Adaptive Resource Control

Machine Learning Approaches to Resource Allocation
in Uncertain and Changing Environments

Ph.D. Thesis Booklet

Balázs Csanád Csáji

Supervisor: László Monostori, D. Sc.

Faculty of Informatics (IK),
Eötvös Loránd University (ELTE)

Doctoral School in Informatics,
Foundations and Methods in Informatics Ph.D. Program,
President: Prof. János Demetrovics, Member of HAS

Computer and Automation Research Institute (SZTAKI),
Hungarian Academy of Sciences (HAS, MTA)

Budapest, Hungary, 2008

1. Introduction

Information technology has been making an explosion-like progress since the middle of the past century. However, as computer science broke out from laboratories and classrooms and started to deal with "real world" problems, it had to face major difficulties. Namely, in practise, we mostly have only *incomplete* and *uncertain* information on the system and the environment that we must work with, additionally, they may even *change* dynamically, the problem may be *non-stationary*. Moreover, we also have to face *complexity* issues, viz., even if we deal with static, highly simplified and abstract problems and it can be known that the solution exists and can be attained in finitely many steps, the problem could still be *intractable*, viz., we might not have enough computation power (or even enough storage space) to achieve it in practise, as this is the case, e.g., with many NP-hard problems.

One way to overcome these difficulties is to apply *machine learning* techniques. It means designing systems which can *adapt* their behavior to the current state of the environment, *extrapolate* their knowledge to the unknown cases and learn how to *optimize* the system. These approaches often use *statistical* methods and satisfy with *approximate*, *suboptimal* but *tractable* solutions concerning both computational demands and storage space.

The importance of *learning* was recognized even by the founders of computer science. It is well known, e.g., that John von Neumann was keen on *artificial life* and, besides many other things, designed *self-organizing* automata. Alan Turing can be another example, who in his famous paper, which can be treated as one of the starting articles of *artificial intelligence* research, wrote that instead of designing extremely complex and large systems, we should design programs that can learn how to work efficiently by themselves.

In the dissertation I considered an important problem with many practical applications, which has all the difficulties mentioned in the previous parts, namely: *resource allocation*.

Resource allocation problems (RAPs) are of high practical importance, since they arise in many diverse fields, such as manufacturing production control (e.g., production scheduling), warehousing (e.g., storage allocation), fleet management (e.g., freight transportation), personnel management (e.g., in an office), scheduling of computer programs (e.g., in massively parallel GRID systems), managing a construction project or controlling a cellular mobile network. RAPs are also central to management science (Powell and Van Roy, 2004). In the thesis I considered optimization problems that include the assignment of a finite set of reusable resources to non-preemtive, interconnected tasks that have stochastic durations and effects. My main objective in the thesis was to investigate efficient decision-making processes which can deal with the allocation of scarce resources over time with a goal of optimizing the objectives. For "real world" applications it is important that the solution should be able to deal with large-scale problems and handle environmental changes, as well.

One of my main motivations for investigating RAPs was to enhance manufacturing production control. Regarding contemporary manufacturing systems, difficulties arise from unexpected tasks and events, non-linearities, and a multitude of interactions while attempting to control various activities in dynamic shop floors. Complexity and uncertainty seriously limit the effectiveness of conventional production control approaches (e.g., deterministic scheduling). In the thesis I applied *machine learning* (ML) techniques to achieve the *subop-*

timal control of a generalized class of stochastic RAPs, which can be vital to an *intelligent manufacturing system* (IMS) (Hatvany and Nemes, 1978). An IMS utilizes the results of *artificial intelligence* research and is expected to solve, within certain limits, unprecedented, unforeseen problems on the basis of even incomplete and imprecise information.

