



ABCDE



## Scientific Report

First name / Family name

Mustafa MISIR

Nationality

Turkish

Name of the *Host Organisation*

INRIA Saclay Île-de-France

First Name / family name  
of the *Scientific Coordinator*

Marc SCHOENAUER

Period of the fellowship

15/10/2012 to 14/10/2013



## I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

I worked on a new approach to portfolio-based algorithm selection, inspired from *Collaborative Filtering (CF)*. CF systems, popularized by the Netflix challenge, make predictions about e.g. the movies that an individual user will most probably like, based on the previous movies she liked, and the movies that have been liked by other individual users.

The contribution is to cast algorithm selection as a CF problem, by stating that a problem instance “likes better” the algorithms with better performance on this particular instance. The proposed algorithm, called *Algorithm Recommender System (ARS)*, leverages the dataset reporting the performances of some algorithms on a set of problem instances, alike the existing successful portfolio approaches, with the main difference that ARS aims at the latent factors (latent descriptive features) explaining the reported performances, whereas the existing methods usually accommodates explicit descriptive features. The main challenge for CF systems is to handle the so-called *cold-start* issue when facing a brand new instance. ARS proceeds by learning to map the explicit features onto the latent descriptive features, thereby supporting the recommendation of a good algorithm for the new problem instance with constant computational cost.

The experimental validation of ARS has been conducted on various Algorithm Selection/Tuning problems including:

- Boolean Satisfiability
- Stochastic Continuous Optimisation (based on the Black-Box Optimization Benchmarking (BBOB) environment introduced in TAO INRIA)
- Constraint Satisfaction
- Hyper-parameter Tuning (collaborating with a CNRS research scientist, Dr. Balázs Kégl)
- Classification with Hyper-parameter Tuning
- Workflows for Classification (collaborating with Dr. Floarea Serban and Prof. Abraham Bernstein from University of Zurich)

All the datasets from these domains were processed first to generate problem instances.

## II – PUBLICATION(S) DURING YOUR FELLOWSHIP

- We are about to complete a journal paper that will be submitted to either IEEE Transactions on Knowledge and Data Engineering or Journal of Machine Learning Research (and first published as an INRIA technical report).



### III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

- Fête Parisienne in Computation, Inference and Optimization: A Young Researchers' Forum at IHES (Orsay, France), March 20, 2013
- DigiCosme Spring School 2013: Program Analysis and Verification at Supelec (Gif-sur-Yvette, France), April 22 – 26, 2013

### IV – RESEARCH EXCHANGE PROGRAMME (REP)

I initially visited the Laboratory for Relational Algorithmics, Complexity and Learning, Department of Software, Universitat Politècnica de Catalunya (UPC), Spain between June 17 and 21, 2013 (The local scientific coordinator is Prof. Ricard Gavaldà). The goal was to extend the application domains of ARS and to pursue collaborative research between INRIA and UPC. During my visit, I had an opportunity to discuss about using ARS in order to determine the best classification algorithms for Datastreams. After initial experiments on some automatically generated data, we have decided to work on real-world data like Twitter. Although we couldn't deliver any publishable results during my ERCIM postdoc period, we are planning to continue working on this topic. ARS also got attention of a couple of researchers working in a relatively new research area called Process Mining. We recently decided to prepare a paper focusing on the application of ARS to this domain.

I am going to visit Dynamic and Distributed Information Systems Group, Department of Informatics, University of Zurich, Switzerland between October 7 and 11, 2013 in the context of REP (The local scientific coordinator is Prof. Abraham Bernstein). We already started working together even though I didn't perform this visit yet. They provided us some datasets about determining best Workflows for a set of well-known classification datasets. The related results will be available in the aforementioned technical report. During my visit, I am going to discuss possible improvements on the current results and further collaborations.