1. Short CV of the advisors

Mauro Castelli

Mauro Castelli has a Ph.D. in Computer Science obtained at the Università di Milano Bicocca (Italy). He is currently an Associate Professor at the Universidade Nova de Lisboa, and a member of the Scientific Council of NOVA Information Management School (NOVA IMS). He is also a researcher at the Information Management Research Center of this university. He participated as a principal investigator, co-principal investigator, or work package leader in different research projects at national and international levels. He is a consultant at Orbit, a startup that aims at providing a machine learning assistant that can understand and answer business queries submitted in the natural language. Finally, in 2022, he has been ranked in the top 2% world scientists by a study of the Stanford University.

He has international collaborations with researchers in more than twenty different countries and with universities recognized as leaders in the area of artificial intelligence. He is a member of ACM (Association for Computing Machinery) and SIGEVO (ACM Special Interest Group on Genetic and Evolutionary Computation). He is the author of more than 150 scientific publications and has presented about fifty seminars, conferences, and communications. He was awarded in 2013 and 2014, in the framework of the main European conference of Artificial Intelligence and Evolutionary Computing (Evo*), for the quality and contribution of his scientific research. He participated in research projects concerning the implementation of machine learning solutions with companies and organizations, such as Critical Software (Portugal), the Champalimaud Foundation (Portugal), the Directorate General of Education and Science Statistics (Portugal), and the Portuguese Ministry of Tourism.

At NOVA IMS, he is/was responsible for the following master-level courses: Deep Learning, Machine Learning, Data Mining, Decision Support Systems, and Big Data. He is also responsible for the following undergraduate courses: Algorithms Design, and Object-Oriented Programming. Finally, he is responsible for the PhD course on Genetic Programming. He was the supervisor of more than 100 master theses and five Ph.D. theses in the field of Machine Learning.

Leonardo Vanneschi

Leonardo Vanneschi is a Full Professor (Professor Catedrático) at the NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal. His main research interests involve Machine Learning, Data Science, Complex Systems, and in particular Evolutionary Computation. His work can be broadly partitioned into theoretical studies on the foundations of Evolutionary Computation, and applicative work. The former covers the study of the principles of functioning of Evolutionary Algorithms, with the final objective of developing strategies able to outperform the traditional techniques. The latter covers several different fields among which computational biology, image processing, personalized medicine, engineering, economics and logistics.

He has published more than 250 contributions, 96 of which in internationally renown scientific journals with impact factor and 11 of which have been honoured with international awards. He has led several research projects in the area, participating in a large number of other projects.

His work has been consistently recognized and appreciated by the international community from 2000 to nowadays. In 2015, he was honoured with the Award for Outstanding Contributions to Evolutionary Computation in Europe, in the context of EvoStar, the leading European Event on Bio-Inspired Computation. Since 2019, he has constantly been ranked in the top 2% world scientists both in the current year and careerwise by a study of the Stanford University.

He has been keynote invited speaker in numerous national and international events. He is a member of the editorial board of three international scientific journals and of the steering committee of several international conferences. He is a member of the steering board of the Species Society. He has regularly organized national and international conferences and workshops. He is the local chair of GECCO 2023.

Since 1999, he has consistently been teaching a large number of disciplines, broadly covering the entire area of Computer Science, for several different types of students (beginners, experts, computer scientists and coming from other areas). In the last decade, his teaching activity was particularly specialized in the areas of Intelligent Systems, Machine Learning and Data Science. He has received several awards and distinctions for his pedagogical activity.

2. Description of the research group

MagIC

The Information Management Center (MagIC) is the research center of Nova Information Management School (NOVA IMS), which is one of the worldwide leading schools in the area of Information Management. NOVA IMS is also ranked in the TOP 5 of the Best Masters and Postgraduate Programs in the world by the Eduniversal and the first European Institution to obtain the ABET accreditation.

With over 50 integrated researchers and more than 1000 published articles since 2011 (752 articles WOS or Scopus indexed), the primary objective of the MagIC working group is to contribute to the advancement of the fields of Information Management and Data Science. The research center is organised around 4 research streams: Data Science, Information Systems, Geoinformatics, and Data-Driven Marketing.

The working group of Data Science has two main branches. One is dedicated to the study of data generation algorithms for imbalanced learning. The other branch, in which the student will be working, is focused on Evolutionary Computation (EC). This includes the theoretical and practical study of EC, the development of EC models and innovative applications of these models.

In 2020, NOVA IMS and MagIC have launched a Data Analytics Lab, directed by Professor Leonardo Vanneschi. This lab is a location where students working in Artificial Intelligence and Data Science can share a very friendly environment, with comfortable space and facilities, ideal for social networking and collaborations. The majority of the students and young researchers of the Data Analytics Lab work in Evolutionary Computation.