Different kinds of RAPs have a huge number of exact and approximate solution methods, for example, (see Pinedo, 2002) in the case of scheduling problems. However, these methods primarily deal with the static (and often strictly deterministic) variants of the various problems and, mostly, they are not aware of uncertainties and changes. Special (deterministic) RAPs which appear in the field of *combinatorial optimization*, e.g., the traveling salesman problem (TSP) or the job-shop scheduling problem (JSP), are *strongly NP-hard* and, moreover, they do not have any good polynomial-time approximation, either. In the stochastic case, RAPs can be often formulated as *Markov decision processes* (MDPs) and by applying *dynamic programming* (DP) methods (Bertsekas, 2001), in theory, they can be solved *optimally*. However, due to the phenomenon that was named *curse of dimensionality* by Bellman, these methods are highly intractable in practice. The "curse" refers to the *combinatorial explosion* of the required computation as the size of the problem increases. Some authors, for example, Powell and Van Roy (2004), talk about even three types of curses concerning DP algorithms. This has motivated *approximate* approaches that require a more tractable computation, but often yield *suboptimal* solutions (Bertsekas, 2005).

It is beyond my scope to give a general overview on different solutions to RAPs, hence, I only concentrate on the part of the literature that is closely related to my approach. My solution belongs to the class of *approximate dynamic programming* (ADP) algorithms which constitute a broad class of discrete-time control techniques. Note that ADP methods that take an actor-critic point of view are often called *reinforcement learning* (RL).

Zhang and Dietterich (1995) were the first to apply an RL technique for a special RAP. They used the $TD(\lambda)$ method with iterative repair to solve a static scheduling problem, namely, the NASA space shuttle payload processing problem. Since then, a number of papers have been published that suggested using RL for different RAPs. The first reactive (closed-loop) solution to scheduling problems using ADP algorithms was briefly described in (Schneider et al., 1998). Riedmiller and Riedmiller (1999) used a *multilayer perceptron* (MLP) based neural RL approach to learn local heuristics. Aydin and Öztemel (2000) applied a modified version of Q-learning to learn dispatching rules for production scheduling. Powell and Van Roy (2004) presented a formal framework for RAPs and they applied ADP to solve them. Later, their solution was parallelized by Topaloglu and Powell (2005).

Recently, *support vector machines* (SVMs) were applied by Gersmann and Hammer (2005) to improve iterative repair (local search) strategies for *resource constrained project scheduling problems* (RCPSPs). An agent-based resource allocation system with MDP-induced preferences was presented in (Dolgov and Durfee, 2006). Beck and Wilson (2007) gave proactive solutions for job-shop scheduling problems based on the combination of *Monte Carlo simulation*, solutions of the associated deterministic problem, and either constraint programming or tabu-search. Finally, the effects of environmental changes on the convergence of RL algorithms was theoretically analyzed by Szita et al. (2002).

2. Main Contributions

The new scientific results of my dissertation can be summarized in six theses which can be organized in two thesis groups. The first concerns with efficiently solving RAPs in presence of *uncertainties*, while the second contains results on managing *changes* in the dynamics.

2.1 Stochastic Resource Allocation

In Chapter 2 of the dissertation I studied RAPs in presence of uncertainties. I also suggested machine learning based solution methods to handle them. My main contributions were:

T 1.1 *I proposed a formal framework for studying stochastic resource allocation problems with reusable resources and non-preemtive, interconnected tasks having temporal extensions. I provided a reformulation of it as a controlled Markov process and I showed that this system was capable of handling both reactive and proactive solutions.*

I defined a formal RAP which was a natural generalization of several standard resource management problems, such as scheduling, transportation and inventory management ones. I reformulated this general RAP as a stochastic shortest path (SSP) problem (a special MDP) having favorable properties, such as, it was aperiodic, its state and action spaces were finite, all policies were proper and the space of control policies could be safely restricted. I defined reactive solutions of stochastic RAPs as control policies of the reformulated problem. I also investigated proactive solutions and treated them as policies of the non-observable MDP corresponding to the reformulated MDP. I analyzed the relation between the optimal cost-to-go of reactive and proactive solutions, too. These results can be found in Section 2.1 of the dissertation.

