FRI-1.405B-1-MIP-01

NETWORK RISKS IN MARKOV DECISION PROCESSES

Acad. Vassil Sgurev, Prof. Dr. Sci
Bulgarian Academy of Sciences,
Inst. of Information and Communication Technologies, Sofia, Bulgaria
E-mail: vsgurev@gmail.com

Ass. Prof. Stanislav Drangajov
Department Intelligent Systems,
Inst. of Information and Communication Technologies - BAS, Sofia, Bulgaria
E-mail: sgrangajov@gmail.com

***Abstract**: A class of Markov Decision Making Processes (MDP) is proposed in this work, considering the network risks. Risk is considered as a product of two measures one of which is the probability for an adverse event at the process' passing through a given state. It is proved that in case of the same values of these probabilities a network flow of risks is received which has one-to-one mapping to the MDP. Relations between these two controllable processes are obtained.*

A case is investigated when the probabilities of adverse events are different for the different states and a method is proposed through which in this case the optimal solutions for the MDP with risks can be found. The results received are confirmed by appropriate numerical examples.

The possible areas of application of the MDP with risks being proposed are pointed out.

***Keywords:** Markov decision processes; risk; network flows; optimization.*

A BRIEF INTRODUCTION

Markov processes take an important place in the probability theory. Their role significantly grew up when in the frame of the scientific area 'Operations research' 'Markov Decision Processes' (MDP) appeared and achieved wide popularity.

In (Sgurev, V., 1993) additional linear constraints were introduced in the MDP and as a result the usage of the concept 'Markov Flows' was proposed.

The concept 'risk' started to be used lately in various dynamic processes, including stochastic ones, and it to be embedded in some mathematical structures used in decision making support systems.

A new class of MDP is considered in the present work – Markov Decision processes with risks.

As the Markov processes have considerable applications, then from the up-to-date aspect, accounting for the risk at their running may be of substantial importance. The way in which the risk would be embedded in the Markov structures plays a very important role in theoretical aspect.

NETWORK FLOW INTERPRETATION OF MDP

An expedient approach is the functioning of a MDP to be considered on an arbitrary network. Let $G(X,U)$ be defined as a directed graph with a set of nodes (vertices) X and a set of edges (arcs) U, whose number is equal to $|X| = n$ and $|U| = m$, respectively. The indices of the graph's nodes are $I = \{i \,/\, x_i \in X\}$, and of the arcs - $I' = \{ x_{ij} \,/\, (i, j) \in U \}$. Indices of nodes connected with outgoing from x_i arcs are denoted by $A_i = \{j/x_{ij} \in U\}$, and of those, which are connected through incoming in x_i arcs – by $B_i = \{j/x_{ij} \in U\}$ (Christofides N., 1986).

Let a set of probability functions $\{p_{ij}^k\}$ be assigned, such that for $i \in I$ and $k \in K_i$ it is true:

$$0 \le p_{ij}^k \le 1; \ \sum_{j \in A_i} p_{ij}^k = 1; \tag{1}$$

where $K_i = \{1, 2, \ldots, k\}$ is a set of the indices of the possible selectable actions from the node of index i.

The probability the Markov process to get into state of index i and from it an action of index $k \in K_i$ to be selected will be denoted by x_i^k. The gain which is obtained if the DMP is in state i and from it action $k \in K_i$ is selected is denoted by a_i^k.

Markov process is turned into a controllable one in the following way:

a) From its initial state of index i_0 at choice of action k the process gets into some of the states of index from A_i and yields a profit of $a_i^k\, x_i^k$. In this way the initial step is implemented.

b) At receiving information in what a new state $j \in A_i$ the process has gotten into a new choice of an action $k \in K_j$ is made and the process makes a new step and gets into a new state of index $s \in A_j$, bringing a new profit of $a_j^k\, x_j^k$.

