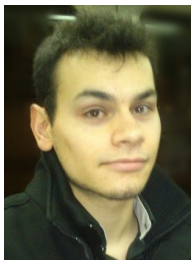
# Matthieu Zimmer

*PhD, Postdoctoral Researcher*

*Born on May 30 1990*
*French Nationality*

*Minhang District*
*Shanghai, China*
☎ *+86 188 2120 3357*
✉ *contact@matthieu-zimmer.net*
*https://matthieu-zimmer.net*

## Professional Experiences

2018 – ongoing
**Postdoctoral Researcher**, UM-SJTU Joint Institute, China - Shanghai.
Supervisor: Paul Weng (Assistant Professor)
Developing deep reinforcement learning algorithms.

2014 – 6 months
**Research internship**, ISIR - UPMC, France - Paris, *AMAC*.
Supervisor: Stéphane Doncieux (Professor UPMC-ISIR)
Study on transfer learning with reward shaping methods within a framework of lifelong learning. Developmental and evolutionary approach. Simulation in C++.
The principle was to first use a direct policy search in the sensorimotor space, i.e. with no pre-defined discrete sets of states nor actions, and then extract from the corresponding learning traces discrete actions and identify the relevant dimensions of the state to estimate the value function. Once this is done, the robot can apply reinforcement learning to be more robust to new domains and, if required, to adapt faster than a direct policy search.

2013 – 3 months
**Research internship**, LIP6, France - Paris, *DECISION*.
Supervisors: Paolo Viappiani (CNRS-LIP6), Paul Weng (UPMC-LIP6)
Bibliographical research, reading articles and state of the art on the integration of knowledge from an expert during reinforcement learning.
An agent (the "teacher") advises another one (the "student") by suggesting actions the latter should take, while learning a specific task in a sequential decision problem; the teacher is limited by a "budget" (the number of times such advice can be given). Implementation in C++ of a new idea : the teacher is also learning, he learns to give advice to propitious moments to the student. He is learning how to teach better. We provided experimental results with the Mountain car domain, showing how our approach outperforms the state-of-the-art heuristics.

2012 – 6 months
**Research internship**, INRIA - LORIA, France - Nancy, *Cortex & Maia*.
Supervisors: Yann Boniface (UL), Alain Dutech (INRIA), Nicolas Rougier (INRIA)
Meta-learning in neural networks.
Deepening ideas developed in articles of consciousness and meta-representations with multilayer perceptrons. How can they judge their own performances and improve them. Introduction to research, neural networks, latex and python.
A first neural network was learning a classification task, while a second one, called higher-order network, learned to bet if the prediction of the first network was correct from its hidden layer neurons. The higher-order network was indeed capable of learning such information, which meant that it can predict when the first network was going to fail. Thus, we proposed several architectures to combine the two networks in order to increase the overall prediction quality of the first network. The source code is available at github.com/matthieu637/anne.

2010 – 3 weeks
**Summer internship**, Mathieu Perrein France, France - Waldweistroff.
C# and WPF development using Microsoft Visual Studio.

## Training

October 2014 – January 2018 **PhD in Artificial Intelligence**, University of Lorraine, France - Nancy.

| | |
|---|---|
| Laboratory | LORIA (University of Lorraine, INRIA, CNRS) |
| Topic | Developmental Reinforcement Learning |
| Supervisors | Alain Dutech (Researcher), Yann Boniface (Associate Professor) |

2012–2014 **Master in Computer Science**, Pierre and Marie Curie University, Paris.

| | |
|---|---|
| Specialization | Artificial Intelligence and Decision |
| Research Training | Intelligent Agents, Learning and Decision |

Magna cum laude

2008–2012 **Bachelor in Computer Science**, University of Lorraine, France - Nancy.
Magna cum laude

2005–2008 **High School Diploma in Sciences**, Lycée Charlemagne, France - Thionville.
Specialization Mathematics

## Teaching Experiences

2014–2017 **Teaching Assistant**, *Engineer school* ENSEM, France - Nancy.

