



ERCIM "ALAIN BENSOUSSAN"  
FELLOWSHIP PROGRAMME



## Scientific Report

First name / Family name	Kalia Orphanou
Nationality	Cypriot
Name of the <i>Host Organisation</i>	CWI
First Name / family name of the <i>Scientific Coordinator</i>	Leen Stougie
Period of the fellowship	01/05/2017 to 30/04/2018

### I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Development of a new variant of the recently-introduced evolutionary algorithm family GOMEA, aimed specifically at solving the problem of learning a Bayesian network structure from data, both from a single-objective as well as a multi-objective perspective. Regarding the single-objective GOMEA, a Bayesian scoring function is used as a fitness function for the optimization. On commonly used datasets of varying size, the performance of GOMEA has been compared to other recently published results and other Genetic and Estimation-of-distribution algorithms. We find out that GOMEA outperforms standard algorithms such as Order-based search (OBS), as well as other EAs, such as Genetic Algorithms (GAs) and Estimation of Distribution algorithms (EDAs), even when efficient local search techniques are added.

Muti-objective GOMEA has been applied to the particular problem of learning Bayesian networks from data using as fitness functions, the Bayesian scoring function and an F-score representing the similarity score of a particular solution to a pre-defined Bayesian network structure build by an expert.

### II – PUBLICATION(S) DURING YOUR FELLOWSHIP

K. Orphanou, D. Thierens, P. A.N Bosman, Learning Bayesian Network Structures with GOMEA, **Accepted to:** *The Genetic and Evolutionary Computation Conference, (GECCO 2018), Kyoto, Japan, July 2018.*

**Abstract:** Bayesian networks (BNs) are probabilistic graphical models which are widely used for knowledge representation and decision making tasks, especially in the presence of uncertainty. Finding or learning the structure of BNs from data is an NP-hard problem. Evolutionary algorithms (EAs) have been extensively used to automate the learning process. In this paper, we consider the use of the Gene-Pool Optimal Mixing Evolutionary Algorithm (GOMEA). GOMEA is a relatively new type of EA that belongs to the class of model-based EAs. The model used in GOMEA is aimed at modeling the dependency structure between problem variables, so as to improve the efficiency and effectiveness of variation. This paper shows that the excellent performance of GOMEA transfers from well-known academic benchmark problems to the specific case of learning BNs from data due to its model-building capacities and the potential to compute partial evaluations when learning BNs. On commonly used datasets of varying size, we find that GOMEA outperforms standard algorithms such as Order-based search (OBS), as well as other EAs, such as Genetic Algorithms (GAs) and Estimation of Distribution algorithms (EDAs), even when efficient local search techniques are added.

### III – ATTENDED SEMINARS, WORKSHOPS, CONFERENCES

### IV – RESEARCH EXCHANGE PROGRAMME (REP)

Fraunhofer Institute SCAI, Bioinformatics Laboratory, (one week visit – 17/04/2018 – 24/04/2018)

**Scientific Coordinator:** Prof. Dr. Martin Hofmann-Apitius

**Administrative Coordinator:** Bettina Toure ([Bettina.toure@zv.fraunhofer.de](mailto:Bettina.toure@zv.fraunhofer.de))

During the REP, I got involved to a project regarding the personalized treatment of colorectal cancer that members of the laboratory are working on.

Colorectal Cancer (CRC) belongs to the third most frequent malignancy in men and women. Around 1.2 million new cases of CRC are estimated worldwide annually. The 5 year survival time in Europe is still below 60%. The clinical response to standard chemo- and radiotherapy varies greatly, and a considerable fraction of patients are therapy resistant. This imposes significant problems for the individual patient and for the whole society, because expensive and ineffective treatments with corresponding adverse side effects are executed. Second line therapies via anti-EGFR and anti-VEGF exist, but strategy of treatment lines and the important question of therapy-free intervals or maintenance therapy are still a matter of debate. Better-personalized patient treatment is seen as an important factor for better disease management.

The overall aim of this project is to address this need via an analysis of clinical data - which allows the clinician to get an integrative, system-wide, temporally resolved and patient centered view on the disease, which can also allow the user to apply filter and search criteria in order to obtain a systematic view on integrated data of specific patients or defined patient sub-groups.

My contribution to the particular project was to apply the variant of GOMEA aimed at solving the problem of learning Bayesian network structure from data, to the clinical data of Colorectal Cancer. The patient's clinical data include information about the treatments, cancer metastasis, and localization of primary and double tumor.

GOMEA has been applied to a subset of 1000/3300 patients, using 52 features, to find out the dependencies between the features that are better characterized from the data (to learn the optimal structure based on a Bayesian scoring function).