This controllable, multistep process is repeated until the MDP executed the necessary number of steps, or did not get into state $i \in I$, which is defined as final, finishing.

If the objective function is defined as

$$L = \sum_{i\in I} \sum_{k\in K_i} a_i^k x_i^k \;\longrightarrow \max \text{ (min)}, \tag{2}$$

then defining of optimal actions $k \in K_i$ for each state $i \in I$ will be reduced to solving of the following network flow programming problem (P. A. Jensen, W. J. P. Barnes, 1987) from (2) to (5) , namely, for each $i \in I$:

$$\sum_{k\in K_i} x_i^k - \sum_{j\in B_i} \sum_{k\in K_i} p_{ji}^k x_j^k = \begin{cases} v_i\,, \text{if } x_i \in S; \\ 0, \text{if } x_i \notin S \cup T; \\ -v_i, \text{if } x_i \in T \end{cases} \tag{3}$$

$$x_i^k \geq 0;\ k \in K_i; \tag{4}$$

$$x_i^k \leq 1;\ k \in K_i; \tag{5}$$

where S is a set of sources, T – a set of consumers at which $(S \cup T) \subseteq X$ and $S \cap T = \emptyset$; $\{v_i / x_i \in S\}$, $\{v_j / x_j \in T\}$ and $\sum_{x_{i\in S}} v_i = \sum_{x_{j\in T}} v_j$; $\emptyset$ - empty set.

If optimal control of MDP is sought for, then at each getting into each state $i \in I$ the optimization problem from (2) to (5) is solved and optimal actions $\{x_i^k / \ k \in K_i\ \}$ are realized, which represent mixed (stochastic) actions from the same state. Това означава, че не се избира една единствена поликика k от K_i, а се определят вероятностите, с които всяка една политика от K_i се използва на дадената стъпка.

In the controllable Markov process thus defined on the network, i.e. on the graph $G(X,U)$, risks arise in some applications due to unprecise realization of the process. It is expedient these risks to be quantitavely assessed by using $\{x_i^k\}$. Then the risk would characterize the losses at their realization, and following the widely accepted approach (Sgurev V., St. Drangajov, 2014) risk r_i^k will be represented as a product of two measures – the value of x_i^k and the probability p_i^k for an adverse event at the realization of the latter, i.e. for each $i \in I$ и $k \in K_i$

$$r_i^k = p_i^k x_i^k; \tag{6}$$

where $0 \leq p_i^k \leq 1$.

It follows from relations from (4) to (6) that for each $i \in I$ and $k \in K_i$

$$0 \leq r_i^k \leq 1. \tag{7}$$

Two cases will be considered, related to the defining of r_i^k:

A. Each quantity $\{p_i^k / i \in I;\ \ k \in K_i\}$ will be supposed to be of one and the same value of the probability for an adverse event, i.e. for each $i \in I$ and $k \in K_i$

$$p_i^k = p. \tag{8}$$

Then

$$r_i^k = p x_i^k. \tag{9}$$

If both sides of relations (2) to (5) are multiplied by p then according (8) and (9) it will be received:

$$Lp = L_r = \sum_{i\in I} \sum_{k\in K_i} a_i^k r_i^k \;\longrightarrow \max \text{ (min)}; \tag{10}$$

subject to the following constraints: for each $i \in I$

$$\sum_{k\in K_i} r_i^k - \sum_{j\in B_i} \sum_{k\in K_i} p_{ji}^k r_j^k = \begin{cases} v_i^r, \text{if } x_i \in S; \\ 0, \text{if } x_i \notin S \cup T\ ; \\ -v_i^r, \text{if } x_i \in T \end{cases} \tag{11}$$

$$r_i^k \geq 0;\ k \in K_i; \tag{12}$$

$$r_i^k \leq 1;\ k \in K_i; \tag{13}$$

where $v_i^r = pv_i$; $i \in S \cup T$.

