Deployment of Finite-stage Markov Decision Processes for Inventory Management Solutions in Enterprise Restructuring

Robert Minovski[1], Katerina Mitkovska-Trendova[2], Delco Jovanoski[3]

[1]*Faculty of Mechanical Engineering, Karpos II b.b., Skopje, Macedonia, mrobert@mf.edu.mk*

[2]*Military Academy, Vasko Karangelevski Str b.b., Skopje, Macedonia, trendov@yahoo.com*

[3]*Faculty of Mechanical Engineering, Karpos II b.b., Skopje, Macedonia, jov@mf.edu.mk*

Abstract:
The decision making process is becoming very complex, especially when it comes to the problem of enterprise restructuring. Intelligent solutions have to be utilized in order to cope efficiently with this issue. For this purpose the model for enterprise restructuring, named as COMPASS was developed. It aims to systematise the complex process of enterprise restructuring and to attach appropriate methods in the key decision making points, supporting the industry praxis with robust, scientifically founded, but practical and easy to use methods/tools. COMPASS is meant to be under continuous development. Now, the improvement of the process of generation of success factors (as a dominantly heuristic activity) and the optimization of the simulation process are the focal points. Here, the improvement of the generation of success factors through an application of finite-stage Markov decision processes (MDP) in the inventory problem is introduced. The objective is to find the optimal inventory policy and to determine the utility function. MDP are a flexible technique for stochastic and dynamic optimization problems and provide a powerful tool for optimizing performance in very different areas and their application in inventory management is only one of the examples.

Keywords
Markov Decision Processes, Inventory Management, Inventory Policy, Enterprise Restructuring, COMPASS.

1. Introduction

Probably the most emphasized feature of today's enterprise environment is its dynamics. The enterprises are functioning in very turbulent market with increased numbers of customers and competition, faster changes of the customer needs, etc. Market globalization, customer satisfaction, market over-saturation, technological (especially informational) prosperity, multi-national and multi-business enterprises are some of the reasons which are dramatically increasing market dynamics. In that direction, the need for quick actions and changes becomes the must for the enterprises that want to survive in such environment. In order to retain the accuracy of those actions enterprises have to have strong support in the key decision making points, powered with different tools and methods that will enable making decisions driven by facts. On one side, these tools and methods have to be more and more sophisticated to cope with the more complex problems, but on the other side they have to be still user friendly which should enable their utilization in praxis.

In that direction the model for enterprise restructuring, called COMPASS (COmpany's Management Purpose ASSistance) was developed, which clearly shows its main intention - to offer the management certain aid in the key decision making points in the complex process of enterprise restructuring. The research institutions are Fraunhofer Institut fuer Produktionstechnik und Automatisierung, Stuttgart, Germany and Faculty of Mechanical

Engineering, University of Ss. Cyril and Methodius, Skopje, R. Macedonia. The basic idea of this model is to make 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 overall improvement of the situation [8].

The (sub)model of performance measurement is the essential part of the methodology. Its basic elements are called subKEs (sub-Key Elements of Success). They are elements that are representing the current performance of the enterprise and can be seen as elements on which the enterprise is building its competitiveness. Examples for some of the subKEs embedded in COMPASS are: Time-Duration, Time-Reliability, Quality-Accuracy, etc., [7].

Phases of COMPASS are given in the Table 1.

Table 1: Phases of the model for enterprise restructuring [5]

#	Content of the phases in the model	Some of the utilised method approaches
1.	Evaluation of the present situation of the enterprise in a measurable form from strategic importance point of view (through measurement of the strategic importance of each subKE).	• AHP method • Team work
2.	Evaluation of the present situation of the enterprise in a measurable form from actual performance point of view (through measurement of the actual performance of each subKE).	• Audit • SWOT
3.	In order to determine the inconsistency of the subKEs from strategic and actual performance point of view I/P (Importance/Performance) matrixes are employed. The result of this phase is the list of Critical Elements (CEs) - subKEs which have unbalance between their importance and performance.	• I/P matrixes • Team work
4.	The beginning of the action generation is in the fourth phase. For every CE, appropriate Success Factor (SF) is induced. Examples for SFs are: shortening the cycle time, smaller lots, layout optimisation, more intensive education and training in some/all departments, ... 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 • Structured knowledge about performance measures
5.	This phase should structure the bunch of previously generated 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
6.	Selection of the optimal solution is determined in the sixth step. Previous phase gives the situation where every certain scenario leads, concerning only subKEs. In this phase, the financial effect of every action is estimated.	• Team work • Pay-back method • Costs/Gain diagram
7.	Implementation of the optimal action.	