162 hours Algorithms and Programming in Python (2nd year integrated preparatory cycle).
Responsible Jean-Philippe Mangeot
Practical work on python imperative programming (pyzo IDE) : searching, sorting and small games (Reversi, Connect Four, ...). I wrote several practical work subject about data structures, Dijkstra, artificial intelligence and networking. I developed a first server (in Java) to create gaming party between two students, so they could challenge their artificial intelligence agent in a tournament determining their grades. During the last year, instead of comparing their agent on small games, we decided that the students had to create autonomous trading agents. Thus, I developed a second server (also in Java) to simulate a stock market exchange. In both cases, the students had simply to interface with the server in python, so they could focus on developing their artificial intelligence. The source code is available at github.com/matthieu637/cpp-2a-info.

8 hours Collaboration and Programming in Java (2nd year integrated preparatory cycle).
I designed and did seminars on Linux command-line, git, continuous integration, object-oriented design, unit testing, threads and synchronization in Java. To let students practice collaborative development, I set up a Jenkins instance communicating with a Github project.

35 hours Algorithmic and Java Object-Oriented Programming (1st year of engineering school).
Responsible Vincent Chevrier
Practical work on search algorithms, object-oriented design and Lego robot navigation.

## Supervision

2017 – 5 months **Achille Fedioun**, University of Lorraine, France - Nancy, LORIA.
With Alain Dutech and Yann Boniface.
End-of-studies internship (Master Computer Science and Engineer school): Reinforcement learning with continuous state and action spaces using model-free actor-critic algorithms with deep neural networks.
Achille had to compare the features of two new algorithms (Qprop and ACER) with ours. He extended one of our actor-critic agents with off-policy multi-step replay using the Retrace algorithm in C++. As experimental validation, he used the cluster of the lab to train deep neural networks on the half-cheetah environment.

2015
5 months

**Nicolas Lefebvre**, University of Lorraine, France - Nancy, LORIA.
With Alain Dutech and Yann Boniface.
End-of-studies internship (Master Cognitive Science): Reinforcement learning with continuous state and action spaces using model-free actor-only algorithms.
Nicolas had to explore if the Power algorithm could be used with neural networks instead of dynamic movement primitives. He developed the Power algorithm inside our C++ framework using Gaussian mixture policies. He experimentally validated his agent on the acrobot environment (double inverted pendulum).

## Computer skills

| | |
|---|---|
| Software | C++, Java, Python, Octave, C, Prolog, C#, OCaml |
| Web | J2EE, PHP, JavaScript & AJAX, HTML & CSS |
| Libraries | Boost, Caffe, SFML, CEGUI, GLib, Apache Commons, JFlex, Java Cup |
| Simulators | ODE, TORCS |
| Storage | PostgreSQL, Oracle, MySQL, SQLite, XML (Schema, DTD, XPath) |
| Utilities | KDevelop, Eclipse, Netbeans, Microsoft Visual Studio, CodeBlocks, Latex, Git |
| OS. | ArchLinux, Debian, Ubuntu |
| Other | Computer cluster (Grid5000, AWS), Shell Bash & CSH, UML, Lua |

## Languages

| | |
|---|---|
| French | Mother tongue |
| English | Fluent |

## Interests

| | |
|---|---|
| Development | Isometric 2D game in team, Dynamic website with applet-server, Server management |
| Others | IT News, Free software, Self Hosting, Hardware and Cryptocurrency |
| Award | NIPS 2017 - Learning to run (top 100) |
| Sport | Badminton (5 years in association) |

## Research

My research mainly focus on reinforcement learning with neural networks (also called deep reinforcement learning). We aimed at designing agents that take decisions in an unknown environment, and learn through their own interaction with this environment to maximize a given criterion. More precisely, I am working on model-free actor-critic algorithms to deal with continuous environments (in state and action), trying to make them more data-efficient without loosing the scalability of neural networks.