Four latter relations define a network flow of controllable Markov risk $\{r_i^k\}$. Its comparison to relations (2) to (9) shows that a one-to-one mapping exists between the two controllable flows – the Markovian one $\{x_i^k\}$ and the Markov flow of risks, i.e., after defining the optimal MDP by relations (2) to (5), all parameters of the Markov flow of risks from (10) to (13) may be received through formula (6), and vice versa.

The linear form L from (2) shows the profit from the realization of the MDP, and L_r from (10) points out the loss from the realizations of risks. The difference L - L_r is and indicator for the decreasing of the profit at accounting for network risks through (8).

B. Let the requirement (8) not be complied with, and the probabilities of adverse events $\{p_i^k\}$ are not equal to each other. So relation (11) is not observed for risks $\{r_i^k\}$ and as so they are not elements of a network flow, i.e., the existence of a Markov flow of risks cannot be guaranteed.

It is proposed in this case risks to be accounted for through the objective function L.. Through $\sum_{i\in I}\sum_{k\epsilon K_i} b_i^k p_i^k x_i^k$ will be denoted the total cost to be incurred in relation to the risks arising in each $i \in I$ when choosing action $k \in K_i$. The coefficients $\{b_i^k\}$ show the value to be paid to cover a unit of risk $r_i^k = px_i^k$. Then it follows that the profit in L should be decreased respectively, and the new objective L' will be equal to:

$$L' = \sum_{i\in I}\sum_{k\epsilon K_i}(a_i^k x_i^k - b_i^k p_i^k x_i^k) = \sum_{i\in I}\sum_{k\epsilon K_i} x_i^k(a_i^k - b_i^k p_i^k) = \sum_{i\in I}\sum_{k\epsilon K_i} c_i^k x_i^k . \quad (14)$$

As $c_i^k = a_i^k - b_i^k p_i^k$, then it follows from (2) and (14) that $L' \leq L$.

This second approach makes it possible to determine the optimal MDP by taking into account risks, for which in the most general case $p_i^k \neq p_j^k$ at $i \neq j$. For this purpose the problem from (3) to (5) should be solved for the linear form L' of (14).

In principle, it is possible to introduce different upper bounds of the capacities of the values $\{x_i^k\}$, and then it will come to using the 'mincut-maxflow' theorem (D. R. Ford, D. R. Fulkerson, 2010) and to the more general concept of controllable Markov flows, as it is realized in (Sgurev, V., 1993). However, in the present paper, more simple and straightforward approaches and models are addressed that are directly related to the risks, which is the main goal of the authors.

NUMERICAL EXAMPLE

A graph $G(X,U)$ with six nodes from X and eight arcs from U is given, as shown in Fig. 1.

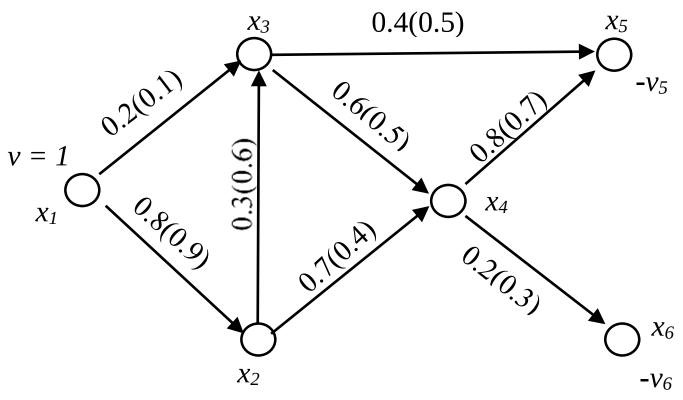

Fig. 1

The parameters of the MDP with risks, defined on the graph $G(X,U)$ have the following values:

a) Transition probabilities

for x_1: $p_{1,2}^1 = 0.8$; $p_{1,2}^2 = 0.9$; $p_{1,3}^1 = 0.2$; $p_{1,3}^2 = 0.1$;