The development of COMPASS is supposed to be a continuous process i.e. its development should never face its end. Our current efforts are directed towards two main issues: (i) the improvement of the process of generation of success factors (phase 4) and (ii) the optimization of the simulation process (phase 5) [9].

In the following text the first issue is elaborated. Namely, if we see closely the phase 4, we can see that it is lacking concrete methods. The decisions in this phase are made mainly heuristically [6]. Because of that, efforts have been made on two levels/directions to improve this phase:

- to utilize knowledge management techniques for better structuring of the knowledge in order to induce the success factors easily and
- to prepare predefined solutions for some of the more frequent success factors.

Such predefined solutions for the inventory management (as a success factor) are prepared by using the Markov Decision Processes (MDPs). MDPs have the two most desirable features: they have sophisticated background and simple user-interfaces can make them easy to use for users with different educational levels.

2. Characteristics of the Markov decision processes

2.1 Basic terms

Markov decision process model has generated a rich mathematical theory. Sometimes this model may appear quite simple and it encompasses a wide range of applications. Under certain mild assumptions, discrete-time MDPs can be applied to a variety of systems, where decisions are made sequentially to optimize a stated performance criterion. We will give an introduction of the Markov decision processes vocabulary in this section. *Decision epochs (t)* are points in time at which decisions are made. MDP binds previous, current, and future system decisions, through the proper definition of system states. *States* are variables that contain the relevant information needed to describe the system. The set of possible states is the state space (S). *Actions* are means by which the decision maker interacts with the system. In other words, when the system is in state i (or we can use s_t as notation), the decision maker chooses an action (decision) k (or a_t) from a certain action set ($A(i)$), which may depend on the observed state. For given state of the system and the chosen action, the immediate reward (or cost) (r or c) is received and it doesn't depend on the history of the process. The chosen action affects both the immediate rewards (costs) and subsequent rewards (costs). Markov decision processes models can include rewards or costs, because there is no essential difference between rewards and costs, since maximizing rewards is equivalent to minimizing costs (except the fact that data for costs are easier to be obtained). For given state of the system (i) and the chosen action, the state at the next decision time point (j) is determined by a transition law given by the transition probability (p_{ij}). *Transition probabilities* are distribution that governs how the state of the process changes as actions are taken over time and they posses the Markov property. In other words, the state variables must be defined in such way that for given current state of the system, the future transitions and rewards (costs) are independent of the past, which is the standard assumption of a Markov process. The planning horizon of the process may be finite, infinite or of random length. *The decision rule* $(d_i(R)=k)$ is a rule for a particular state that prescribes an action for each decision epoch. The *value function* (*utility function*) helps to determine maximum total expected reward (or minimum total expected cost) or other stated performance criterion. Collectively, decision time points (epochs), states, actions, rewards (costs) and transition probabilities form a Markov decision process (there are also other formal definitions of MDPs). *Policy* (R) is a collection of decision rules for all states. Under a fixed policy, the process behaves according to a Markov chain. The goal of an MDP is to provide an optimal policy, a decision strategy to optimize a particular criterion in expectation, which differs MDPs from other stochastic modelling techniques, used only to evaluate the consequences of a fully specified stochastic model.

2.2 Markov decision processes solution techniques

There are various factors to consider for choosing the right technique for solving a MDP problem. Classification can be made according to the formulation of the problem as a finite or infinite horizon (stage, period) problem.

