Application Of Finite Stage Markov Decision Process In Policy Determination For Employees' Motivation

Ph. D, Associate professor, R. Minovski
University "Ss Cyril and Methodius", Faculty of Mechanical Engineering, Skopje, Republic of Macedonia
Graduate mathematician, teaching assistant, K. Mitkovska-Trendova
Military Academy "General Mihailo Apostolski", Skopje, Republic of Macedonia
Ph. D, Full professor, Jovanoski D.
University "Ss Cyril and Methodius", Faculty of Mechanical Engineering, Skopje, Republic of Macedonia

Abstract: *Enterprise restructuring is becoming very important in today's dynamic environment. In that direction systematic and efficient approach concerning the enterprise restructuring is more than important for gaining competitiveness of the enterprises. In this article, one methodology for enterprise restructuring (COMPASS) is briefly presented. It is an open methodology, meaning that it is giving the frames of the restructuring process and it allows implementation of various additional methods, tools, etc. We are continually trying to upgrade this methodology with sophisticated but easy to use methods that will improve its usability. One very important aspect of the restructuring is coping with the resistance to changing. Here, the utilization of Markov decision processes considering the policy determination for employees' motivation is presented. The focus will be on solution techniques for the finite stage problems.*

Key words: *Enterprise restructuring, COMPASS, Markov decision processes, employees' motivation, finite stage problems.*

1. Introduction

The problem of decision making is getting more and more actual. That is a result of the increased dynamics of the market, which is becoming its main feature. Not only valid, but also fast decision making on all levels of management is becoming imperative. In that sense, the need for enterprise restructuring has become very important issue, and so is the creation of a model for fast determination of the economical and technical capabilities of the enterprise and suggestions for the future development of the enterprise and obtaining its competitive capabilities.

The model for enterprise restructuring presented in this article is called COMPASS (COmpany's Management Purpose ASSistance). Its main intention is to offer aid to the enterprise management in direction of systematisation of the complex process of enterprise restructuring and locating method approaches that will help in the key points of the decision making. The basic idea of the model is to obtain a (sub)model of performance measurement, which will enable determination of the inconsistency of the importance and performance of all segments of the enterprise and on that basis to generate quantified alternative and then optimal actions for improvement of the situation.

The main aim of the COMPASS is to systematise the complex process of enterprise restructuring. The main benefit from its utilisation is to get the clear picture about the enterprise situation through utilisation of certain methods and tools. COMPASS tends to use methods and tools that are both simple and robust. This should accomplish both their implementation and should solve detected problems. Utilisation of those methods and tools forces analysis of the situation, which should bring to its better understanding. Implementation of the exact methods and tools is not the main point. COMPASS doesn't have intention to frame the process of enterprise restructuring. On the contrary. Since the methodology has huge target group of enterprises in order to aid their restructuring, our efforts at this stage are directed to obtain auxiliary methods, which will give more reliable basis for decision making in those points. So, COMPASS is free to accept any additional methods and models in concrete cases of implementation. Table 1 contains the phases of COMPASS.

The picture about the actual situation in the enterprise is described through the variables of the COMPASS-subKEs. SubKEs are presented in one matrix, called I/P (Importance/Performance) (the third phase of COMPASS).The output of this matrix is the list of Critical Elements-subKEs which have unbalance between their importance and performance. I/P matrices are genuinely gap analysis, presented in portfolio way, which improves the transparency of the analysis. For every Critical Element (CE) appropriate Success Factor (SF) is inducted. SFs are various kinds of actions which should lead to improved situation in the enterprise. At the initial development phase of COMPASS the generation of the SFs is done heuristically [6].

Our limited experience from the on-going verification of the methodology in few enterprises located these main problems: finance for realisation of the improvement actions and employees' obstruction for the changes. The practice in management of changes pinpoint that the best way to avoid the second obstacle is to be transparent in all phases and explain to the employees the aims of the restructuring. In that way they may become the alliances instead of opponents to the changes.