for x_2: $p_{2,3}^1 = 0.3$; $p_{1,3}^2 = 0.6$; $p_{2,4}^1 = 0.7$; $p_{2,4}^2 = 0.4$;

for x_3: $p_{3,4}^1 = 0.6$; $p_{3,4}^2 = 0.5$; $p_{3,5}^1 = 0.4$; $p_{3,5}^2 = 0.5$;

for x_4: $p_{4,5}^1 = 0.8$; $p_{4,5}^2 = 0.7$; $p_{4,6}^1 = 0.2$; $p_{4,6}^2 = 0.3$.

b) Arc ratings at different actions

$$a_1^1 = 9;\ a_1^2 = 10;\ a_2^1 = 3;\ a_2^2 = 2;$$
$$a_3^1 = 3;\ a_3^2 = 4;\ a_4^1 = 14;\ a_4^2 = 15.$$

c) Flows

$v_1 = 1$; $T = \{x_5, x_6\}$; v_5 and v_6 – variables.

A. Case #1 at observing (11).

Then the equations from (2) to (15) have the following form:

$$x_1^1 + x_1^2 = 1;$$
$$x_2^1 + x_2^2 - 0.8x_1^1 - 0.9x_1^2 = 0;$$
$$x_3^1 + x_3^2 - 0.2x_1^1 - 0.1x_1^2 - 0.3x_2^1 - 0.6x_2^2 = 0;$$
$$x_4^1 + x_4^2 - 0.7x_2^1 - 0.4x_2^2 - 0.6x_3^1 - 0.5x_3^2 = 0;$$
$$-0.4x_3^1 - 0.5x_3^2 - 0.8x_4^1 - 0.7x_4^2 + v_5 = 0;$$
$$-0.2x_4^1 - 0.3x_4^2 + v_6 = 0;$$
$$v_5 + v_6 = 1;$$
$$x_1^1 \geq 0;\ x_1^2 \geq 0;\ x_2^1 \geq 0;\ x_2^2 \geq 0;\ x_3^1 \geq 0;\ x_3^2 \geq 0;\ x_4^1 \geq 0;\ x_4^2 \geq 0;$$

and the objective function is in the following form:

$$L = 9x_1^1 + 10x_1^2 + 3x_2^1 + 2x_2^2 + 3x_3^1 + 4x_3^2 + 14x_4^1 + 15x_4^2 \rightarrow \max.$$

The solution to this problem of network-flow programming leads to the following mixed optimal.

$$x_1^1 = 0;\ x_1^2 = 1;\ x_2^1 = 0.9;\ x_2^2 = 0;\ x_3^1 = 0.37;\ x_3^2 = 0;\ x_4^1 = 0;\ x_4^2 = 0.85;\ v_5 = 0.74;\ v_6 = 0.26.$$

The optimal linear form is equal to $L = 26.59$.

It is assumed that for the MDP the probability of an adverse event is the same for all states and is equal to 0.18. Then, the missed profit when taking into account the risk is equal to

$$L_r = pL = 26.59 \text{ x } 0.18 = 4.79,$$

and the overall reduction of the profit in L for the MDP with risks amounts to

$$L' = L - L_r = 26.59 - 4.79 = 21.80.$$

B. If the probabilities of adverse events have different values of $\{p_i^k\}$, the risk is taken into account in the objective functions (2) and (14). Then

$$c_i^k = a_i^k - b_i^k p_i^k = a_i^k - d_i^k;$$

where $d_1^1 = 2$; $d_1^2 = 1.8$; $d_2^1 = 0.5$; $d_2^2 = 0.4$; $d_3^1 = 0.7$; $d_3^2 = 0.7$; $d_4^1 = 4.5$; $d_4^2 = 5.3$.