T 1.2 *I suggested methods based on the combination of approximate dynamic programming, simulated annealing and either hash tables or kerner regression, in order to compute and represent reactive solutions. I confirmed the effectiveness of this approach with results of numerical experiments on both benchmark and industry related problems.*

In order to compute a good approximation of an optimal control policy, I suggested ADP methods, particularly, fitted Q-learning. Regarding value function representation, I studied two approaches: hash tables and support vector regression (SVR), especially, ν-SVRs. In both cases, I defined the inputs as numerical feature vectors. Since the problem to be faced was an SSP, I used off-line learning after each episode. An episode consisted of a state-action-cost trajectory, generated by simulation. Regarding controlling the ratio of exploration and exploitation during simulation I applied the Boltzmann formula. These ideas are described in Sections 2.2.1 and 2.2.2 of the dissertation. I also presented results of numerical experiments on both benchmark and industry-related data. I measured the performance of my approach on hard benchmark flexible job-shop scheduling problems and I also demonstrated the scaling properties of my method by experiments on a simulated factory producing mass-products. These experiments are presented in Sections 4.1.2 and 4.1.4.

T 1.3 *I provided further improvements based on rollout algorithms, action space decomposition, clustering and distributed sampling, in order to speed up the computation of a solution. I presented results of numerical experiments to support their effectiveness.*

The suggested improvements were: application of limited lookahead rollout algorithms in the initial phases to guide the exploration and to provide the first samples to the approximator; decomposing the action space to decrease the number of available actions in the states; clustering the tasks to reduce the length of the trajectories and so the variance of the cumulative costs; as well as two methods to distribute the proposed algorithm among several processors having either shared or distributed memory architecture. These approaches are contained in Sections 2.2.3 and 2.2.4. I presented results of numerical experiments concerning the improvements in Sections 4.1.3 and 4.1.5. These experiments illustrate the effects of clustering depending on the size of the clusters and the speedup relative to the number of processors.

2.2 Varying Environments

In Chapter 3 of the dissertation I analyzed environmental changes. I also investigated value function based RL methods in varying environments. My main contributions were:

T 2.1 *I deduced bounds in discounted MDPs concerning the dependence of the optimal value function and value functions of (stationary, Markovian, randomized) control policies on the transition-probabilities, the immediate-costs and the discount factor.*

I proved that the value function of a (stationary, Markovian, randomized) control policy in a discounted MDP Lipschitz continuously depended on the immediate-cost function (Theorem 11). A similar result was already known for the case of transition-probability functions, however, I presented an improved bound for that case, as well (Theorem 10). I also presented value function bounds (Theorem 12) for the case of changes in the discount factor and demonstrated through an example that this dependence was not Lipschitz continuous. Then (with Lemma 14) I extended these results to optimal value functions, too. These theorems can be found in Section 3.1.

T 2.2 *I introduced a new MDP model, called (ε, δ)-MDP, in order to study varying environments. It allows asymptotically bounded changes in the transition-probabilities and the immediate-costs. I proved that changes in the discount rate could be incorporated into the immediate-costs, thus, discount changes did not have to be modeled.*

In order to study changing environments, I introduced (ε, δ)-MDPs (Definition 23) that were generalizations of classical MDPs and ε-MDPs. In this extended model the transition-probability function and the immediate-cost function might change over time, provided that the accumulated changes remain asymptotically bounded, viz. bounded in the limit. I showed (Lemma 24) that potential changes in the discount factor could be incorporated into the immediate-cost function, thus, discount changes did not have to be considered. These contributions are presented in Section 3.2.2.

T 2.3 *I proved a general convergence theorem for time-dependent stochastic iterative algorithms. As a corollary, I deduced an approximation theorem for value function based reinforcement learning (RL) methods working in (ε, δ)-MDPs. I also illustrated these results through three classical RL algorithms as well as numerical experiments.*

I analyzed stochastic iterative algorithms where the value function update operator might change over time. I proved a relaxed convergence theorem for this kind of algorithm (Theorem 26). As a corollary, I got an approximation theorem for value function based RL methods working in (ε, δ)-MDPs (Corollary 27). Furthermore, I illustrated my results through three classical RL algorithms. I deduced relaxed convergence properties in (ε, δ)-MDPs for asynchronous value iteration, Q-learning and TD(λ) – temporal difference learning. In order to demonstrate the results, I presented two simple stochastic iterative algorithms, a "well-behaving" and a "pathological" one. These contributions are described in Sections 3.2.3 and 3.2.4. I also presented results of numerical experiments which highlight some features of working in varying environments. I showed two experiments concerning adaptation in Section 4.2.