As we previously mentioned, the idea is to improve COMPASS with scientifically funded methods which are going to help the generation of more reliable SFs. In that direction, Markov decision processes are used here to support the policy determination. This article intents to give an introduction of their application in management of changes. Here we use a finite-stage Markov decision processes model in order to illustrate its application on our example. The objective is to find the optimal informing policy and to determine the utility function. There is no need for mathematical theory background for using this algorithm. It is easily applied using programs in Excel, MATLAB, LINDO/LINGO, CPLEX, …etc.

Table 1: Phases of the model for enterprise restructuring [6]

#	Content of the phases in the model	Some of the utilised method approaches
	Elucidation of the present situation of the enterprise in a measurable form from strategic importance point of view. The measurement of this issue is done through subKEs. AHP method is implemented (3).	AHP method Team work (Workshop)
	Explanation of the present situation of the enterprise in a measurable form from actual performance point of view. The measurement of this issue is done through subKEs. Specific methodology for auditing is created – SAudit (1), which is followed with the specially created procedure for evaluation.	SAudit SWOT Interview
	In order to determine the inconsistency of the subKEs from strategic and actual performance point of view I/P matrixes are employed. The result of this phase is the list of Critical Elements – subKEs which have unbalance between their importance and performance.	I/P matrixes (Gap analysis) Team work (Workshop)
	The beginning of the action generation is in the fourth phase. For every Critical Element (CE), appropriate Success Factor (SF) is induced. Examples for Success Factors are: shortening the cycle time, smaller lots, layout optimisation, more intensive education and training in some/all departments, standardisation, automation, ... So, SFs can be defined as various kinds of actions which should lead to improved situation in the enterprise. The generation of the SFs is done heuristically.	Structured knowledge about method approaches Forms for performance measures Matrixes KE -functional areas
	This phase should structure the bunch of SFs. The idea is to simulate the situation after the implementation of every possible set of SFs through the implementation of the particular procedure for scenarios generation and analysis.	Scenario technique Qualitative MICMAC method Simulation
	Selection of the optimal solution is determined in the sixth step. Previous phase gives the situation where certain scenario leads, concerning only subKEs. In this phase, the financial effect of every action is estimated.	Team work (Workshop) Pay-back method Costs/Gain diagram
	The seventh phase covers the implementation of the optimal action – no specific methods or procedures are foreseen for this phase in the present stage of development of the model.	

2. Application of Markov decision processes in policy determination for employees' motivation

Here is presented an application of dynamic programming in solving stochastic processes with a finite number of states, for decision making. Markov decision processes (MDPs) are a method for formulating and solving stochastic and dynamic decisions. MDPs provide a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of the decision maker. MDP model is a model for sequential decision making under uncertainty, which takes into account both the outcomes of current decisions and future decision making opportunities. This model encompasses a wide range of applications. MDP is a discrete time stochastic control process characterized by a set of states, and in each state there are several actions from which the decision maker must choose. For a state s and an action a, a state transition function determines the transition probabilities to the next state. The decision maker earns a reward (or cost) for each visited state. The states possess Markov property.

In other words, MDP is a tuple $(S, A, P_{\cdot}(\cdot,\cdot), R(\cdot))$, where S is the state space, A is the action space, $P_a(s,s') = P_r(s_{t+1} = s' | s_t = s, a_t = a)$ is the probability that action a in state s at time t will lead to state s' at time $t+1$, and $R(s)$ is the expected immediate reward (or cost) received after transition to state s' from state s, with transition probability $P_a(s,s')$. There is not an essential difference between rewards and costs, since maximizing rewards is equivalent to minimizing costs. But in practice costs are easier to determine. One of the characteristics in an MDP is the planning horizon of the process, which may be finite, infinite or of random length. Our task is to maximize some cumulative function of the rewards (or to minimize some cumulative function of the costs).