### PhD Context

Developmental robotics emerged in the 2000s, when researchers initiate to equip robots with sophisticated learning algorithms without providing pre-defined representations and with less specific knowledge given *a priori*. This thesis is part of this current by adding the hypothesis that the goal of the agent is to maximize a reward signal: it learns by reinforcement. Its body is located in a rich and continuous environment, it does not manipulate discrete symbols, and therefore does not have a countable set of actions or preconceived states. The models learned by the agent are nonlinear (neural networks); it must build its own representations through its many interactions with the environment, without relying on a set of preconceived basic functions. While many reinforcement learning algorithms are discrete or rely on linear models of basic functions, the main question addressed here is how to long-term learn by reinforcement in a continuous space of states and actions, with nonlinear models and less specific knowledge. To address this problem, a final hypothesis is formulated: the body of the agent, and thus the difficulty of the problem it solves, grows with time, in order to allow a partially guided exploration of the research space.

### Research I would like to conduct

Through the interaction with several researchers, my personal conviction to build a general intelligence artificial contains the following points:

- a situated agent acting in a rich environment with complex interactions is a key element for intelligence,
- robotic is not a necessary condition, the rich environment can be simulated,
- giving less *a priori* (human) knowledge is important to not impedes agent capacities to find surprising solutions. Neural networks let them build their own representations, specific to the task they have to solve, given their sensors and effectors; representations that humans might not have designed as specific. Moreover, neural networks and gradient descent allow scalability in learning, which is an important skill to deal with decision-making in rich environments,
- a discrete set of states or actions is also a pre-defined knowledge that the agent should discover itself,
- morphology scaffolding is interesting to control the complexity of the learning : sensors and effectors of the agent should grow over time.

Thus, I would like to work on research interested in those different points.

## Publications

### International Journals

Matthieu Zimmer and Stephane Doncieux. Bootstrapping q-learning for robotics from neuro-evolution results. *IEEE Transactions on Cognitive and Developmental Systems*, 2017.

### International Conferences

Matthieu Zimmer, Yann Boniface, and Alain Dutech. Developmental reinforcement learning through sensorimotor space enlargement. In *The 8th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*, September 2018.

Matthieu Zimmer, Yann Boniface, and Alain Dutech. Neural fitted actor-critic. In *ESANN - European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, April 2016.

### International Workshops

Matthieu Zimmer, Yann Boniface, and Alain Dutech. Off-policy neural fitted actor-critic. In *Deep Reinforcement Learning Workshop, NIPS 2016*, 10 December 2016.

Matthieu Zimmer, Yann Boniface, and Alain Dutech. Toward a data efficient neural actor-critic. In *European Workshop on Reinforcement Learning*, 4 December 2016.

Matthieu Zimmer, Paolo Viappiani, and Paul Weng. Teacher-student framework: a reinforcement learning approach. In *AAMAS Workshop Autonomous Robots and Multirobot Systems*, 2014.

### National Conferences (with reviewing committee)

Matthieu Zimmer, Yann Boniface, and Alain Dutech. Vers des architectures acteur-critique neuronales efficaces en données. In *Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes*, July 2016.

### Theses and Various Reports

Matthieu Zimmer. *Developmental reinforcement learning*. PhD thesis, University of Lorraine, January 2018.

Matthieu Zimmer and Stéphane Doncieux. Construction automatique d'état et d'actions en apprentissage par renforcement. Master's thesis, University Pierre and Marie Curie, 2014.

Matthieu Zimmer, Yann Boniface, Alain Dutech, and Nicolas Rougier. Dans quelle mesure un système apprenant peut prendre conscience de ses performances et altérer son comportement. Research Report, 2012.

## Other communications

| | |
|---|---|
| November 2018 | Team presentation of the developmental layers developed during my PhD (Shanghai) |
| October 2018 | Team presentation of the NFAC framework developed during my PhD (Shanghai) |
| October 2016 | Journal club presentation of a chosen article - Unifying Count-Based Exploration and Intrinsic Motivation (Nancy) |
| June 2016 | Team presentation of my current research on deep reinforcement learning (Nancy) |
| January 2016 | Presentation for the doctoral students' day (Nancy) |
| October 2014 | Team presentation of my pre-doctoral research (Nancy) |