3. Scientific Publications

3.1 Journal Articles

1. Schuh, G.; Monostori, L.; Csáji, B. Cs.; Döring, S.: Complexity-based modeling of reconfigurable collaborations in production industry, *Annals of the CIRP – Manufacturing Technology*, Vol. 57., No. 1., 2008 (in print)
   - Impact factor: 0.989

2. Argyros, A.; Bártfai, G.; Eitzinger, Ch.; Kemény, Zs.; Csáji, B. Cs.; Kék, L.; Lourakis, M.; Reisner, W.; Sandrisser, W.; Sarmis, T.; Umgeher, G.; Viharos, Zs. J.: Smart sensor based vision system for automated processes, *International Journal of Factory Automation, Robotics and Soft Computing*, Thomson Scientific Journal, Vol. 3., 2007, pp. 118–123.

3. Csáji, B. Cs.; Monostori, L.; Kádár, B.: Reinforcement learning in a distributed market-based production control system, *Advanced Engineering Informatics*, Vol. 20, No. 3, July 2006, pp. 279–288.
   - Impact factor: 1.295
   - Indep. citations: 2

4. Monostori, L.; Csáji, B. Cs.: *Stochastic dynamic production control by neurodynamic programming*, *Annals of the CIRP – Manufacturing Technology*, Vol. 55, No. 1, 2006, pp. 473–478.
   - Impact factor: 0.989
   - Indep. citations: 3

5. Kádár, B.; Monostori, L.; Csáji, B. Cs.: Adaptive approaches to increase the performance of production control systems, *CIRP Journal of Manufacturing Systems*, Vol.

34, No. 1, 2005, pp. 33–43.
- Indep. citations: 3

6. Monostori, L.; Csáji, B. Cs.; Kádár, B.: Adaptation and learning in distributed production control, *Annals of the CIRP – Manufacturing Technology*, Vol. 53, No. 1, 2004, pp. 349–352.
- Impact factor: 0.973
- Indep. citations: 8

3.2 Book Chapters

7. Argyros, A.; Bártfai, G.; Eitzinger, Ch.; Kemény, Zs.; Csáji, B. Cs.; Kék, L.; Lourakis, M.; Reisner, W.; Sandrisser, W.; Sarmis, T.; Umgeher, G.; Viharos, Zs. J.: Smart sensor based vision system for automated processes, In book: *Emerging Technologies, Robotics and Control Systems*, editor: Salvatore Pennacchio, Vol. 2., pages 24–29, International Society for Advanced Research, 2007

8. Csáji, B. Cs.; Monostori, L.: Stochastic reactive production scheduling by multi-agent based asynchronous approximate dynamic programming, Lecture Notes in Computer Science; 3690: *Lecture Notes in Artificial Intelligence*, Proceedings of the 4th International Central and Eastern European Conference on Multi-Agent Systems (CEEMAS), September 15–17, Budapest, Hungary, 2005, pp. 388–397.
- Impact factor: 0.251
- Indep. citations: 2

9. Csáji, B. Cs.; Küng, J.; Palkoska, J.; Wagner, R.: On the automation of similarity information maintenance in flexible query answering systems; *Lecture Notes in Computer Science*, Vol. 3180: Proceedings of the 15th International Conference on Database and Expert Systems Applications, (DEXA), 2004 pp. 130–140.
- Impact factor: 0.402

10. Csáji, B. Cs.; Kádár, B.; Monostori, L.: Improving multi-agent based scheduling by neurodynamic programming, Lecture Notes in Computer Science; 2744: *Lecture Notes in Artificial Intelligence*, Proceedings of the 1st International Conference on Holonic and Multi-Agent Systems for Manufacturing (HoloMAS), 2003, pp. 110–123.
- Indep. citations: 9