Summarized ratings are equal to:

$$c_1^1 = 7;\ c_1^2 = 8.8;\ c_2^1 = 2.5;\ c_2^2 = 1.6;\ c_3^1 = 2.3;\ c_3^2 = 3.3;\ c_4^1 = 9.5;\ c_4^2 = 9.7.$$

The objective function L_c is determined by

$$L_c = 7x_1^1 + 8.8x_1^2 + 2.5x_2^1 + 1.6x_2^2 + 2.3x_3^1 + 3.3x_3^2 + 9.5x_4^1 + 9.7x_4^2 \rightarrow \max.$$

The solution of the network-flow optimization problem thus defined, results in the following optimal values of $\{x_i^k\}$, namely:

$$x_1^1 = 0;\ x_1^2 = 1;\ x_2^1 = 0.9;\ x_2^2 = 0;\ x_3^1 = 0.37;\ x_3^2 = 0;$$
$$x_4^1 = 0;\ x_4^2 = 0.85;\ v_5 = 0.74;\ v_6 = 0.26;\ L_c = 20.17.$$

In Fig. 2 and Fig. 3. the optimal solutions for the two numerical examples are shown – **A.** и **B.**, respectively.

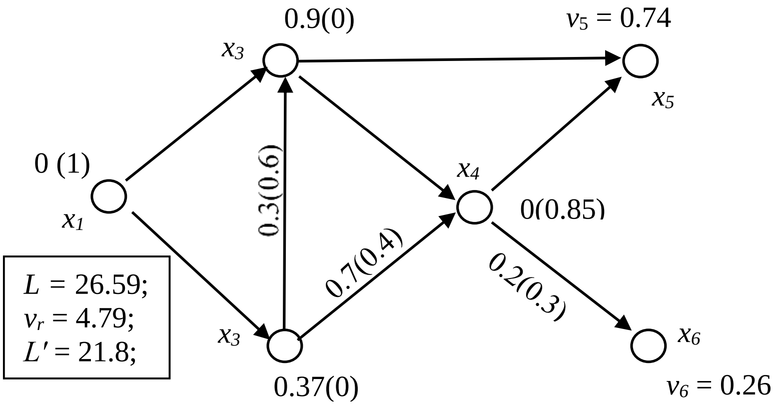

Fig. 2

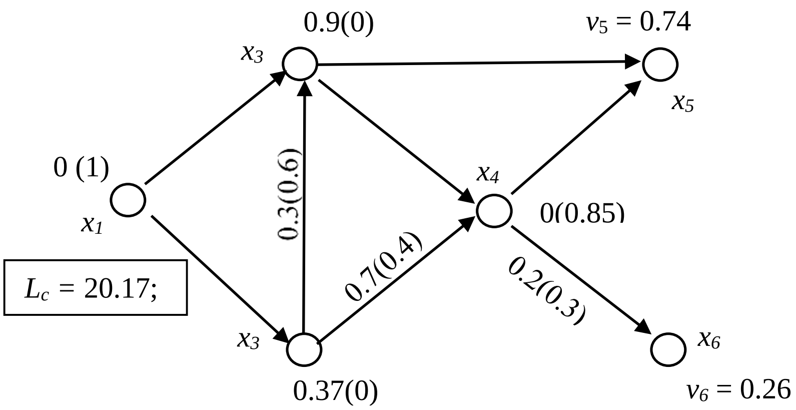

Fig. 3

Since the probabilities of adverse events $\{p_i^k\}$ are selected so that their arithmetic mean is close to the constant probability

$$p \sim p_i^k;\ i \in I;\ k \in K_i;$$

and the arcs'ratings are close to each other for the two numerical experiments – **A.** and **B.**, the optimal solutions $\{x_i^k\}$ in both cases coincide with each other.

Characteristically, for the specific data of both experiments, pure policies are optimal rather than mixed ones, as can be seen from the null values of some $\{x_i^k\}$ in Fig. 2 and 3.

