length and cost of the selected action, respectively.
- If after selected done action, improvement observed only in one section, $R = 9/(L*C)$.
- If after selected done action, no improvement observed or the position of organization be worst $R = -0.5*(9/(L*C))$.
- If after selected done action, the agent leaved the environment $R = -50$.

Solve the strategic planning problem: One of the simplest model-free methods to solve the reinforcement learning problem is *Q*-learning method. In *Q*-learning method, the system states and actions should be determined. This act is done in previous section. The agent chooses an action in each place of environment

Table 1: Different places of nine strategies

Internal environment assessment		
Competitive strategies	Competitive-aggressive strategies	Aggressive strategies
Defensive-competitive strategies	Compositional strategies	Conservative-aggressive strategies
Defensive strategies	Defensive-conservative strategies	Conservative strategies
External environment assessment		

randomly, and does it. After doing each action, the agent receives a reinforcement value based on introduced reinforcement function. The updating formula of Q -value in each state is computed basis on Eq. (1):

$$Q(S, \alpha) = Q(S, \alpha) + \alpha [r + \gamma \max_b Q(S_n, b) - Q(S, \alpha)]$$

α and γ are learning rate and discount factor, respectively. Algorithm (1) shows all steps of Q -learning algorithm, that we used it.

Algorithm 1: Steps of Q -learning algorithm:

- Set $t = 0$, and consider s -a random state for start position.
- Repeat the follow actions, for a specified time.
- Choose action $a \hat{I} A(x)$ randomly and apply it for system. Suppose that new sate is s_n .
- Update Q -value by using Eq. (1).
- Set $s = s_n$.
- The algorithm will finished, if the agent reaches to its goal or condition of the problem be reversal, otherwise go to step 2.

The Q -table will be complete, by considering mentioned algorithm. After this work, the present position of the organization should be specified and based on this, some action will select to maximize sum of the rewards that the agent receives from present position to final goal.

RESULTS AND CONSEQUENCES

In this section we use mentioned method in previous section on Noshahr-Chaloos Islamic Azad University’s data. This data is gathered in 2006. The program assumptions for implement are mentioned follow:

- University environment is consist of six internal sections; training, researching, student, official and financial, constructional, cultural and four external sections; political, economical, social, cultural.
- We can define utmost 10 evaluating indexes for each section.
- Each index has a coefficient between [0 1] that shows its significance.
- Each index has a weight between [1 5] that shows the efficiency of organization present strategy for this index.
- The university can do utmost 40 actions. The actions of number [1 10] is related to defensive strategy, the actions of number [11 20] is related to conservative strategy, the actions of number [21 30] is related to competitive strategy and the actions of number [31 40] is related to aggressive strategy.
- Every action has a length in interval [1 3] and a cost in interval [-5 +5].

- Effect of doing every action on 10 sections is specified (one 40*10 matrix).

The inputs of program are:

- The inputs of each 10 sections are: indexes of each section, coefficient and weight of each index.
- One array 40*1 that shows the length of each action.
- One array 40*1 that shows the cost of each action.
- One matrix 40*10 that shows the effect of each action on 10 sections.

After receiving the inputs of the program, necessary processing is done on them and the outputs are:

- The present position of the organization.
- The optimum path to move from present position of the organization toward its vision.
- The proposed actions for moving over optimum path.

After receiving the coefficient and weight of each index for all 10 sections, the general strategic position of university computes by algorithm (2).

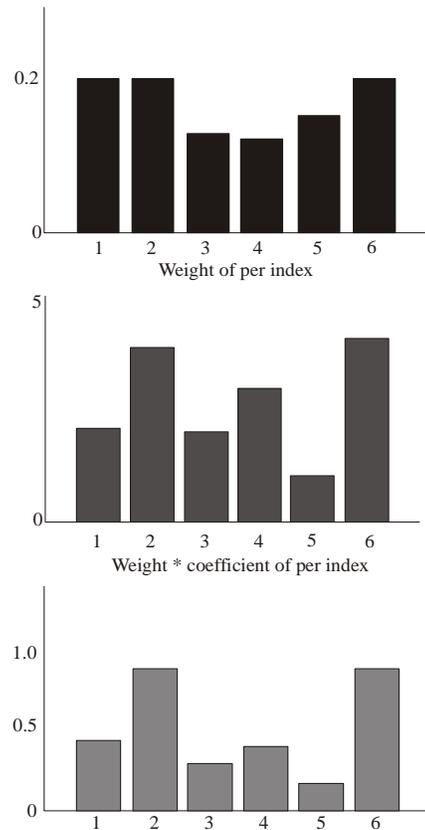


Fig. 2: Condition of external cultural section; Number of indexes are six (First row shows the coefficient of each index, 2nd row shows the weight of each index and 3rd row shows the total value of each index)

Algorithm 2: Situation determination algorithm of each section and whole organization in SWOT matrix:

- Condition of each 10 sections is computed this manner:
 - Determine the result of multiplying coefficient * weight for each index.
 - Condition of each section is equal with sum of multiplying coefficient * weight for each index, the result of it, is divided to the total numbers of section indexes.
- Whole condition of the internal environment achieves by computing the average of 6 internal environment sections assessment.
- Whole condition of the external environment achieves by computing the average of four external environment sections assessment.
- The general position of university will determine in SWOT matrix by specifying the general position of internal and external environment.