3.3 Conference Papers

11. Csáji, B. Cs.; Monostori, L.: A complexity model for networks of collaborating enterpises, *17th IFAC World Congress*, July 6–11, 2008; Seoul, Korea (accepted)

12. Monostori, L.; Csáji, B. Cs.: Complex adaptive systems (CAS) approach to production systems and organisations; *41st CIRP Conference on Manufacturing Systems*; May 26–28, 2008; The University of Tokyo, Japan (keynote paper)

13. Egri, P.; Csáji, B. Cs.; Kemény, Zs.; Monostori, L.; Váncza, J.: Komplexität der Bedarfsprognosen und ihre Wirkungen in kooperativen Logistiknetzwerken; *10th Paderborner Frühjahrstagung, Reagible Unternehmen in dynamischen Märkten*; March 26, 2008; Paderborn, Germany (accepted)

14. Csáji, B. Cs.; Monostori, L.: Modeling networks of collaborating enterprises as complex systems, Preprints of the *IFAC Workshop on Modelling, Management and Control* (MIM'07), November 14–16, 2007, Budapest, Hungary, pp. 7–12.

15. Monostori, L.; Csáji, B. Cs.: Production structures as complex adaptive systems, Proceedings of the *40th CIRP International Seminar on Manufacturing Systems*, May 30 – June 1, 2007, Liverpool, United Kingdom.

16. Csáji, B. Cs.; Monostori, L.: Adaptive sampling based large-scale stochastic resource control, Proceedings of the *21st National Conference on Artificial Intelligence* (AAAI-06), July 16–20, 2006, Boston, Massachusetts, pp. 815–820.

17. Csáji, B. Cs.; Monostori, L.: Adaptive algorithms in distributed resource allocation, Proceedings of the *6th International Workshop on Emergent Synthesis*, (IWES), Kashiwa, The University of Tokyo, Japan, August 18–19, 2006. pp. 69–75.
    - Indep. citations: 1

18. Viharos, Zs. J.; Kádár, B.; Monostori, L.; Kemny, Zs.; Csáji, B. Cs.; Pfeiffer, A.; Karnok D.: Integration of production-, quality- and process monitoring for agile manufacturing, Proceedings of the *13rd IMEKO World Congress*, Metrology for a Sustainable Development, September, 17–22, Rio de Janeiro, Brazil, 2006

19. Csáji, B. Cs.; Monostori, L.: Stochastic approximate scheduling by neurodynamic learning, *16th IFAC World Congress*, July 3–8, 2005, Prague, Czech Republic.
    - Indep. citations: 1

20. Pfeiffer, A.; Kádár, B,; Csáji, B. Cs.; Monostori, L.: Simulation supported analysis of a dynamic rescheduling system, *IFAC Symposium on Manufacturing, Modelling, Management and Control*, October 21–22, 2004, Athens, pp. 24–29.

21. Csáji, B. Cs.; Kádár, B.; Monostori, L.; Pfeiffer, A.: Simulation supported agent-based adaptive production scheduling, *International IMS Forum*; Global Challenges in Manufacturing, May 17–19, 2004, Cernobbio, Lake Como, Italy, pp. 658–665.

22. Csáji, B. Cs.; Monostori, L.; Kádár, B.: Learning and cooperation in a distributed market-based production control system, Proceedings of the *5th International Workshop on Emergent Synthesis*, (IWES), May 24–25, Budapest, 2004, pp. 109–117.
    - Indep. citations: 6

23. Kádár, B.; Monostori, L.; Csáji, B. Cs.: Adaptive approaches to increase the performance of production control systems, Proceedings of the *36th CIRP International*

Seminar on Manufacturing Systems, Progress in Virtual Manufacturing Systems, June 3–5, 2003, Saarbrcken, Germany, pp. 305–312.
- Indep. citations: 4

24. Monostori, L.; Kádár, B.; Csáji, B. Cs.: The role of adaptive agents in distributed manufacturing, Proceedings of the *4th International Workshop on Emergent Synthesis* (IWES'02), May 9–10, 2002, Kobe, Japan, pp. 135–142.
- Indep. citations: 1

3.4 Papers not Connected to the Thesis

25. Csáji, B. Cs.; Rédei, M.: A racionális demokratikus véleményösszegzés korlátairól, *Magyar Filozófiai Szemle*, Vol. 1., 2008 (accepted)

26. Gilles, M.; Ballin, D.; Csáji, B. Cs.: Efficient clothing fitting from data; *12nd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, February 2–6, Plzen, Czech Republic, 2004, pp. 129–136.