3. Description of the work to be carried out by the student

Geometric Semantic Genetic Programming (GSGP) [1] is a promising variant of Genetic Programming (GP) that uses Geometric Semantic Operators (GSOs) instead of the standard crossover and mutation of GP. GSOs generate offspring with specific properties on the semantic space, and due much of their success to their ability to induce a unimodal error surface on the training data for any supervised learning problem. This ability bestows on GSGP a noteworthy optimization ability and opens a wide spectrum of possible research tracks concerning the ability of GSOs to generalize to unseen data [2]. The main limitations of the current version of GSGP are:

- Geometric Semantic Crossover (GSC) looks like a genetic operator characterized by several weaknesses, and the most used settings of GSGP nowadays mainly rely on Geometric Semantic Mutation (GSM) as the main operator guiding the exploration of the search space [3].
- By their very definition, GSOs generate offspring that are larger than their parents, leading to a quick growth in the code of the programs in the population. As a consequence, the final models produced by GSGP are usually black-box models, that are very hard to interpret by humans.
- Traditional GSOs are usually not able to explore the semantic space in a uniform way; in other words, several points in the semantic space are harder to obtain than others and usually remain mainly unexplored during the GSGP search.

The objective of the work carried out by the student will be to study these limitations of GSGP, proposing methods to counteract them, with the final objective of proposing new and more robust variants of GSGP.

The initial focus of the work will be on the first of the previous three points. One of the main motivations for the weakness of GSC is the fact that GSC is an effective operator only if the target is inside the convex hull defined by the population in the semantic space. The first objective of the study will be to deepen our understanding of this problem by defining simple and efficient methods to test the presence or absence of a point inside a convex hull that can outperform the existing ones in terms of efficiency. Then, the study will focus on methods to efficiently evolve populations whose convex hull contains the target point. The idea is to define a new GSGP variant that clearly identifies two phases: the first phase that takes place when the target point is not yet inside the population's convex hull, whose objective is to "move" the population until its convex hull includes the target point, and a second phase that takes place when the target point is inside the population's convex hull.

Inspired by the work [4], presented at EuroGP 2023, the idea is to use a mixture of non-semantic genetic operators, plus GSM, in the first phase, and mainly only GSC in the second phase. A better understanding and definition of how and when this alternation of operators should take place is also a focus of this study.

Concerning the other two weaknesses of GSGP, we hypothesize that an alternation of semantic and non-semantic genetic operators could help us to generate simpler, and thus more interpretable final models;

a hypothesis that will be studied by the student during his stage at NOVA IMS. Also, an objective of the student's work will be to verify if the semantic space will be explored more uniformly with this mixture of genetic operators, compared to standard GSGP, and draw appropriate conclusions aimed at establishing relationships between the exploration of the semantic space and the performance of the algorithm.

References

[1] Moraglio, A., Krawiec, K., Johnson, C.: Geometric semantic genetic programming. In: Coello, C.C., Cutello, V., Deb, K., Forrest, S., Nicosia, G., Pavone, M. (eds.) Parallel Problem Solving from Nature - PPSN XII. Lecture Notes in Computer Science, vol. 7491, pp. 21–31. Springer (2012)

[2] Vanneschi, L.: In: Schutze, O., Trujillo, L., Legrand, P., Maldonado, Y. (eds.) An Introduction to Geometric Semantic Genetic Programming, pp. 3–42. Springer, Cham (2017)

[3] Castelli, M., Silva, S., Vanneschi, L.: A C++ framework for geometric semantic genetic programming. Genetic Programming and Evolvable Machines 16(1), 73–81 (2015)

[4] Ferreira, J., Castelli, M., Manzoni, L., Pietropolli, G. (2023). A Self-Adaptive Approach to Exploit Topological Properties of Different GAs' Crossover Operators. In: Pappa, G., Giacobini, M., Vasicek, Z. (eds) Genetic Programming. EuroGP 2023. Lecture Notes in Computer Science, vol 13986. Springer, Cham. https://doi.org/10.1007/978-3-031-29573-7_1

4. Other relevant information

Lisbon is the capital and the largest city of Portugal. It is recognised as an alpha-level global city because of its importance in finance, commerce, media, entertainment, arts, international trade, education and tourism. It is a very multifaceted and multi skilled city. Morevoer, it is a safe town, which foster the dissemination of culture all around - a perfect environment for students. The climate of Lisbon is great and the city was considered the 23rd World's Best City to Live by Monocle Magazine (July / August 2011).

NOVA IMS is located in an excellent neighborhood of Lisbon, with public transportation and all facilities nearby. The School was founded in 1989 and currently it provides education at the highest level to more than 3.000 students from more than 80 countries. They are undergraduates, post-graduates, masters and PhD degree students. The vocation of NOVA IMS is to contribute in a pioneering and leading way to the conversion of data into value through Data Science, by an integrated and properly articulated commitment to teaching, research, and third mission initiatives. NOVA IMS é driven by what is genuinely central to the best higher education institutions. In the journey of affirmation for the quality of NOVA IMS is formed by the attraction of great students, teachers, and researchers, accompanied by support services provided by great professionals; a human community that through an agile, uncomplicated, vibrant, motivating environment, based on meritocracy and a strong internal cohesion, is encouraged to dream and fulfill dreams in a context of great intellectual and creative freedom.