Figure 2 illustrates the number of indexes, weight and coefficient of each index and total value of each index for external cultural section. There are the pictures like this, for each of ten university sections.

Figure 3a shows the general condition of six internal environment sections and four external environment sections. After specifying the general condition of 10 internal and external sections, the general position of university specify in SWOT matrix. This operation shows in Fig. 3b. Table 2 shows the general value of each 10 internal and external environment sections of university, also, it shows the general value of external and internal environment. After specifying the general strategic position of university in SWOT matrix, the program starts to learn by using algorithm (1), and proposes some actions to reach its vision on lowest cost and time (place (27 , 27)), based on present position of university, its own previous experiences and specified parameters. Figure 4(b) shows the proposed optimum path for Noshahr-Chaloos Islamic Azad University on its present position. Also, Fig. 4a shows the proposed actions to reach the final goal of university from its present position.

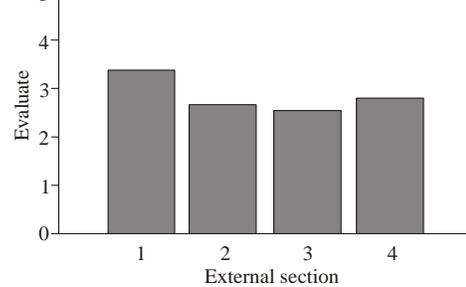
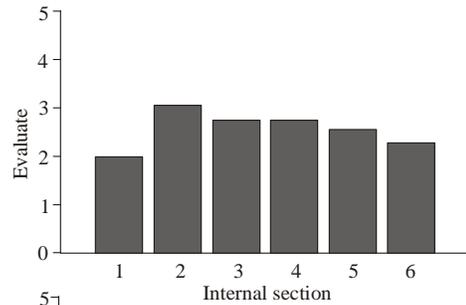
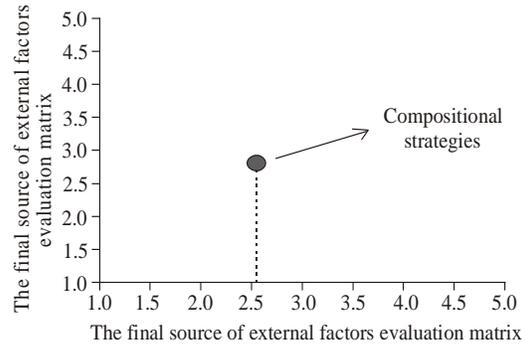


Fig. 3: a: General strategic position of the university in SWOT matrix. b: 1st row shows the general condition of six internal environment sections and bottom row shows the general condition of four external environment sections

As it is shown in Fig. 4a, because of present position of university placed in the place of compositional strategies, thereupon, it can do each of 40 actions

Table 2: The general condition of 10 university sections and general value of internal and external environment

External environment		Internal environment	
The cost of the expert,s perspective	49	Cultural	2.76
The cost of optimum	54	Political	2.74
The cost of optimum path	2.82	Economical	2.49
The total value of the external environment		Social	3.31
The total value of the internal environment		Constructional	2.56
		Cultural	2.33
		Official and financial	2.7
		Student	2.67
		Researching	3
		Training	1.93

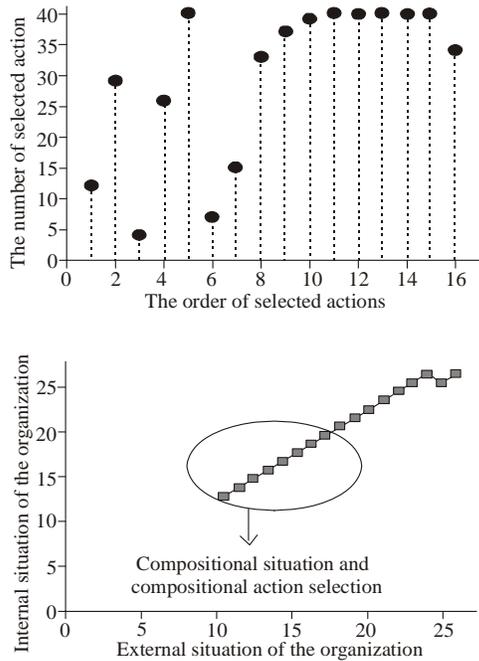


Fig. 4: a: Shows the proposed actions for present position of the university. b: Shows the proposed optimum path for Noshahr-Chaloos Islamic Azad University

based on its condition. But, the position of university improved after doing some actions and placed in aggressive strategies. After this, the university should do only aggressive actions. As it is mentioned before, the numbers of aggressive actions is in interval [31 40]. Figure 4a clearly shows that the program only proposes actions with the number of greater than 30, after entering university in the place of aggressive strategies. Since, the number of states in each place of Table 1 are 81 and the numbers of actions that can select on them are 10, therefore, the actions iteration selection is not impossible, by considering their appropriate length and cost.

