ABCDE

Scientific Report

First name / Family name	Gergely Neu
Nationality	Hungarian
Name of the *Host Organisation*	INRIA
First Name / family name of the *Scientific Coordinator*	Daniil Ryabko
Period of the fellowship	01/09/2013 to 31/08/2014

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

I have spent a very successful year with the SequeL team at INRIA Lille. Besides pursuing my original research project concerning online learning in non-stationary Markov Decision Processes (MDPs), I have collaborated with a number of local colleagues and worked on a large variety of new topics in online optimization and bandit theory. Below, I present the main results of my research in each studied topic.

Non-stationary MDPs

In this topic, the goal is to propose efficient learning algorithms for the well-studied problem of Reinforcement Learning in MDPs, with an important relaxation of the traditional model: we allow the reward function to change in an arbitrary fashion as time progresses. During the course of my fellowship, I have worked on different variants of this problem, with some results described below.

- Based on my previous work on the topic, I proposed an efficient learning algorithm for a general class of this problem that attains the best possible performance guarantees, assuming perfect knowledge of the underlying MDP (except for the reward sequence, of course). Joint work with Alexander Zimin (now with IST Austria), published as [1].
- I also studied a variant of the problem where the sequence of rewards and transition functions are unknown functions of some side-observation, reducing the learning problem to identifying these unknown functions. This allows for efficient representation of large, continuous action and state spaces, capturing a large range of practical problems. I gave an efficient algorithm for this problem in [5], joint work with Yasin Abbasi-Yadkori (now with the Queensland University of Technology).
- I have also made progress with the problem of treating non-stationary when relaxing the assumption of perfect knowledge of the MDP. The new approach combines recent work on a Bayesian learning algorithm known as Thompson sampling and my recent papers [1,10]. The results obtained so far are very promising, although this work is not yet ready for publication. The current version benefited from discussions with Nathaniel Korda (now with the University of Oxford) and Rémi Munos (INRIA).

Online optimization

Studying the worst-case performance of online optimization algorithms has been a very active area of research in the past decades. A well-known problem with the obtained guarantees is, however, that they can be way too pessimistic in many cases of interest, as the worst cases rarely ever occur in practice. There has been substantial interest lately in proposing online optimization algorithms that can efficiently exploit easy problem instances and perform much better than overly careful algorithms, while still performing near-optimally in the worst cases. In my paper [4], I proposed a general algorithmic scheme that allows achieving the best of both worlds in the above sense, solving a very recent open problem posed at this year's Conference on Learning Theory (COLT 2014). Joint work with Amir Sani and Alessandro Lazaric (both of INRIA).

Bandit models

Multi-armed bandit models provide a popular framework to formalize (non-stationary) sequential learning tasks with partial information. In my recent work, I have contributed to the state of the art in different variants of the bandit problem. These contributions are the following.

- I have proposed a framework for bandit problems with combinatorial decision spaces where the decision sets are allowed to change in time. I have proposed an efficient algorithm that is capable to deal with a wide class of changing decision sets, progressing the state of the art on previously well-studied variants of this framework. The results appear in the submission [2], joint work with Michal Valko (INRIA).
- I have proposed a new algorithm for learning in an important recent partial observability model. The central element of this new algorithm is a new procedure for bias-variance tradeoff in multi-armed bandit models called "implicit exploration". The resulting algorithm is much more efficient (both computationally and information-theoretically) than previously proposed ones. The results appear in the submission [3], joint work with Tomás Kocák, Michal Valko and Rémi Munos (all of INRIA).

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

During the course of my fellowship, I have published the following conference paper (even though it is a result of a previous collaboration):

[1] A. Zimin and **G. Neu**: *Online Learning in Markov Decision Processes by Relative Entropy Policy Search*. In Advances in Neural Information Processing Systems 26 (NIPS), pp. 1583-1591, 2013.

More importantly, I have submitted the following papers to the same top-tier machine learning conference:

[2] **G. Neu** and M. Valko: *Online Combinatorial Optimization with Stochastic Decision Sets and Adversarial Losses*. Submitted to NIPS 2014. (review scores: 6,7,8)

[3] T. Kocák, **G. Neu**, M. Valko and R. Munos: *Efficient Learning by Implicit Exploration in Bandit Problems with Side Observations*. Submitted to NIPS 2014. (review scores: 7,7,8)

[4] A. Sani, **G. Neu** and A. Lazaric: *Exploiting Easy Data in Online Optimization*. Submitted to NIPS 2014. (review scores: 6,8,9)

Review scores are on a scale of 1 through 10. Papers with an average score above 7 usually fall in the upper 20% quantile and go on to be accepted. Furthermore, I have co-authored the following paper, ready to be submitted to a suitable conference:

[5] Y. Abbasi-Yadkori and **G. Neu**: *Online Learning in MDPs with Side Information*. Arxiv preprint, 2014.

I have also submitted the following journal publications during my fellowship:

[6] L. Devroye, G. Lugosi and **G. Neu**: *Random-Walk Perturbations for Online*

Combinatorial Optimization. Submitted to IEEE Transactions on Information Theory, 2014.

[7] **G. Neu**, A. György, and Cs. Szepesvári: *The Online Loop-free Stochastic Shortest-Path Problem*. Submitted to Mathematics of Operations Research, 2014.

Finally, the following paper has also been technically published during my fellowship:

[8] **G. Neu** and G. Bartók: *An Efficient Algorithm for Learning with Semi-Bandit Feedback*. In Proceedings of the 24th International Conference on Algorithmic Learning Theory (ALT), pp. 234-248, 2013.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

I have attended the following major machine learning conferences during the course of my fellowship:

- The 24th International Conference on Algorithmic Learning Theory (ALT), Singapore, October 6-9, 2013. I gave a talk on my paper [8].
- The 26th Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, USA, December 5-10, 2013. I presented a poster on my paper [1] and gave an extended talk at the workshop on "Perturbations, Optimization, and Statistics" about my paper [8].

Furthermore, I have attended the following local workshops:

- Hermès workshop on recommendation systems, 27 March, 2014, Lille, France. I gave a talk on a preliminary version of my paper [2].
- Workshop on "Kernel Methods for Big Data", 31 March - 2 April, 2014, Lille, France.
- Workshop on "Challenging problems in Statistical Learning", 7-8 April, 2014, Paris, France.

I was also in charge of organizing the seminar series of the SequeL group at INRIA Lille.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

As my first REP visit, I have visited the lab of Prof. Andreas Krause at ETH Zürich, Switzerland between 2-8 May, 2014, where I gave a talk to his research group on a preliminary version of my paper [2]. During most of my visit here, I was mainly working on the journal version of my paper [8] with Gábor Bartók (then employed at Prof. Krause's group). I have also had some very inspiring discussions with Prof. Krause about recent work done in his lab and about future directions of research in online optimization.

I have paid my second REP visit to CNR IMATI led by Dr. Bruno Betrò in Milan, Italy, between 21-25 July, 2014. Although we couldn't find a convenient time for me to give a talk for the entire group, I have interacted with all the researchers who were present at the time. Dr. Betrò introduced me to their ongoing projects concerning various applications of machine learning and statistics. This visit was enlightening for both sides: while getting familiar with some practical aspects of my research field, I have also shared my experiences concerning the newest trends in machine learning – more specifically, feature learning, stochastic optimization and sequential prediction.