If the problem is formulated as a finite horizon problem, the choice of the solution technique does not depend on discounting of the rewards (costs), or whether the objective is to maximize (minimize) expected total reward (cost) or to maximize (minimize) expected average reward (cost) – in this case they are equivalent. Here, the solution technique is backwards induction solution technique, which uses recursive equations of the dynamic programming. The same solution technique can be presented with directed graph (network, where the nodes are the states, and the numbers on the arcs are the rewards or the costs), and the problem of finding the shortest path (and its length) can be solved as a Markov decision process over a finite horizon. This solution technique is also referred to as value iteration approach, or method of successive approximations. This approach is used for quick finding of at least an approximation of the optimal policy.

But for infinite horizon, solution technique does depend on discounting and the objective. Here, exhaustive policy enumeration (which is impractical and infeasible if the state space and the proper action sets generate a huge number of stationary policies), value iteration, policy iteration (policy improvement algorithm) and linear programming approach can be considered.

In the exhaustive enumeration method, the (long-run) expected average cost per unit time is usually used as a measure of performance for finding the optimal policy:

$E(C)=\sum_{i=0}^{M} C_{ik}\pi_i$, where $k=d_i(R)$ for each i and $(\pi_0,\pi_1,\ldots,\pi_M)$ represents the steady-state distribution of the state of the system under the policy R .

In the linear programming approach the unknowns of the problem are modified, such that the optimal solution automatically determines the optimal action k , when the system is in state i , and the collection of all the optimal actions defines the optimal policy.

The policy iteration is an iterative approach, which consists of two steps, starting with an arbitrary stationary policy, and then determining a new improved one. The process ends when two successive policies are identical.

3. Finite horizon problems

We consider a problem solved directly with dynamic programming recursive equations, a model developed for solving a finite horizon (finite-stage) MDP problem. In the finite horizon case one has to control a system with, in general, non-stationary rewards (costs) and non-stationary transition probabilities, over a finite planning horizon of N periods, since the steady-state values are not reached yet. We consider the total expected reward (or cost) as utility (value) function. An optimal Markov policy, with deterministic, but in general non-stationary decision rules exists, and such an optimal policy can be obtained by backwards induction, which is based on the principle of optimality and it is an iterative approach starting at the end of the planning horizon. Here we give the idea behind backwards induction and the formalization of the same, illustrated on an inventory example, in order to find a policy that minimizes a certain cost value function. The idea is first to observe the last time period for all the possible states and decide the best action for all those states, which enables us to gain an optimal value for that state in that period. Next, we go in the next-to-last period for all the possible states and decide the best action for those states, knowing and using the optimal values of being in various states at the next time period. And we continue this process until we reach the present time period.

3.1 Inventory problem statement and its Markov decision processes model

Inventories usually have bad connotation because of the tied investments, costs for the care of the stored material/products and they are a subject of spoilage and obsolescence. There are many approaches developed that are aiming to reduce inventory levels and to increase the efficiency of the shop floor. Some of the most popular approaches are just-in-time manufacturing, lean manufacturing, flexible manufacturing, etc. Despite that, inventories do have positive aspects, such as providing a stable source of input required for production, reduction of ordering costs, reduction of the impact of the variability of the production rates in a plant, protection against failures in the processes, better customer service, variety and easy availability of the products, …etc. Inventories have practical and economic importance and they are significant portion of almost any company's assets. There is no unique model to handle the problem of making the optimal decisions regarding the inventory policy. The questions for the inventory policy are amenable to quantitative analysis associated with the inventory theory, including many inventory models and methods using different mathematical solution techniques. Here, an application of Markov decision processes, implemented on an inventory model example, very close to real inventory situations and applicable in practice is described. In the following text, first we state the very common inventory problem that is going to be solved as a discrete-time MDPs problem, with its characteristics and simplifying assumptions. Than, we discuss why it possesses the Markov property and give the general model for solving the problem.