MDPs are an extension of Markov chains (or Markov chains are special case of MDPs), and the difference is the addition of actions (allowing choice) and rewards (giving motivation). MDP reduces to a Markov chain if the action to take were somehow fixed for each state. The solution to a MDP is called policy, which gives the action to take for a given state, regardless of prior history, that optimizes the performance of the system. This fixes the action for each state, and we get a Markov chain. The performance is measured by a utility function.

Our objective is to implement MDP model in COMPASS, in policy determination for employees' motivation, previously mentioned as one of the main located problems in the observed enterprises.

2. 1. Scope of the Markov decision problem – En enterprise example

We introduce a simple example, to use it as a basis for the model explanation. Besides its simplicity, the example is an equivalent to a number of important applications and the idea of the example can be adapted to represent this applications. In this way, basis is made for comparison of the actions used in order to improve the condition. This will contribute in optimal scenario selection.

We observe a continuously restructuring enterprise. At the beginning of every year, analysis is made in order to check its condition. Among that analysis are the tests for the transparency effect over the employees' motivation. According to the test results, we can classify the state of the employees' motivation for the new year as strong, average and weak. Let us assume that in years, it is noticed that we can assume that the employees' motivation in the actual year depends only on the employees' motivation in the past year (Markov process). Also we assume that we are in a condition to present the transition probabilities for one year period, from one motivation state to another, with the transition matrix:

$$\text{this year system's state} \begin{cases} 1 \\ 2 \\ 3 \end{cases} \overbrace{\begin{matrix} 1 & 2 & 3 \end{matrix}}^{\text{NEXT YEAR SYSTEM'S STATE}} \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix} = P^1$$

The correspondence between the motivation (strong, average and weak) and the states (1, 2 and 3), respectively, is:

System's state	Employees' motivation
1	strong
2	average
3	weak

It is assumed that because of the inertness of the uninformed employees, the transition probabilities in P^1 show that the motivation in the actual year can't be better from the one in the last year. For example, if employees' motivation this year is average (state 2), next year's motivation can stay average with probability of 0.5 or, to become weak (state 3), also with probability of 0.5.

The enterprise management can change the transition probabilities P^1 with taking other available directions of actions. The actions are means by which the decision maker interacts with the system, and when he observes that the system is in a certain state, he chooses an action from a certain action set, which may depend on the observed state. It is known that employees' informing about the enterprise restructuring, improves their motivation. If they don't do that, the transition probabilities will stay as it is given in P^1. But, if they take other directions of actions, as slightly informing or radically informing the employees, we get the following transition matrices P^2 and P^3, respectively:

$$P^2 = \left\|p_{ij}^2\right\| = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.25 & 0.55 & 0.2 \\ 0.1 & 0.55 & 0.35 \\ 0.05 & 0.3 & 0.65 \end{bmatrix} \end{array} \quad \text{and} \quad P^3 = \left\|p_{ij}^3\right\| = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix} \end{array}$$

We notice that in this example the action set consists of three actions, and they can be applied to every state.

It is possible to improve the employees' motivation in the new matrices P^2 and P^3, compared to the prior year. To put the decision problem in perspective, the management associates a return function (or a reward structure) with the transition from one state to another, and this function expresses the gain or the loss during one year period, depending on the states between which the transition is made. Since the management has the options for radically informing, slightly informing and not informing the employees at all, it is expected the gains and the losses to vary depending on the decision made by the management. Given the state of the system and the chosen action, an immediate reward (or cost) is earned (there is not an essential difference between rewards and costs, because maximizing rewards is equivalent to minimizing costs and both models can be found in MDPs sources). So, we get matrices R^1, R^2, and R^3 with the gains (immediate rewards) in hundreds of euros, associated with the matrices P^1, P^2, and P^3, respectively. We apply R^1 when there is no informing, R^2 when the informing is slight, and R^3 if the informing is radical.

$$r_{ij}^1\right\| = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & -1 \end{bmatrix} \end{array}, \quad R^2 = \left\|r_{ij}^2\right\| = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 6 & 5 & -1 \\ 7 & 4 & 0 \\ 6 & 3 & -2 \end{bmatrix} \end{array}, \quad R^3 = \left\|r_{ij}^3\right\| = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 5 & 4 & -2 \\ 8 & 3 & -1 \\ 5 & 4 & -3 \end{bmatrix} \end{array}.$$

