



Master's thesis in
Information- and Communication Technology

Title:
**Empirical Evaluation of the
Bayesian Learning Automaton Family**

Candidate:
Terje Brårdland
Thomas Norheim

Supervisor:
Ole-Christoffer Granmo,
University of Agder



Introduction

The two-armed bandit problem is a classical optimization problem where a player sequentially selects and pulls one of two arms attached to a gambling machine, and each arm pull results in either a reward or penalty to the player. Each arm is associated with a certain fixed reward probability which is unknown to the player. The player needs to sequentially select and play an arm and receive a reward or a penalty in order to discover its true reward probability. The overall goal for the player is reward maximization, and the player needs to balance between exploiting existing knowledge or obtaining new knowledge by trying different arms. In the long run it may be beneficial to risk short term loss to gain greater certainty about the reward probability associated with each arm.

The problem as described above is the simplest case of the more general k -armed bandit problem and is concerned with the exploration vs. exploitation dilemma often found in optimization problems. A wide range of schemes, based on many different principles and theoretical foundations, has been proposed to solve this problem.

In this thesis we focus our attention on a series of schemes collectively referred to as the *Bayesian Learning Automaton (BLA) family*, originally introduced by Granmo with the Bayesian Learning Automaton designed for Bernoulli distributed reward. We investigate the theoretical foundation of the BLA proposed by Granmo and derive and introduce three new members to the BLA family, designed for Poisson and normally distributed reward.

Bayesian Learning Automata

The key concept of the Bayesian Learning Automata (BLA) is the use of Bayesian statistics both to represent a belief about the unknown reward distribution associated with each arm, and to balance the exploration and exploitation effort in the automata.

Bayesian methods may result in complex and tedious calculations which could make such approaches intractable. However, the BLA approach makes extensive use of conjugate prior distributions, which reduce complex and tedious computations to simple updating rules of parameters associated with these conjugate prior distributions. With these simple updating rules the prior belief about the reward distributions is updated to reflect the posterior belief after a reward is received.

Arm selection is performed by random sampling from the posterior belief about the reward distributions, and the arm that returns the highest sample value is selected. The probability that a BLA selects a particular arm can be interpreted as the probability that it is the optimal arm, given the reward received so far on this arm. Thus, a BLA gradually shifts its arm selection focus towards the arm which most likely is the optimal arm as rewards are received.

Results

We have performed an extensive evaluation of the Bayesian Learning Automaton family against competing schemes in the k -armed bandit problem with Bernoulli, Poisson and normally distributed reward. We also introduce the players in selected games from game theory, which from a player's point of view may be perceived as a bandit problem.

Player / Reward distribution	Bernoulli $p_o = 0.9$ $p_i = 0.8$	Poisson $\lambda \in [0,10]$	Normal $\mu \in [0,1]$ $\sigma^2 \in [0,1]$
BLA Bernoulli	0.988	-	-
BLA Poisson	-	0.980	-
BLA Normal	0.982	0.972	0.955
En-Greedy	0.988	0.928	0.858
Pursuit	0.960	0.893	0.842
UCB1-Tuned	0.977	-	-
UCB1-Normal	0.797	0.910	0.810
Poker	0.916	0.874	0.849

Table 1: Overall probability of selecting optimal arm over 100 000 iterations

In Table 1 we present selected results from the thesis for the 10-armed bandit problem with Bernoulli, Poisson and normally distributed reward. As the results from the table indicate, the members of the BLA family offer similar or better performance than its competing schemes in the selected experiments.

In Figure 1 we present the development of the regret, i.e the expected loss due to the optimal arm is not always selected, in the 2-armed bandit problem with normally distributed reward. As depicted, BLA Normal clearly offers the lowest regret when the number of iterations is sufficiently large in the selected experiment.

The presented results reflect the overall performance of the BLA in all experiments we conducted with the k -armed bandit problem, although some competing schemes were able to beat the BLA in specific experiments.

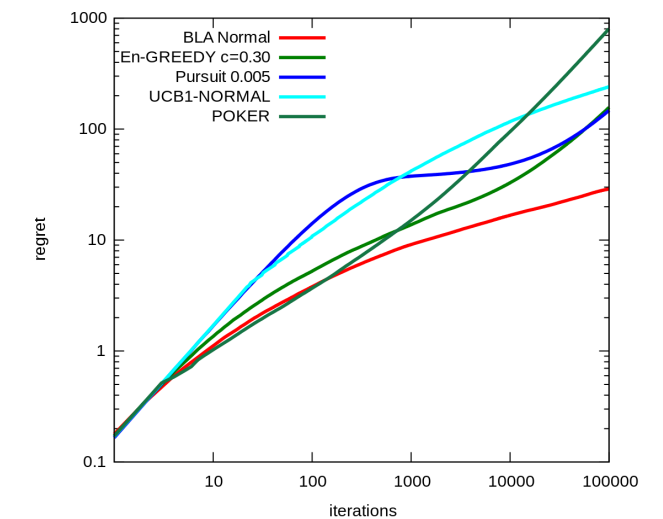


Figure 1: Regret for the 2-armed bandit problem with normally distributed reward

Also in the experiments with selected games from game theory the BLA offer impressive performance, despite the reward distributions are not fixed as in the k -armed bandit problem.

Conclusion

Through extensive experiments we show that the Bayesian Learning Automaton family overall outperforms all other comparable learning schemes in the k -armed bandit problem. The Bayesian Learning Automata are also among the top performers in all the games they were introduced to in this thesis.

Thus, we believe that the Bayesian Learning Automaton family is an important addition to the field of bandit playing algorithms and is an important area for further research.