Chapter 2
Batch Reinforcement Learning

Sascha Lange, Thomas Gabel, and Martin Riedmiller

Abstract. Batch reinforcement learning is a subfield of dynamic programming-based reinforcement learning. Originally defined as the task of learning the best possible policy from a fixed set of a priori-known transition samples, the (batch) algorithms developed in this field can be easily adapted to the classical online case, where the agent interacts with the environment while learning. Due to the efficient use of collected data and the stability of the learning process, this research area has attracted a lot of attention recently. In this chapter, we introduce the basic principles and the theory behind batch reinforcement learning, describe the most important algorithms, exemplarily discuss ongoing research within this field, and briefly survey real-world applications of batch reinforcement learning.

2.1 Introduction

Batch reinforcement learning is a subfield of dynamic programming (DP) based reinforcement learning (RL) that has vastly grown in importance during the last years. Historically, the term 'batch RL' is used to describe a reinforcement learning setting, where the complete amount of learning experience—usually a set of transitions sampled from the system—is fixed and given a priori (Ernst et al, 2005a). The task of the learning system then is to derive a solution—usually an optimal policy—out of this given batch of samples.

In the following, we will relax this assumption of an a priori fixed set of training experience. The crucial benefit of batch algorithms lies in the way they handle a batch of transitions and get the best out of it, rather than in the fact that this set is fixed. From this perspective, batch RL algorithms are characterized by two basic

Sascha Lange · Thomas Gabel · Martin Riedmiller
Albert-Ludwigs-Universtität Freiburg, Faculty of Engineering, Georges-Köhler-Allee 079, D-79110 Freiburg, Germany
e-mail: {slange,tgabel,riedmiller}@informatik.uni-freiburg.de

M. Wiering and M. van Otterlo (Eds.): Reinforcement Learning, ALO 12, pp. 45–73.
springerlink.com

constituents: all observed transitions are stored and updates occur synchronously on the whole batch of transitions ('fitting'). In particular, this allows for the definition of 'growing batch' methods, that are allowed to extend the set of sample experience in order to incrementally improve their solution. From the interaction perspective, the growing batch approach minimizes the difference between batch methods and pure online learning methods.

The benefits that come with the batch idea—namely, stability and data-efficiency of the learning process—account for the large interest in batch algorithms. Whereas basic algorithms like Q-learning usually need many interactions until convergence to good policies, thus often rendering a direct application to real applications impossible, methods including ideas from batch reinforcement learning usually converge in a fraction of the time. A number of successful examples of applying ideas originating from batch RL to learning in the interaction with real-world systems have recently been published (see sections 2.6.2 and 2.6.5).

In this chapter, we will first define the batch reinforcement learning problem and its variants, which form the problem space treated by batch RL methods. We will then give a brief historical recap of the development of the central ideas that, in retrospect, built the foundation of all modern batch RL algorithms. On the basis of the problem definition and the introduced ideas, we will present the most important algorithms in batch RL. We will discuss their theoretical properties as well as some variations that have a high relevance for practical applications. This includes a treatment of Neural Fitted Q Iteration (NFQ) and some of its applications, as it has proven a powerful tool for learning on real systems. With the application of batch methods to both visual learning of control policies and solving distributed scheduling problems, we will briefly discuss on-going research.

2.2 The Batch Reinforcement Learning Problem

Batch reinforcement learning was historically defined as the class of algorithms developed for solving a particular learning problem—namely, the batch reinforcement learning problem.

2.2.1 The Batch Learning Problem

As in the general reinforcement learning problem defined by Sutton and Barto (1998), the task in the batch learning problem is to find a policy that maximizes the sum of expected rewards in the familiar agent-environment loop. However, differing from the general case, in the batch learning problem the agent itself is not allowed to interact with the system during learning. Instead of observing a state s,