Notice that the elements r_{ij}^2 and r_{ij}^3 of R^2 and R^3 take in account the costs for employees' informing, such as lost working hours while the informing has been done, materials, means and qualified staff for the training and the informing, ...,etc. For example, if the system was in state 1 and stayed in state 1 during the next year, its gain would be $r_{11}^2 = 6$ compared to $r_{11}^1 = 7$, when there was no informing. Thus, *R* gives the net reward after the cost for the informing is factored in.

So, the management has a decision problem for policy determination. First, they must know whether the informing will continue for a limited number of years or indefinitely. These situations are referred to as finite-stage and infinite-stage decision problems. The management uses the outcome of the analysis tests in both cases, to determine the best course of action that maximizes the expected revenue.

The management may also be interested in evaluating the expected revenue resulting from a prespecified course of action for a given state of the system. For example, informing should be done whenever the employees' motivation is weak (state 3). In this case, the decision process is said to be represented by a stationary policy, and we get Markov chain. We must notice that each stationary policy is associated with different transition and return matrices, constructed from the matrices P^1, P^2, and P^3, and R^1, R^2, and R^3.

We notice that after we evaluate all possible stationary policies of the decision problem which is equivalent to an exhaustive enumeration process and can be used only if the number of stationary policies is reasonably small, we can apply appropriate analysis for choosing the best policy, which we give in another article about the same example, but for the infinite-stage problem. However, this can be impractical even for the problems with limited size, because the number of policies can be too large. What we need is a method which determines the best policy systematically, without enumeration of all policies in advance.

2. 2. Finite-stage model

Suppose that the enterprise management plans to "retire" from the market in *N* years. So the management is interested in determining the optimal course of action for each year (three actions from the action set), that will return the highest expected revenue at the end of *N* years (finite planning horizon).

Let k=1, 2 and 3 represent the three courses of action (alternatives) available to the management. The matrices P^k and R^k represent the transition probabilities and reward function for alternative k, and are given previously. The management problem is expressed as a finite-stage dynamic programming (DP) model. In order to generalize the problem, we define

m (= 3) number of states at each stage (year)

$f_n(i)$ = optimal expected revenue of stages *n, n+1,* K, *N* given that i is the state of the system at the beginning of year *n*.

Let

$$v_i^k = \sum_{j=1}^{m} p_{ij}^k r_{ij}^k \quad , \qquad (1)$$

is the expected one-step transition revenue for the alternative *k* and the state *i* for that year.

The DP recursive equation can be written as

$$f_N(i) = \max_k \{v_i^k\}$$

$$f_n(i) = \max_k \left\{ v_i^k + \sum_{j=1}^{m} p_{ij}^k f_{n+1}(j) \right\}, \quad n = 1, 2, \ldots, N-1. \qquad 2)$$

We solve the management problem using the data summarized in the matrices P^1, P^2, P^3, R^1, R^2, and R^3, given a planning horizon of 3 years (*N = 3*). For the computations we can also use the programs we mentioned above using available interactive tables or make our own. For convenience we present the results in tables for the dynamic programming calculations, and discuss the results.

The optimal solution shows that the policy should be the same for all three years, and that is if the system is in state 1 (strong motivation), no informing should be done. But if the system is in state 2 (average motivation) and state 3 (weak motivation), the management should apply radical informing (*k* = 3)*. The total expected revenues for the three years are $f_1(1) = 10.891$, if the state of the system in year 1 is strong motivation, $f_1(2) = 9.15$, if it is average, and $f_1(3) = 5.289$, if it is weak.