Let's assume that COMPASS is utilized by certain trading company. The analysis of the first 3 steps points out that the subKEs called Time-Duration, Time-Reliability and Costs-Materials show unbalance of their strategic importance (evaluated as high) and its actual performance (evaluated as weak)[1]. After that the auxiliary indicators[2] were analysed. The indicator of supplier performance had excellent value (showing that the company can obtain the product almost promptly and with out any problems). The indicator for the lost sales had very high value (showing that the company had lost significant sales because of shortage of products when there was a need for them). The indicator for holding costs had very high value (showing that the company had the products when they were not needed). On the basis of these facts, it was decided that the ordering policy i.e. the inventory policy should be a matter of improvement. In other words, the inventory policy was detected as a success factor.

Next thing to do is to define the inventory problem statement for the company. Let's assume that the inventory problem statement contains the following: (i) inventory problem with no backlogging, (ii) single item, (iii) variable weekly demand with known probability distribution (Poisson distribution with mean 1), (iv) trade off between holding costs and lost sales costs (surplus demand is lost), (v) instantaneous delivery, (vi) limited storage capacity of *M* units, (vii) only whole units can be sold, (viii) costs and demand distribution do not change from week to week.

The random variable X_t is the state (the number) of the inventories at the end of week $t\left(t=0,1,2,\ldots\right)$. X_0 represents the number of items on hand at the outset. So, the state at time t equals the number of items at the end of week t. The random variables X_t are dependent. The random variable $D_t\left(t=1,2,\ldots\right)$ represents the demand for the item and is the number of items that would be sold in week t if the inventory is not depleted. Otherwise it includes lost sales. For the relevance of this model we assume that D_t are independent and identically distributed random variables. Given that the current state is $X_t=i$, X_{t+1} depends only on D_{t+1}. Since it is independent of any history of the system prior to time t, the stochastic process $\{X_t\}\left(t=0,1,\ldots\right)$ has the Markovian property.

In the following text, the essential notations and equations for the Markov decision process model are given.

The connection between the consecutive states is $s_{t+1}=s_t+a_t-\min\{D_t,s_t+a_t\}$;

[1] COMPASS evaluates the strategic importance and actual performance of each subKE in more details; since the evaluation of subKEs is not the focal point of this text, only rough description on these issues is given.

[2] Auxiliary indicators are a significant part of the performance measurement system of COMPASS [7]. They are divided as I (influencing the condition of the certain subKE) and B (being influenced by the condition of the certain subKE) and at the moment they (especially the I-indicators) present the essential basis for generation of the success factors.

The probability of the demand to take a certain value is $P(D_{t+1} = j) = p_j$, $j = 0,1,2,\ldots$;

If we go to minimize the relevant costs, we can use the following recursive equations to solve our example, just to illustrate costs determination and the iterations:

$$f_N(i) = \min_k \{C_{ik}\}$$

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

The transition probabilities are given with $p_{sj}^a = P\{s_{t+1} = j \mid s_t = s, a_t = a\} = \begin{cases} p_{s+a-j} & j \le s+a \\ \sum_{i=s+a}^{\infty} p_i & j = 0 \\ 0 & j > s+a \end{cases}$

In some MDPs problems, a thorough inspection is done at every decision epoch that results in classifying the condition into one of the possible states. After historical data on these inspection results are gathered, statistical analysis is done on how the state of the system evolves from one decision epoch to the next. This is how the relative frequency (probability) of each possible transition from the state in one epoch to the state in the following epoch and the associated reward (cost) are obtained. These transition probabilities form the transition probability matrix. MDPs are very data intensive, since the transition probabilities are governing the stochastic process and the rewards (costs) are allowed to vary according to the decision made at each decision epoch. As we mentioned before, the transition probabilities in our inventory example depend on demand probability distribution and are determined using probability theory, so the polices are randomized.

3.2 Inventory problem solution

The reason why this approach is attractive is that there is a quick method of finding an optimal policy when the process has only N periods to go – that is the probabilistic dynamic programming.

In order to illustrate the solution procedure, in this part we give concrete data for the example of the inventory model given previously. Some of them are: the inventory level fluctuates between a minimum of 0 items and a maximum of 3 items, so the possible states of the system at time t (the end of week t) is the state space $S = \{0,1,2,3\}$, (Table 2) i.e. the only possible values for the random variable X_t are 0, 1, 2 or 3.