The Summary of present position of Noshahr-Chaloos Islamic Azad University is described follow:

The weaknesses and strengths points of studied university surveyed based on sextet internal indexes: researching, student, official and financial, cultural and constructional and the following results have been obtained:

- In training section, five indexes was evaluated, the university has weaknesses in four indexes and assessed mean in only one index.
- In researching section, five indexes was evaluated, the university has weaknesses in two indexes and has strengths in two indexes and assessed mean in one index.

- In student section, four indexes was evaluated, the university has weaknesses in two indexes, has strength in one index and assessed mean in one index.
- In official and financial section, nine indexes was evaluated, the university has weaknesses in five indexes, has strengths in three indexes and assessed mean in one index.
- In cultural section, four indexes was evaluated, the university has weaknesses in three indexes and has strength in one index.
- In constructional section, three indexes was evaluated, the university has weaknesses in two indexes and has strength in one index.

In general, six internal section of university was evaluated. The results show that training section was assessed weakness and researching, student, official and financial, cultural and constructional sections were assessed mean.

The opportunities and threats (due to environmental agents) of this university surveyed by considering four external sections-social, economical, political and cultural and the following results were obtained:

- In social section, six indexes were evaluated, the university has strength in one index and assessed mean in five indexes.
- In economical section, eight indexes were evaluated, all of indexes assessed mean.
- In political section, five indexes was evaluated, the university has weaknesses in two indexes and has strengths in three indexes.
- In cultural section, seven indexes were evaluated, all of indexes assessed mean.

In general, four external sections of university were evaluated. The university is assessed mean in all indexes. The achieved grad from evaluating weaknesses and strengths points, opportunities and threats show that we should use compositional strategies with higher emphasis on defensive strategies to change the position of university to a desirable position.

Since, the case study university is placed in compositional strategies position, so the program first proposed compositional actions and after that the university went out from this place, it proposed aggression actions.

DISCUSSION AND CONCLUSION

The results of implementing strategic planning by using reinforcement learning on the data of case study university show, general strategic position of university is placed in compositional strategies. These results show, the

program found appropriate path to reach to university vision. Although using intelligent methods to computerize the strategic planning is on the beginning of its long path, but, results show that it has a very good future.

Summary of intelligent strategic planning benefits are:

- An organization can observe its present strategic conditions simply.
- Ability to expand the number of actions that an organization can do.
- Ability to expand the number of evaluating indexes for each section.
- Ability to increase or decrease the number of each nine strategy places.
- The ability of program generalization to another organization except university with a partial change.
- Staff costs and times reduction by using computers in strategic planning.
- The organization can observe its optimum path to reaching its vision.
- The organization can observe approximate cost to reaching its vision.
- All of measurements show in some diagrams and it is simple for anybody to understand it.

REFERENCES

- Bryson, M.J., 2004. Strategic Planning for Public and Nonprofit Organization. Jossey Bass., United kingdom, ISBN 0787967556.
- Dyson, G.R., 2004. Strategic development and SWOT analysis at the University of Warwick. *Eur. J. Oper. Res.*, 152: 631-640.
- Fred, R.D., 2004. Strategic Management: Cases. 10th Edn., Prentice Hall, ISBN-0131503472.
- Haricombe, L.J. and B.J. Boettcher, 2004. Using libQUAL data in strategic planning: Bowling green state university. *J. Library Admin.*, 40 (3/4):181-195
- James, M.W., S. Wong, N. Thomas and A.P.C. Chan, 2010. Strategic planning for the sustainable development of the construction industry in Hong Kong. *Habitat Int.*, 34(2): 256-263.
- Littman, M.L., L.P. Kaelbling and A.W. Moore, 1996. Reinforcement learning: A survey. *J. Artif. Intell. Res.*, 4: 237-285.
- Lorie, K.S. and B. Fischer, 2011. Creating a nursing strategic planning framework based on evidence. *Nurs. Clin. N. Am.*, 46(1): 11-25.
- Mitchell, T., 1997. Machine Learning. Machine Learning. McGraw-Hill Science/ Engineering/ Math, .
- Ogasawara, M., 2002. Strategic planning of the graduate and undergraduate education in a research university in Japan. *Higher Educ. Policy*, 15(1): 55-60.
- Shoemaker, L.K. and B. Fischer, 2011. Creating a nursing strategic planning framework based on evidence. *Nurs. Clin. North Am.*, 46(1): 11-25.
- Shorb, S.R. and L. Driscoll, 2004. Meets strategic planning at the University of Florida. *J. Lib. Admin.*, 40(3/4): 173-180.
- Zhaoguang, H., T. Xiandong, Y. Fan, M. Yang, Q. Wen, B. Shan and H. Xinyang, 2010. Integrated resource strategic planning: Case study of energy efficiency in the Chinese power sector. *Energ. Policy*, 38(11): 6391-6397.