In both cases - **A.** and **B.**, the values obtained for L' and L_c differ by 7% from each other. All this shows that both methods can work effectively. If the parameters $\{p_{ij}\}$ are averaged, the simpler model **A.** will give approximately the same results as the model in case **B.** The numerical experiments performed support the research results obtained in the present work.

The proposed controllable Markov processes with risks can find applications in a various areas of material production, services and public practice - in information and communication technologies, production processes and systems, organizational, financial and economic structures and processes, military, research and etc.

CONCLUSION

1. Introduction of network risks in Markov Decision Making Processes (MDP) is proposed in the present work. The risk is considered as product of multiplication of two measures: the probability the process to get into state *i* and the selection of a distinct action from this state, and the probability of an adverse event when passing through this state.

2. A case study has been investigated where the probability of adverse events is the same across all process's states and it is proved that this leads to a network Markov flow of risks, which one-to-one mapped with a DMP with no risks. A number of results a received related to these two one-to-one mapped processes.

3. A case is considered when the probabilities of adverse events are different in different states of the Markov process and it has been proved that in this case it is impossible to have a network Markov flow of risks. A method for determining the optimal solutions for this class of MDP with risks is proposed.

4. A numerical example has been performed in which the optimal mixed policies for Markovian control of risks are determined from the specific data and the feasibility of the proposed theoretical results for MDP with risks has been proved.

5. The possible areas for the application of such MDP with risks are pointed out.

REFERENCES

Sgurev, V., Markov Flows, Sofa, Publishing House of Bulgarian Academy of Sciences, 1993, ISBN 954-430-103-8, https://search.rsl.ru/ru/record/01000556731 (in Russian)

Sgurev V., St. Drangajov - Intelligent Control of Flows with Risks on a Network, Proceedings of the 7th IEEE International Conference Intelligent Systems - IS’14, September 24–26 2014, Warsaw, Poland, ISSN 2194-5357, ISSN 2194-5365 (electronic), ISBN 978-3-319-11309-8, ISBN 978-3-319-11310-4 (eBook), DOI 10.1007/978-3-319-11310-4, Volume 2: Tools, Architectures, Systems, Applications, Springer International Publishing, Switzerland, P. Angelov et al. (eds.), Advances in Intelligent Systems and Computing vol. 323, 2014, pp. 27-35.

D. R. Ford, D. R. Fulkerson, Flows in Networks, Princeton University Press Princeton, NJ, USA©2010, ISBN: 0691146675 9780691146676, https://dl.acm.org/citation.cfm?id=1942094

Christofides N., Graph theory: An Algorithmic Approach, London [etc.]: Academic Press, 1986.

H. Mine and S. Osaki, Markovian Decision Processes, Copyright © 1971 Society for Industrial and Applied MathematicsRead, ISSN (print): 0036-1445, ISSN (print): 0036-1445

Eitan Altman, Constrained Markov Decision Processes, Chapman and Hall/CRC Published March 30, 1999, ISBN 9780849303821 - CAT# C0382

Sgurev, V., Network Flows with General Constraints, Publishing House of the Bulgarian Academy of Sciences, Sofia, 1991 (in Bulgarian)

Ruszczyński, Risk-averse dynamic programming for Markov decision processes, © Springer and Mathematical Optimization Society 2010, A. Math. Program. (2010) 125: 235. https://doi.org/10.1007/s10107-010-0393-3

Ravindra K. Ahuja, Thomas L. Magnanti, James B. Orlin, Network Flows: Theory, Algorithms, and Applications, Pearson, 2013, ISBN 1292042702, 9781292042701

P. A. Jensen, W. J. P. Barnes, Network Flow Programming, Krieger Pub Co, 1987, ISBN-13:978-0894642104, ISBN-10:0894642103, https://www.amazon.com/Network-Flow-Programming-Paul-Jensen/dp/0894642103Christofides N., Graph theory: An Algorithmic Approach, London [etc.]: Academic Press, 1986.