The action sets (given in Table 3) are: $A(0) = \{0,1,2,3\}$, $A(1) = \{0,1,2\}$, $A(2) = \{0,1\}$, $A(3) = \{0\}$, since we have a limited space of $M = 3$ units, so $s_t + a_t \le M$ $(t = 0,1,2,\ldots)$.

Table 2 Classifying the condition of the system in four states

State	Condition
0	0 items on hand
1	1 item on hand
2	2 items on hand
3	3 items on hand

Table 3 Classifying the actions depending on states

Decision	Action	Relevant states
0	0 items to order	0,1,2,3
1	1 item to order	0,1,2
2	2 items to order	0,1

3	3 items to order	0

An important modelling decision concerns which distribution to use for demand. Transition probabilities depend on probability distribution of the demand. For models with small state space or when the expected demand in a time interval is small the recommended distributions are Poisson or exponential [4]. Here we take Poisson distribution with a mean of 1:

$$P\{D_{t+1}=j\}=\frac{(1)^j e^{-1}}{j!}\,(j=0,1,\dots).$$

For our purposes we compute:

$$p_0=P\{D_{t+1}=0\}=0.368,\ p_1=P\{D_{t+1}=1\}=0.368,\ p_2=P\{D_{t+1}=2\}=0.184,$$
$$P\{D_{t+1}\geq 3\}=0.080.$$

The transition probability data according to model formulation given above can be calculated as shown with the given “matrices” and they are not given in a sense of a transition probability matrices, but only a data matrices for the allowed transitions, and we put “–“ in the place where the transitions are infeasible. Only the matrix P^0 can be observed as transition matrix for the policy: do not make orders no matter the state. As we’ll see later, state 0 is absorbing state.

Transition matrices can be obtained for a certain stationary policy with determined and fixed decision rules, which is not the case in our representation used only for transparency. The decision making process evaluating the expected revenue (lost) resulting from a predefined course of action for a given state of the system is said to be represented by a stationary policy. Policy enumeration is a method for problem analysis and for best policy choice. Evaluation of all possible stationary policies of the decision problem which is equivalent to an exhaustive enumeration process, can be used only if the number of stationary policies is reasonably small and mostly for the infinite horizon problems. We emphasize that this method can be impractical, even for the limited size problems. In this example, 24 stationary policies are possible, but the question is which of them are practical. In the finite case, when the conditional transition probabilities are not yet steady-state probabilities, the unconditional probabilities can be obtained for every stationary policy if the probability distribution of the initial state is specified.

$$P^0=\left\|p_{ij}^0\right\|=\begin{bmatrix}\sum_{i=0}^{\infty}p_i & 0 & 0 & 0\\ \sum_{i=1}^{\infty}p_i & p_0 & 0 & 0\\ \sum_{i=2}^{\infty}p_i & p_1 & p_0 & 0\\ \sum_{i=3}^{\infty}p_i & p_2 & p_1 & p_0\end{bmatrix},\ P^1=\left\|p_{ij}^1\right\|=\begin{bmatrix}\sum_{i=1}^{\infty}p_i & p_0 & 0 & 0\\ \sum_{i=2}^{\infty}p_i & p_1 & p_0 & 0\\ \sum_{i=3}^{\infty}p_i & p_2 & p_1 & p_0\\ - & - & - & -\end{bmatrix},$$

$$P^2=\left\|p_{ij}^2\right\|=\begin{bmatrix}\sum_{i=2}^{\infty}p_i & p_1 & p_0 & 0\\ \sum_{i=3}^{\infty}p_i & p_2 & p_1 & p_0\\ - & - & - & -\\ - & - & - & -\end{bmatrix},\ P^3=\left\|p_{ij}^3\right\|=\begin{bmatrix}\sum_{i=3}^{\infty}p_i & p_2 & p_1 & p_0\\ - & - & - & -\\ - & - & - & -\\ - & - & - & -\end{bmatrix}.$$