Table 2: Results for stage 3

i	v_i^k			**Optimal solution**	
	$k=1$	$k=2$	$k=3$	$f_3(i)$	k^*
1	5.3	4.05	3.7	5.3	1
2	3	2.9	3.2	3.2	3
3	-1	-0.1	0.6	0.6	3

Table 3: Results for stage 2

i	$v_i^k + p_{i1}^k f_3(1) + p_{i2}^k f_3(2) + p_{i3}^k f_3(3)$			Optimal solution	
	$k=1$	$k=2$	$k=3$	$f_2(i)$	k^*
1	8.14	7.255	7.27	8.14	1
2	4.9	5.4	6.3	6.3	3
3	-0.4	1.515	2.71	2.71	3

Table 4: Results for stage 1

i	$v_i^k + p_{i1}^k f_2(1) + p_{i2}^k f_2(2) + p_{i3}^k f_2(3)$			Optimal solution	
	$k=1$	$k=2$	$k=3$	$f_1(i)$	k^*
1	10.891	10.092	10.193	10.891	1
2	7.505	8.1275	9.15	9.15	3
3	1.71	3.9585	5.289	5.289	3

If we compare this results with the results obtained in the infinite-stage case, we can see that the same policy is recommended by applying all the three solution techniques for the infinite horizon.

Dynamic programming solution from above is sometimes called value iteration approach, because from the recursive equation nature, the values of $f_n(i)$ are determined by iterations. It can be often referred to as backwards induction solution technique.

3. Conclusions

MDPs are an application of dynamic programming in solving stochastic processes with a finite number of states, for decision making. Dynamic programming is a very useful technique for making a sequence of interrelated decisions and is providing a great computational savings over other solution techniques, decomposing mathematical programming problems into smaller and computationally simpler subproblems. One way to recognize a situation that can be formulated as a dynamic programming problem is to notice the basic features of its basic structure. The problem can be devided into stages, with a policy decision required at each stage. Given the current state, an optimal policy for the remaining stages is independent of the policy decisions adopted in previous stages (principle of optimality). Any problem lacking the Markovian property cannot be formulated as a dynamic programming problem. A recursive relationship that identifies the optimal policy for stage *n*, given the optimal policy for stage *n+1* is available.

Finite-stage dynamic programming model is for decision problems with finite number of periods. But, many decision situations consist of very large number of periods or last indefinitely. That is why the infinite-stage dynamic programming model is developed. As *N* grows large, the corresponding optimal policies will converge to an optimal policy for the infinite-period problem. Although the method of successive approximations may not lead to an optimal policy for the infinite-stage problem after a few iterations, it never requires solving a system of equations. This is its advantage over the policy improvement and linear programming solution techniques, for its iterations can be performed simply and quickly. But it definitely obtains an optimal policy for an *n*-period problem after *n* iterations [7].

As the problem size increases, i. e. the state and/or the action space become larger, it becomes computationally very difficult to solve the Markov decision processes problem. For each action and state pair, we need a transition probability matrix and a reward function, which are enormous data requirements.

Finite-stage MDPs problems are more likely to be found in practice, where it is not usual to have infinite planning horizon, and there is a recursive nature to the prblem. This is the reason why we set out this model. On the other hand, it is easy to analyze and understand, and there is no need to know the mathematical theory for its implementation, since there is the opportunity for applicable computer programs.

References

1. Taha H. A. Operations Research: An Introduction. Maximillan Publishing Company, New York, Third edition 1982.
2. Kallenberg L. Markov decision processes. University of Leiden, fall 2007,
3. http://www.math.leidenuniv.nl/~lnmb/courses/Lecture notes-2007.pdf
4. Schaefer A. Markov decision processes. EWO Seminar, October 26, 2006. http://www.egon.cheme.cmu.edu/ewocp/docs/SchaeferMarkovdecisionprocesses.pdf
5. Jensen P. A. , Bard J. F. Operations Research, Models and Methods, http://www.me.utexas.edu/~jensen/ORMM/index.html
6. Minovski R. Creation of the Model for Overall Enterprise Restructuring, Faculty of Mechanical Engineering, Skopje (in Macedonian), 1996
7. Hillier F. S. Liebermann G. J. Introduction to Operations research. McGraw-Hill, New York, 8-th edition 2005