27. Csáji, B. Cs.: In defense of the symmetry of true and false; Proceedings of the *6th Interdisciplinary Symmetry Congress and Exhibition* of ISIS (International Society for the Interdisciplinary Study of Symmetry), Symmetry: Art & Science, October 22–29, Tihany, Hungary, 2004, pp. 46–49.

3.5 Articles under Review

28. Csáji, B. Cs.; Monostori, L.: Value function based reinforcement learning in changing Markovian environments, *Journal of Machine Learning Research* (submitted in 2007)

29. Csáji, B. Cs.; Monostori, L.: Adaptive stochastic resource control: a machine learning approach, *Journal of Artificial Intelligence Research* (submitted in 2007)

30. Kemény, Zs.; Csáji, B. Cs.; Viharos, Zs., J.: Timing parameter optimization for vision-based monitoring of automated production lines, Journal of *Mechanical Systems and Signal Processing* (submitted in 2008)

Summary

category	**own articles**	**impact factor**	**ind. citations**
journal	6	4.246	16
book chapter	4	0.653	11
conference	14	0	13
other	(3)	(0)	(0)
under review	(3)	-	-
total	24 (30)	4.899	40

References

Aydin, M. E. and Öztemel, E. (2000). Dynamic job-shop scheduling using reinforcement learning agents. *Robotics and Autonomous Systems*, 33:169–178.

Beck, J. C. and Wilson, N. (2007). Proactive algorithms for job shop scheduling with probabilistic durations. *Journal of Artificial Intelligence Research*, 28:183–232.

Bertsekas, D. P. (2001). *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, Massachusetts, 2nd edition.

Bertsekas, D. P. (2005). Dynamic programming and suboptimal control: A survey from ADP to MPC. *European Journal of Control*, 11(4–5):310–334.

Dolgov, D. A. and Durfee, E. H. (2006). Resource allocation among agents with MDP-induced preferences. *Journal of Artificial Intelligence Research*, 27:505–549.

Gersmann, K. and Hammer, B. (2005). Improving iterative repair strategies for scheduling with the SVM. *Neurocomputing*, 63:271–292.

Hatvany, J. and Nemes, L. (1978). Intelligent manufacturing systems - a tentative forecast. In Niemi, A., editor, *A link between science and applications of automatic control; Proceedings of the 7th IFAC World Congress*, volume 2, pages 895–899.

Pinedo, M. (2002). *Scheduling: Theory, Algorithms, and Systems*. Prentice-Hall.

Powell, W. B. and Van Roy, B. (2004). *Handbook of Learning and Approximate Dynamic Programming*, chapter Approximate Dynamic Programming for High-Dimensional Resource Allocation Problems, pages 261–283. IEEE Press, Wiley-Interscience.

Riedmiller, S. and Riedmiller, M. (1999). A neural reinforcement learning approach to learn local dispatching policies in production scheduling. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence, Stockholm*, pages 764–771.

Schneider, J. G., Boyan, J. A., and Moore, A. W. (1998). Value function based production scheduling. In *Proceedings of the 15th International Conference on Machine Learning*, pages 522–530. Morgan Kaufmann, San Francisco, California.

Szita, I., Takács, B., and Lőrincz, A. (2002). ε-MDPs: Learning in varying environments. *Journal of Machine Learning Research (JMLR)*, 3:145–174.

Topaloglu, H. and Powell, W. B. (2005). A distributed decision-making structure for dynamic resource allocation using nonlinear function approximators. *Operations Research*, 53(2):281–297.

Zhang, W. and Dietterich, T. (1995). A reinforcement learning approach to job-shop scheduling. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1114–1120. Morgan Kauffman.