If we calculate the values of the transition probabilities, we obtain

$$P^0 = \left\| p_{ij}^0 \right\| = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.08 & 0.184 & 0.368 & 0.368 \end{bmatrix}, \quad P^1 = \left\| p_{ij}^1 \right\| = \begin{bmatrix} 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.08 & 0.184 & 0.368 & 0.368 \\ - & - & - & - \end{bmatrix},$$

$$P^2 = \left\| p_{ij}^2 \right\| = \begin{bmatrix} 0.264 & 0.368 & 0.368 & 0 \\ 0.08 & 0.184 & 0.368 & 0.368 \\ - & - & - & - \\ - & - & - & - \end{bmatrix}, \quad P^3 = \left\| p_{ij}^3 \right\| = \begin{bmatrix} 0.08 & 0.184 & 0.368 & 0.368 \\ - & - & - & - \\ - & - & - & - \\ - & - & - & - \end{bmatrix}.$$

Statistical analysis can prove that these transition probabilities are unaffected by also considering what the states were in prior weeks, which is the “lack of memory” property called Markovian property.

For every transition probability a cost data should be computed according to the model formulas and initial values. The holding cost $h(s_t + a_t)$, the lost sales cost $l(D_t)$, and the order cost $O(a_t)$ depend on the variables in brackets. Purchasing cost could also be included in total costs. The values of these types of costs for the example are given in Table 4 (the details how to obtain certain values will be skipped in this occasion).

Table 4 Expected total cost data for a week, caused by a certain decision

Decision (k)	**State** (i)	**Holding cost, €**	**Lost sales cost, €**	**Order cost, €**	**Purchasing cost, €**	**Total cost for a week, €,** (C_{ik})
0	0	0	6320	0	0	6320
	1	10	2640	0	0	2650
	2	20	800	0	0	820
	3	30	0	0	0	30
1	0	10	2640	10	1000	3660
	1	20	800	10	1000	1830
	2	30	0	10	1000	1040
2	0	20	800	20	2000	2840
	1	30	0	20	2000	2050
3	0	30	0	30	3000	3060

For the algorithm transparency we use the following tables (Table5-Table8). These tables contain the values from the recurrence equations given in the previous section. Let $N = 4$ (weeks) – we have a reason for choosing bigger value to approximate the optimal policy.

Table 5 Stage 4

State (i)	C_{ik}				Optimal solution	
	$k=0$	$k=1$	$k=2$	$k=3$	$f_4(i)$	k^*
0	6320	3660	2840	3060	2840	2
1	2650	1830	2050	-	1830	1
2	820	1040	-	-	820	1
3	30	-	-	-	30	0

The first approximation calls for ordering 2 items if the inventory level is 0, ordering 1 item if the inventory level is 1 or 2, and not to put an order if the inventory level is 3, and that is the optimal policy for this stage.

Table 6 Stage 3

State (i)	$C_{ik} + p_{i0}^k f_4(0) + p_{i1}^k f_4(1) + p_{i2}^k f_4(2) + p_{i3}^k f_4(3)$				Optimal solution	
	$k=0$	$k=1$	$k=2$	$k=3$	$f_3(i)$	k^*
0	9160	7285.248	4564.96	3936.72	3936.72	3
1	5118.32	3554.96	2926.72	-	2926.72	2
2	2544.96	1916.72	-	-	1916.72	1
3	906.72	-	-	-	906.72	0

The second approximation calls for ordering 3 items if the level is 0, 2 items if the level is 1, 1 item if the level is 2, and 0 items if the inventory level is 3. That is the optimal policy for this stage. This policy recommendation continues in the next two stages (Table 7 and Table 8) and that is probably the optimal policy for this example for the infinite horizon case, but we can not prove this unless we use other approaches. $f_n(i)$ is the expected total cost from the stages $n, n+1, \ldots, N$, if the process starts at state i at the beginning of the week n.

Table 7 Stage 2

State (i)	$C_{ik} + p_{i0}^k f_3(0) + p_{i1}^k f_3(1) + p_{i2}^k f_3(2) + p_{i3}^k f_3(3)$				Optimal solution	
	$k=0$	$k=1$	$k=2$	$k=3$	$f_2(i)$	k^*
0	10256.72	7225.04	5661.68	4952.48	4952.48	3
1	6215.04	4651.68	3942.48	-	3942.48	2
2	3641.68	2932.48	-	-	2932.48	1
3	1922.48	-	-	-	1922.48	0

Table 8 Stage 1

State (i)	$C_{ik} + p_{i0}^k f_2(0) + p_{i1}^k f_2(1) + p_{i2}^k f_2(2) + p_{i3}^k f_2(3)$				Optimal solution	
	$k=0$	$k=1$	$k=2$	$k=3$	$f_1(i)$	k^*
0	11272.48	8240.8	6677.44	5968.24	5968.24	3
1	7230.8	5667.44	4958.24	-	4958.24	2
2	4657.44	3948.24	-	-	3948.24	1
3	2938.24	-	-	-	2938.24	0

Total expected costs for the four observed weeks are $f_1(0) = 5968.24$ if the state at the beginning of the week 1 is 0, $f_1(1) = 4958.24$ if the state is 1, …etc.

Instead of summary of this part, it can be stated that all these calculations can be easily "packed" in various software packages dealing with spreadsheets. So, the managers on the basis of few (important) inputs (like, holding cost, the lost sales cost, etc.) can make the decisions easily. One can argue that those important inputs are not easy to be obtained. That is true, but it is also true that it can be seen as one of the additional advantages of the approach in a sense that it is forcing the responsible management to stay focused on monitoring the important things of the business.

4. Conclusions

Here, the utilization of finite-stage Markov decision processes in one model for enterprise restructuring is described. More concretely, their utilization for solving inventory management problems is presented.
At the end, we will try to discuss the dilemma whether the finite-stage problems corresponds to the real world situations. 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, since its iterations can be performed simply and quickly. But it definitely obtains an optimal policy for an n-period problem after n iterations [10].
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. There are some methods that are more memory-efficient than policy iteration and value iteration algorithms. There are some solution techniques that find near-optimal solutions in short time. For each action and state pair, we need a transition probability matrix and a reward function, which are enormous data requirements. Infinite-stage Markov decision process problems can be formulated and solved as linear programs. Also we can find approximate solution techniques that are promising. There are extensions of Markov decision processes, because of their limitations. But the finite-stage Markov decision processes problems are more likely to be found in reality of inventory management, where there is a recursive nature of the problem. In practice it is not usual to have infinite planning horizon. That is the reason why we set out this model which is very close to real situations in inventory management, but yet easy to analyze and understand.

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, http://www.math.leidenuniv.nl/~lnmb/courses/Lecture notes-2007.pdf
3. Schaefer A.: *Markov decision processes*, EWO Seminar, October 26, 2006 http://www.egon.cheme.cmu.edu/ewocp/docs/SchaeferMarkovdecisionprocesses.pdf
4. Jensen P. A. , Bard J. F.: *Operations Research, Models and Methods*, http://www.me.utexas.edu/~jensen/ORMM/index.html
5. Minovski R. Jovanoski D.: *Utilization of the Importance/Performance (I/P) Matrices in a Model for Enterprise Restructuring*. Proceedings of the 44th European Quality Congress Volume 1, pp. 79-86, Budapest, 2000
6. Minovski R.: *Creation of the Model for Overall Enterprise Restructuring*, Doctoral Thesis, Faculty of Mechanical Engineering, Skopje, 1999, (in Macedonian)
7. Minovski R., Jovanoski D. and Zeh K-P: *COMPASS – ein universelles Modell zur Unternehmensstrukturierung*, Refa – Nachrichten, Heft 5/Oktober, pp.53-58, Darmstadt, 2002
8. Minovski R., Jovanoski D., Muthsam H.: *MORE – a Methodology for Overall Restructuring of the Enterprises*, Journal of "Industrial Engineering & Management", Vol. 4, No. 4, Shanghai, 1999, (in Chinese)
9. Minovski R., Jovanoski B.: *Methodology for overall company restructuring and the simulation as added value*, Wing journal, 3/2008, pp., 12-17, Graz, 2008
10. Hillier F. S. Liebermann G. J.: *Introduction to Operations research*, McGraw-Hill, New York, Eight edition 2005